Use of Real-Time Information to Predict Future Arrivals in the Emergency Department

Yue Hu, PhD*; Kenrick D. Cato, RN, PhD; Carri W. Chan, PhD; Jing Dong, PhD; Nicholas Gavin, MD, MBA; Sarah C. Rossetti, RN, PhD; Bernard P. Chang, MD, PhD

*Corresponding Author. E-mail: yh2987@columbia.edu.

Study objective: We aimed to build prediction models for shift-level emergency department (ED) patient volume that could be used to facilitate prediction-driven staffing. We sought to evaluate the predictive power of rich real-time information and understand 1) which real-time information had predictive power and 2) what prediction techniques were appropriate for forecasting ED demand.

Methods: We conducted a retrospective study in an ED site in a large academic hospital in New York City. We examined various prediction techniques, including linear regression, regression trees, extreme gradient boosting, and time series models. By comparing models with and without real-time predictors, we assessed the potential gain in prediction accuracy from real-time information.

Results: Real-time predictors improved prediction accuracy on models without contemporary information from 5% to 11%. Among extensive real-time predictors examined, recent patient arrival counts, weather, Google trends, and concurrent patient comorbidity information had significant predictive power. Out of all the forecasting techniques explored, SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous factors) achieved the smallest out-of-sample the root mean square error (RMSE) of 14.656 and mean absolute prediction error (MAPE) of 8.703%. Linear regression was the second best, with out-of-sample RMSE and MAPE equal to 15.366 and 9.109%, respectively.

Conclusion: Real-time information was effective in improving the prediction accuracy of ED demand. Practice and policy implications for designing staffing paradigms with real-time demand forecasts to reduce ED congestion were discussed. [Ann Emerg Med. 2023;81:728-737.]

Please see page 729 for the Editor's Capsule Summary of this article.

Readers: click on the link to go directly to a survey in which you can provide **feedback** to *Annals* on this particular article.
A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter
Copyright © 2022 by the American College of Emergency Physicians.
https://doi.org/10.1016/j.annemergmed.2022.11.005

INTRODUCTION

Background and Importance

Across the United States, the focus on developing computational/machine learning models to predict demand for patient care in the emergency department (ED) has been growing within the emergency medicine field. Over the years, a variety of prediction techniques has been examined. Early studies have utilized time series models to forecast future arrivals based on recent arrival count information.[1-6] Additional studies have used other prediction models with exogenous predictors, such as linear regression, regression tree, etc.[7-11] There have also been recent efforts that explored techniques to combine time series models with exogenous features.[12,13] In addition to using appropriate prediction techniques, it is important to identify what information is most relevant in predicting emergency department demand, especially because a vast amount of information is now made available by electronic health records and various other data sources. Most of the existing literature has used classic predictors such as seasonality, holidays, weather, and previous arrival counts. A few other studies have examined limited real-time information beyond weather and previous arrival counts, including ambulance diversion status and physician workload.[14-16] However, to our knowledge, little research has explored the comprehensive patient-level and regional data that are now more readily available. Such data could provide novel additional information and improve ED demand prognostication.

An important motivation behind these developments is that predictive information about ED demand can be used to improve operational efficiency in resource allocation and

Editor's Capsule Summary

What is already known on this topic

Models to predict emergency department (ED) volumes for staffing tend toward simple analyses of prior experience.

What question this study addressed

Does adding additional information, including holidays, weather, and search terms on Google trends, increase the accuracy of volume forecasts?

What this study adds to our knowledge

Inclusion of additional information yielded 5-11% more accurate predictions of ED volumes for a single academic site. A small number of variables contributed most to improved accuracy.

