dHealth 2019 – From eHealth to dHealth
D. Hayn et al. (Eds.)

doi:10.3233/978-1-61499-971-3-57

Improving the Prediction of Emergency Department Crowding: A Time Series Analysis Including Road Traffic Flow

Jens RAUCH[a,1], Ursula HÜBNER[a], Mathias DENTER[b] and Birgit BABITSCH[c]
[a]*Health Informatics Research Group, Osnabrück University AS, Germany*
[b]*Klinikum Osnabrück, Germany*
[c]*New Public Health, University Osnabrück, Germany*

Abstract. Background: Crowding in emergency departments (ED) has a negative impact on quality of care and can be averted by allocating additional resources based on predictive crowding models. However, there is a lack in effective external overall predictors, particularly those representing public activity. Objectives: This study, therefore, examines public activity measured by regional road traffic flow as an external predictor of ED crowding in an urban hospital. Methods: Seasonal autoregressive cross-validated models (SARIMA) were compared with respect to their forecasting error on ED crowding data. Results: It could be shown that inclusion of inflowing road traffic into a SARIMA model effectively improved prediction errors. Conclusion: The results provide evidence that circadian patterns of medical emergencies are connected to human activity levels in the region and could be captured by public monitoring of traffic flow. In order to corroborate this model, data from further years and additional regions need to be considered. It would also be interesting to study public activity by additional variables.

Keywords. emergency hospital service, patients, forecasting, regression analysis

1. Introduction

Crowding in emergency departments (ED) is associated with poor patient care, higher mortality and negative impact on patient-safety [1]. Crowding occurs when the need for emergency services exceeds available resources for patient care in the emergency department, hospital, or both [2]. An issue of misaligned demand and supply [3], it is therefore vital to relieve overstrained EDs by curbing patient volume as well as adjusting ED resources to better meet demand during highly busy hours. With respect to patients, there are many advances in understanding their motives and patterns of frequent use informing increasing demand [4]. These can be used for countering improper ED use and raising awareness among patients. On the other hand, it is of equal importance to tackle resource alignment on the hospital's side, since there is often a mismatch between staffing rosters and patient demand [5]. Growing adoption of health IT technology, however, holds the chance to integrate detailed forecasting models into ED processes and resource management, as electronic health care records become widely available [6].

[1] Corresponding Author: Jens Rauch, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.rauch@hs-osnabrueck.de.

Several recent studies could show that relatively simple regression models already give good predictions on forthcoming ED crowding measures [7,8,9]. So far, more complex models provide only somewhat more accurate results, despite their higher modelling capacities. This is foremost due to the lack of appropriate external covariates [10]. While there is a large portion of emergency research devoted to the study of surges in patient volume in case of catastrophic events, pandemic outbreaks or seasonal fluctuations, little has been published with regard to covariates of regular variations in intra-day ED occupancy [3]. Indeed, only a small number of indicators for upcoming ED service demand have been studied. Among the covariates often used are weather and calendric data, so far. The use of weather data is justified by the fact that weather affects a number of conditions, which are likely to lead to medical emergencies, e.g. certain air masses increase asthma hospital admissions significantly [11]. There were few attempts to include other public data. A recent study investigated the predictive website traffic on a public health portal [12]. Other studies revolve around the impact of mass gatherings on the demand for emergency services. Only recently, first results of a systematic examination regarding the effect of mass gatherings on medical emergency prevalence were presented [13].

Common to all these studies is the use of variables, which are inherently connected to public activity level. The use of calendric data is usually intended to reflect the influence that weekends, holidays and seasons have on public activity patterns. Aside from its physiological effects [14], weather evidently has effects on public activities, too [15]. However, both kinds of data are only indirectly connected to overall activity levels.

It seems reasonable to including more proximate measures of public activity into forecasting models of ED crowding. For instance, daily commuters can make up a considerable portion of the number of people in a city during specific times and working hours [16]. Accordingly, it can be hypothesized, that they also contribute to medical incidents subject to emergency care. Data readily available about population movement and sojourn is given by public recordings of traffic data. In this paper, we propose hourly traffic flow as a direct and openly available measure of public activity and aim at investigating its effect on the prediction of the overall hourly ED load.