How this is relevant to clinical practice

Increased prediction accuracy comes at a cost of complexity. These analyses portray what additional information is most useful in predicting ED volumes, and so develop a hypothetical foundation for the inclusion of "big data" in ED staffing.

better meet patients' needs.[17] Such proactive planning is particularly relevant for nurse staffing, as nurses provide a substantial portion of patient care and are an increasingly limited resource in the ED (eg, because of nursing shortages exacerbated by burnout and quitting during the coronavirus [COVID-19] pandemic).[18-20] Inefficient and inadequate staffing is often associated with ED crowding, reduced quality of care, clinician burnout, and reduced hospital revenue.[21-26] In the current nurse staffing practice, EDs typically divide a day into multiple shifts. The ED manager staffs most of the nurses for shifts weeks to months in advance. A few hours before the nursing shift, the ED manager could call in extra nurses with incentive pays if s/he senses a higher patient volume that renders the planned staffing level insufficient (after taking into account staffing fluctuations because of sick calls and personal emergencies). We refer to the former as base staffing and the latter as surge staffing. The ED demand forecasts synchronized with these 2 versus staffing decision epochs can greatly facilitate these decisions. Because overtime/surge staff are more expensive and less convenient for nurses, it is important to understand how much we can improve the prediction accuracy at the surge stage (when we can use more real-time information) than at the base stage (when limited information about the shift is available). A recent study shows that even a small accuracy improvement at the surge stage can lead to effective prediction-driven 2-stage (base and surge) nurse staffing policies.[27] However, little is known about whether (and if so, by how much) real-time information improves prediction accuracy in practice.

Goals of This Investigation

The goal of this study was to explore and evaluate rich real-time information (including previous arrival counts, temporal and seasonal variations, holidays, weather, electronic health records, and Google trends) and a variety of prediction techniques. By comparing prediction models with and without real-time predictors, we assessed the gain in prediction accuracy from real-time information. Lastly, we described how these 2 types of prediction models (with and without real-time information) could both contribute to a prediction-driven staffing framework.

METHODS

Study Setting and Objective

We conducted a retrospective study using data obtained from electronic health records for an adult ED in a large academic hospital in New York City. A total of 284,550 adult patients who arrived at the ED from noon January 1, 2018, through 11:59 PM January 31, 2021, were included in the analysis.

At the hospital, each day was divided into 2 main 12-hour nursing shifts that started at 7:00 AM and 7:00 PM, respectively. To facilitate relevant operational decisionmaking (eg, nurse staffing decisions), the subject of prediction was the shift-level arrival count, defined as the total number of patients who arrived at the ED during each shift. Many hospitals have more nursing shifts than the 2 listed above. In those cases, we can divide the day into non-overlapping intervals and predict the interval-level arrival count similarly.

Model fitting and selection were performed using 1 year of data from January 1, 2018, to January 31, 2019, which we hereafter refer to as the training set. Model performance was tested on the subsequent 1-year data from February 1, 2019, to January 31, 2020, which we refer to as the test set. The remaining data from February 1, 2020, to January 31, 2021, contained the outbreak of the COVID-19 pandemic, and we, thus, refer to it as the COVID test set. Because patient volume was highly unpredictable during the pandemic and the pandemic is likely a unique generational event, we relegate the results and discussions regarding the COVID test set to Appendix E1 (available at http://www.annemergmed.com). The training, test, and COVID test sets were fixed across all prediction models. This study was approved by the Columbia University institutional review board: protocol IRB-AAAT6452.

Data Source

We used 3 sources of data: patient electronic health records, weather data published by the National Centers for Environmental Information, and Google trends.[28,29] These data sources were selected based on past work, extant models, and our own novel hypotheses. Although the importance of weather information has been well established in the literature,the prediction power of real-time patient electronic health records and Google trends has been relatively underexplored.[14-16]

The data extracted from the patient electronic health records specified for each patient: (i) the patient's clinical time stamps in the ED, including arrival time, first evaluation time, admission decision time, and departure time; (ii) the arrival source of the patient, eg, walking in or by ambulance; (iii) the patient's chief complaint(s), ie, reason of visit; (iv) the patient's Emergency Severity Index (ESI); (v) laboratory tests and imaging ordered: indicators for whether laboratory tests, computed tomography (CT), magnetic resonance imaging (MRI), ultrasonography, and radiographs were ordered; (vi) indicator for whether the patient was admitted into the hospital; (vii) the Charlson comorbidity index based on a list of 17 comorbidities; (viii) age; and (ix) indicator for whether the patient left without being seen.