2. Methods

We conducted a retrospective study and used historical ED data extracted from the Electronic Health Record of Klinikum Osnabrück, an academic teaching hospital with 660 beds serving the town and region of Osnabrück, Lower Saxony, Germany. The ED has about 40,000 cases per year and is operated 24 hours a day on 365 days a year. Since exported data was anonymised, no ethical statement needed to be obtained. Data covered the period from January 1 until December 15, 2017. Historical traffic data for the same period was obtained from the German Federal Highway Research Institute (Bundesanstalt für Straßenwesen), which collects hourly data about the direction and number of vehicles passing by measuring stations on federal roads and motorways. There is a total of six measuring stations in the area of Osnabrück, covering the major traffic axes for motorised vehicles from and to Osnabrück. All data sources were mirrored in a PostgreSQL 10 research database. Data analysis was carried out using the statistics software R.

ED occupancy is a common measure for ED crowding [17]. Thus, hourly totals of patients within the ED was the variable of interest, while traffic flow as a measure of activity should act as a preceding covariate. Cross-correlation analysis was performed to examine correlation of traffic flow with the number of ED occupancy of concurrent ED cases. Incremental 1-hour lags of the traffic data were compared stepwise to ED occupancy to assess the temporal preceding effect of traffic flow on ED crowding. Since ED occupancy follows a circadian pattern, we limited the number of lags to the preceding 23 hour traffic values. The pronounced periodicity of both time series may lead to an overestimated linear relationship, i.e. covariate time series of ED occupancy that follow the same circadian pattern might not improve a seasonal autocorrelative model [18]. Therefore, two SARIMA models were fitted to the ED occupancy data, one with and a second without traffic data as external predictors. The SARIMA model parameters were determined by inspection of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots of the seasonally differenced time series. Additionally, candidate parameters were validated by fitting the parameterized models on the time window January 1 to September 30 and comparing their performance with respect to the remaining period. Traffic flow data from the three roads that showed the highest cross-correlation with ED occupancy were used as predictors (Tab. 1).

A running cross-validation was employed to compare the performance of the models [19]. In this procedure, each model was fitted repeatedly, beginning from a fixed starting point of a connected time frame, which was iteratively expanded in 1 hour steps. For each iteration both models (without and with traffic as predictor) were fitted and prediction errors were calculated for a time horizon of up to six hours, resulting in six prediction values (one per hour). Data from the last quarter of the year (October 1 to December 15) was selected for this procedure. Prediction accuracy was calculated as root-mean-square error (RMSE) and mean absolute error (MAE).

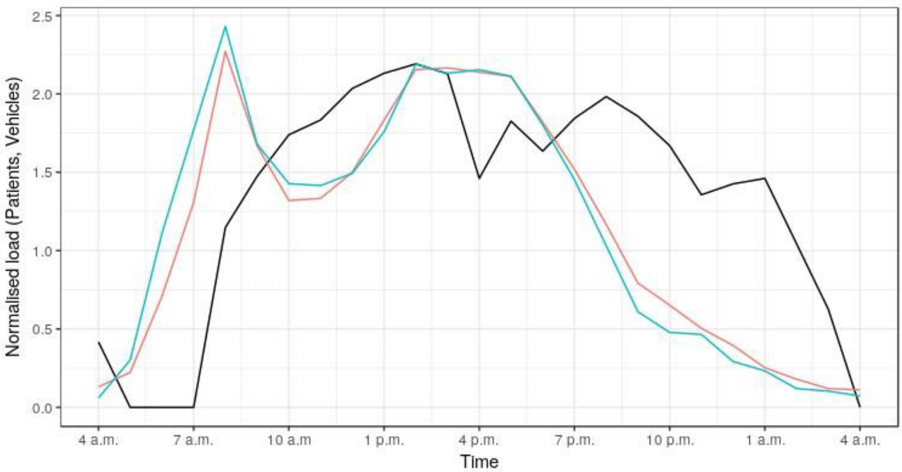

Figure 1. Typical intraday variation in traffic flow and ED occupancy (data from January 20, 2017), normalised by respective day mean (red: motorway A33 and green: federal road B51, black: ED occupancy)

3. Results

An exemplary intraday time series of ED occupancy and two selected traffic densities spanning a 24-hour period, beginning at 4 a.m. is shown in Figure 1. Obviously, traffic flow took a similar shape and was preceding ED occupancy. Increases in ED occupancy were foregone by increases in traffic load. There was a mean of 12.7 patients (± 7.86 SD) and a maximum of 37 patients in the emergency department during the period of investigation. Maximum Pearson correlation coefficients from all traffic measuring stations with their respective lags are given in Table 1 together with the approximate distance to the city centre. We also observed that in almost all cases traffic flow towards the city had a higher correlation than traffic flow in the opposite direction. The best explanation of the variance (r_{max}) was given by traffic flow on motorway A33 at Hellern and both stations on the federal road B51 with maximum correlation coefficients of .73, .71 and .71 respectively, preceding ED occupancy by two hours in all three cases. Accordingly, traffic values from these three roads were used as external predictors in the SARIMA model.