In addition to the patient electronic health records, we obtained retrospective daily weather information, including the minimum temperature, precipitation, snow, wind, and a hot-weather indicator for whether the maximum temperature exceeds 86 °F (30 °C).

The last source of data came from Google trends, which specified, for each day, the relative Google search volume for the words "flu," "emergency room," "abdominal pain," "respiratory infection," "chest pain," "depression," "heart attack," "abuse," "disorder," "weather," and "hospital" in New York State. We came up with a list of keywords based on existing studies and our own novel hypotheses. Araz et al[30] established that the Google trends for "flu" were able to forecast influenza-like-illness-related ED visits. Tuominen et al[31] found that the Google trends for "ED" facilitated prediction. The other Google trends keywords were constructed based on our own hypotheses. Because the most frequent reasons for ED visits were abdominal pain, respiratory infection, and chest pain, we hypothesized that the Google search volumes for these keywords were positively correlated with ED visits.[32] In addition, we hypothesized that the search volumes for "depression," "heart attack," "abuse," and "disorder" signaled relevant illnesses in the neighborhood. Moreover, the Google search record for "weather" might reflect citizens' subjective perception of weather conditions which might influence their stay-at-home/travel plans. Lastly, similar to "ED," a higher Google search volume for "hospital" might indicate that more patients were seeking care.

When selecting the data sources, we tried to be comprehensive by including as much potentially relevant information as possible. Later in the model training and feature selection section, we discuss procedures to train different prediction models and identify relevant predictors.

Data Processing

We processed the data into shift-level predictors. The data regarding day versus night, day of the week, month, season, near-holiday indicators, weather, and Google trends were readily available at the shift level. As for the data from electric health records, we constructed the following 3 categories of shift-level predictors.

The first category was the previous arrival counts, which specified for each shift the arrival count 1 day ago and 7 days ago, as well as the moving average of the shift-level arrival count over the last 30 days. More precisely, the arrival count on the previous day was the total number of patients who arrived during the previous 24 hours. The arrival count on the previous nth day was the 2 shifts between the previous 24*(n-1)th and 24*(n)th hour.

The second category of predictors was the patient comorbidity information, which we processed into the following 3 sets. The first set specified for each comorbidity the total number of patients with that comorbidity on the previous day, ie, during the previous 2 shifts, and the sum and the weighted sum of Charlson comorbidity indexes for all patients on the previous day. The second set contained similar information as the first set, but instead of considering the previous day, calculated the average daily number of patients with each comorbidity over the last 3 days, as well as the average daily sum and weighted sum of Charlson comorbidity indexes for all patients over the last 3 days. The third set calculated for each comorbidity, the percentage of patients with that comorbidity over the last 3 days, as well as the average sum and weighted sum of Charlson comorbidity indexes per patient over the last 3 days. The difference between the second and third sets was that the third set considered average comorbidity measures on the individual level and was not influenced by how many patients arrived over the last 3 days. The motivation to consider comorbidity information over the last 3 days was because of the existing findings that patients with certain comorbidities are more likely to be readmitted to the ED within 72 hours.[33,34] These 3 sets of information were likely to be correlated. Because it was a priori unclear which specification had the most predictive power, we left it to the model training and feature selection procedures to sift out redundant information and identify important features.

The third category of predictors was the recent ED volume and patient severity information on the previous day (ie, during the 24 hours before the focal shift). This included the total number of patients who arrived by ambulance, the total number of patients with ESI from 1 to 5, the total number of laboratory tests, CT, MRI, ultrasounds, and XR ordered, the total number of patients admitted to the hospital, the total number of patients whose age exceeds 65 years old, the total number of patients whose age exceeds 80 years old, the total number of patients who left without being seen, the average waiting time (from arrival time to first evaluation time), the average treatment time (from first evaluation time to discharge decision time), and the average boarding time (from discharge decision time to departure time) on the previous day. Intuitively, the waiting and boarding times captured how busy the ED was on the previous day.