Table 1. Maximum correlation coefficients from cross-correlation analysis of road traffic and ED occupancy. Distance refers to driving distance from measuring station to Osnabrück centre. Traffic from roads included in this study is bolded.

Measuring station	Distance (km)	r_{max}	Lag (h)
B68 Lechtingen	6.0	.70	-3
B51 Ostercappeln	**13.2**	**.71**	**-2**
B51 Glandorf	**20.9**	**.71**	**-2**
A33 Hellern	**5.6**	**.73**	**-2**
A33 Fledder	6.5	.70	-3
A33 Handorf	10.2	.70	-3

Fitting of the SARIMA model to the training data period (January 1 to September 30) yielded optimal model parameters $(1,0,0),(0,1,1)_{24}$ from analysis of ACF and PACF plots. The time series was subjected to first order seasonal differencing, since no trend but strong 24-hour seasonality was present. Inspecting the ACF of the differenced time series showed a fair amount of decay. The respective PACF showed a sharp cut-off and positive autocorrelation at lag 1. Thus, an AR term was added. Since ACF showed a negative correlation at lag 24, an SMA term was added. An iterative comparison of the models by the Bayesian Information Criterion in the parameter space $(AR,MA,SMA,SAR) \in \{0,...,5\}^2 \times \{0,1,2\}^2$ confirmed, that this model was indeed optimal.

We used the same parameters for the SARIMA model with the external predictors "road traffic on the federal road B51" and "road traffic on the motorway A33 Hellern". Overall cross-validated RMSE and MAE values for all six time horizons are given in Table 2. Improvement for shorter time horizons was better. E.g. for lag 1, the SARIM model with traffic had roughly 20 % lower RMSE (4.04 vs. 3.21) and for lag 2 RMSE was about 10 % lower (4.20 vs 3.77). Inclusion of external predictors improved the prediction for all lags as could be shown (Tab. 2).

Table 2. Model comparison from cross validation over the period October 1 to December 15, 2017.

Measure	Model	Lags					
		1	2	3	4	5	6
RMSE	SARIMA	4.04	4.20	4.22	4.23	4.23	4.23
	SARIMA+traffic	3.21	3.77	3.98	4.10	4.15	4.18
MAE	SARIMA	3.11	3.23	3.24	3.25	3.25	3.25
	SARIMA+traffic	2.32	2.88	3.05	3.14	3.18	3.21

4. Discussion

In this study, prediction of hourly ED occupancy could substantially be improved when including traffic flow data into a seasonal autoregressive forecasting model. To the best of our knowledge, this is the first study making use of regional traffic data as a preceding indicator for expected ED service demand. Other studies that predicted hourly rates of ED crowding by including external covariates resorted to variables from calendric, weather and air pollution data sets. Model improvements by these data, however, fell short of expectations [20]. Moreover, the effect of weather data, which is most commonly used, showed to be highly dependent on regional and climatic features and is thus a measure unfit for generalizable assertion about its impact on ED service demand [10].

In [7], SARIMA models without external predictors where fitted to occupancies of different EDs. They achieved MAEs for one hour lags between 2.4 and 5.4 patients. Similarly, a recent study that employed deep neural networks and made use of weather and calendric data to predict patient volume had RMSEs between 4 and 5 patients for hourly predictions [21]. By and large, the present results are lower than in the other studies and show that public activity measured by road traffic can be successfully used as an effective additional predictor.

Public activity had been measured in previous studies by local mass gathering events [22,13]. For example, [22] found an increase in ED cases of about 1/1000 per participant in a mass gathering event. However, not including time series data. It might also be argued that mass gathering is a very specific predictor, i.e. one that is correlated with specific medical conditions, e.g. alcohol intoxication. We contend that road traffic acts as a general predictor of ED occupancy.

In opposition, the predictive value of road traffic cannot be causally attributed in the same way, apart from an increased risk of traffic accidents. If road traffic was only related to car accidents the predictive value would have been much lower as car accidents account for only a small proportion of the ED cases.

The quest for a general predictor that mimics the seasonal course of ED occupancy leads to findings about circadian patterns, such as heart rate, blood pressure and serotonine level [14] which are governed by the wake-sleep-rhythm and individual activity [23]. They were indeed associated with medical emergencies, a consistent finding replicated worldwide [24]. Interestingly, not all peaks, e.g. morning peak of stroke events, could be explained by biological features such as presence of hypertension, dyslipidemia, diabetes mellitus, or cigarette smoke. [24]. Individual activity patterns also seem to play a distinctive role. Notably, it was found that some medical emergencies amongst the working population follow a pattern that differed from a nonworking subgroup [25].

Obviously, biological and behavioural circadian patterns are correlates of emergency events on an individual level and thus of ED use. This study revealed circadian public activity measured by road traffic to be a meaningful predictor of overall ED occupancy on a regional level. These findings do not claim any causal relationship. Thus, more research is needed to explain the underlying mechanisms.

However, indicators such a road traffic are powerful because they are available on a regional level and thus able to predict a regional phenomenon, i.e. ED occupancy. It is therefore a natural next step, to draw on datasets other than road traffic that are indicative of public activity levels. Especially mobile cellular location data might be accessible for prediction of upcoming ED demand and could further improve the results. If so, they could corroborate the hypotheses of public activity as a correlate of ED occupancy. Apart from exogenous factors like public activity, a comprehensive predictive model of ED crowding evidently needs to incorporate factors that are endogenous to the hospital processes, e.g. staffing rosters and bed occupancy. The present study, however, focussed on the effectiveness of road traffic data as an easily publicly available predictor. Its predictive power can be expected to generalize to all EDs in a given region.

This study is of course limited in that data from the six measuring stations is a specific sample of traffic activity around the city, since neither usage of public transport nor side road traffic was given. Also, we used ED occupancy as the only measure for ED crowding. Occupancy belongs to the central throughput measures that inform ED workload. Yet, there are several other aspects that should also be taken into consideration, but which were not present in our data, e.g. ED capacity and hospital efficiency. Furthermore, it remains to be examined in what way the present findings generalize to other regions.

5. Conclusion

To the best of our knowledge, this is the first study, which examines regional traffic data as indicator for urban critical health events that require emergency treatment. It could be shown that road traffic as an external overall covariate can indeed contribute to a substantial improvement in forecasting crowding in emergency departments. Fundamentally, the effects might be explained by an inherent relation to human activity levels that previously were found to be related to medical emergencies.

6. Acknowledgements

We thank Tobias Sonnenberg and the Klinikum Osnabrück for the provision of emergency data and their collaboration. This work is funded by the state of Lower Saxony, project ROSE, the learning health care system (ZN 3103).

References

[1] B.C. Sun, R.Y. Hsia, R.E. Weiss, D. Zingmond, L.-J. Liang, W. Han, H. McCreath and S.M. Asch, Effect of Emergency Department Crowding on Outcomes of Admitted Patients, *Annals of emergency medicine* 61(6) (2013), 605–611.
[2] Crowding, *Annals of Emergency Medicine* 47(6) (2006), 585, ISSN 0196-0644, 1097-6760. doi:10.1016/j.annemergmed.2006.02.025.
[3] N.R. Hoot and D. Aronsky, Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions, *Annals of Emergency Medicine* 52(2) (2008), 126–1361, ISSN 0196-0644. doi:10.1016/j.annemergmed.2008.03.014.
[4] J. Rauch, J. Husers, B. Babitsch and U. Hübner, Understanding the Characteristics of Frequent Users of Emergency Departments: What Role Do Medical Conditions Play?, *Studies in health technology and informatics* 253 (2018), 175–179.
[5] R. Champion, L.D. Kinsman, G.A. Lee, K.A. Masman, E.A. May, T.M. Mills, M.D. Taylor, P.R. Thomas and R.J. Williams, Forecasting Emergency Department Presentations, *Australian Health Review* 31(1) (2007), 83–90, ISSN 1449-8944. doi:10.1071/ah070083.
[6] R.S. Evans, Electronic Health Records: Then, Now, and in the Future, *Yearbook of medical informatics* Suppl 1 (2016), 48–61, ISSN 0943-4747. doi:10.15265/IYS-2016-s006.
[7] M. Hertzum, Forecasting Hourly Patient Visits in the Emergency Department to Counteract Crowding, *The Ergonomics Open Journal* 10(1) (2017), 1–13.
[8] L. Zhou, P. Zhao, D. Wu, C. Cheng and H. Huang, Time Series Model for Forecasting the Number of New Admission Inpatients, *BMC medical informatics and decision making* 18(1) (2018), 39.
[9] W. Whitt and X. Zhang, A Data-Driven Model of an Emergency Department, *Operations Research for Health Care* 12 (2017), 1–15, ISSN 2211-6923. doi:10.1016/j.orhc.2016.11.001.
[10] R. Calegari, F.S. Fogliatto, F.R. Lucini, J. Neyeloff, R.S. Kuchenbecker and B.D. Schaan, Forecasting Daily Volume and Acuity of Patients in the Emergency Department, *Computational and Mathematical Methods in Medicine* 2016 (2016).
[11] P.F. Jamason, L.S. Kalkstein and P.J. Gergen, A Synoptic Evaluation of Asthma Hospital Admissions in New York City, *American Journal of Respiratory and Critical Care Medicine* 156(6) (1997), 1781–1788, ISSN 1073-449X. doi:10.1164/ajrccm.156.6.96-05028.
[12] A. Ekstrom, L. Kurland, N. Farrokhnia, M. Castren and M. Nordberg, Forecasting Emergency Department Visits Using Internet Data, *Annals of emergency medicine* 65(4) (2015), 436–442.
[13] J. Ranse, A. Hutton, T. Keene, S. Lenson, M. Luther, N. Bost, A.N. Johnston, J. Crilly, M. Cannon and N. Jones, Health Service Impact from Mass Gatherings: A Systematic Literature Review, *Prehospital and disaster medicine* 32(1) (2017), 71–77.
[14] R. Manfredini, O. La Cecilia, B. Boari, J. Steliu, V. Michelini, P. Carli, C. Zanotti, M. Bigoni and M. Gallerani, Circadian Pattern of Emergency Calls: Implications for ED Organization, *The American journal of emergency medicine* 20(4) (2002), 282–286.
[15] T. Horanont, S. Phithakkitnukoon, T.W. Leong, Y. Sekimoto and R. Shibasaki, Weather Effects on the Patterns of People's Everyday Activities: A Study Using GPS Traces of Mobile Phone Users, *PloS one* 8(12) (2013), 81153.
[16] R. Patuelli, A. Reggiani, S.P. Gorman, P. Nijkamp and F.-J. Bade, Network Analysis of Commuting Flows: A Comparative Static Approach to German Data, *Networks and Spatial Economics* 7(4) (2007), 315–331, ISSN 1572-9427. doi:10.1007/s11067-007-9027-6.
[17] L.I. Solberg, B.R. Asplin, R.M. Weinick and D.J. Magid, Emergency Department Crowding: Consensus Development of Potential Measures, *Annals of Emergency Medicine* 42(6) (2003), 824–834, ISSN 01960644. doi:10.1016/S0196-0644(03)00816-3.
[18] R.H. Shumway and D.S. Stoffer, Time Series Regression and Exploratory Data Analysis, in: *Time Series Analysis and Its Applications*, Springer, 2011, pp. 47–82.
[19] C. Bergmeir, R.J. Hyndman and B. Koo, A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction, *Computational Statistics & Data Analysis* 120 (2018), 70–83.
[20] M. Wargon, B. Guidet, T.D. Hoang and G. Hejblum, A Systematic Review of Models for Forecasting the Number of Emergency Department Visits, *Emergency Medicine Journal* 26(6) (2009), 395–399.

[21] S. Jiang, K.-S. Chin and K.L. Tsui, A Universal Deep Learning Approach for Modeling the Flow of Patients under Different Severities, *Computer Methods and Programs in Biomedicine* 154 (2018), 191–203, ISSN 0169-2607. doi:10.1016/j.cmpb.2017.11.003.

[22] J. Ranse, S. Lenson, T. Keene, M. Luther, B. Burke, A. Hutton, A.N. Johnston and J. Crilly, Impacts on In-Event, Ambulance and Emergency Department Services from Patients Presenting from a Mass Gathering Event: A Retrospective Analysis, *Emergency Medicine Australasia* (2018).

[23] R. Manfredini, M. Gallerani, F. Portaluppi and C. Fersini, Relationships of the Circadian Rhythms of Thrombotic, Ischemic, Hemorrhagic, and Arrhythmic Events to Blood Pressure Rhythms, *Annals of the New York Academy of Sciences* 783(1) (1996), 141–158, ISSN 1749-6632. doi:10.1111/j.17496632.1996.tb26713.x.

[24] R. Manfredini, F. Manfredini, B. Boari, R. Salmi and M. Gallerani, The Monday Peak in the Onset of Ischemic Stroke Is Independent of Major Risk Factors, *The American journal of emergency medicine* 27(2) (2009), 244–246.

[25] C. Spielberg, D. Falkenhahn, S.N. Willich, K. Wegscheider and H. Voller, Circadian, Day-of-Week, and Seasonal Variability in Myocardial Infarction: Comparison between Working and Retired Patients, *American Heart Journal* 132(3) (1996), 579–585, ISSN 0002-8703.