Model Evaluation

We focused on 2 measures of forecast accuracy for shift-level arrival counts—the root mean square error (RMSE) and the mean absolute prediction error (MAPE). Let $(y_1, y_2, \ldots, y_n)$ be the vector of observed arrival counts for a total of n shifts, and let $(\widehat{y}_1, \widehat{y}_2, \ldots, \widehat{y}_n)$ be the corresponding vector of predicted arrival counts given by the prediction model. The RMSE was the square RMSE between the predicted and observed values:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y}\text{i} - \text{yi})^2}{n}}$$

The MAPE was the average percentage error of the prediction:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\widehat{y}\text{i} - \text{yi}|}{\text{yi}}$$

Both RMSE and MAPE are standard measures of prediction accuracy.[1-13] Hereafter, we refer to the RMSE (MAPE) calculated on the training set as the training RMSE (MAPE) and on the test set as the test RMSE (MAPE). In addition to the overall RMSE and MAPE, we also examined the over-estimation and under-estimation errors separately.

Model Training and Feature Selection

Using the predictors developed in the Data Processing section, we examined various prediction models. For the baseline models without real-time information, as we had relatively few predictors, we trained linear regression and regression tree models only. As we incorporated more real-time information, in addition to linear regression and regression tree, we trained more sophisticated models, including extreme gradient boosting (XGBoost), seasonal autoregressive integrated moving average (SARIMA), and SARIMA embedded with linear regression (SARIMAX). Comparatively, linear regression and regression tree models are highly interpretable statistical models but may be inadequate for nonlinear or autocorrelated data. The SARIMA and SARIMAX models are time series models that are effective in modeling seasonal trends and autocorrelation. XGBoost is a sophisticated black-box model for complex and nonlinear relationships but is less interpretable than the other models.[10] To select the relevant features for linear regression, we used a modulated 2-way stepwise model selection method based on Akaike's information criterion (AIC). For the regression tree and XGBoost, we used 10-fold cross-validation for hyperparameter tuning. For the time series models, we used a variation of the Hyndman-Khandakar algorithm[35] to determine the hyperparameters. Detailed training and feature selection procedures for each model are provided in Appendix E2 (available at http://www.annemergmed.com).

RESULTS

Models without Real-Time Information

We referred to the linear regression model without real-time information as LR1. The significant covariates in LR1 were day versus night, day of the week, month, and holidays. On the test set, LR1 achieved an RMSE of 16.425 and an MAPE of 9.627%. Table 1 lists the estimated coefficients for the covariates in LR1. We refer to the tree model without real-time information as TR1, which had hyperparameters cp = 0.01 and maxdepth = 7. Figure 1 illustrates the structure of TR1. TR1 performed similarly to LR1 on the test set and achieved test RMSE of 16.644 and test MAPE of 9.353%.

Models with Real-Time Information

Linear regression. We referred to the linear regression model with real-time information as LR2. It contained the following predictors: day versus night, day of the week, season, holidays, weather, the total number of arrivals 1 and 7 days ago, the moving average of daily arrival count over the last 30 days, Google trends for "flu," "respiratory infection," "depression," "heart attack," "abuse," "weather," and "hospital, and the average daily numbers of patients with comorbidity "HP" (hemiplegia or paraplegia),

Table 1. Estimated 95% confidence intervals for the coefficients of covariates in LR1, LR2, and ARIMAX (3, 1, 4).

Covariate	LR1	LR2	ARIMAX
(Intercept)	(82.954, 93.912)	(40.041, 165.262)	NA
Monday day	(113.957, 125.610)	(114.435, 125.615)	(113.382, 128.438)
Monday night	(3.784, 15.437)	(3.620, 16.707)	(5.719, 16.199)
Tuesday day	(91.385, 103.079)	(91.313, 104.781)	(92.285, 107.536)
Tuesday night	(0.288, 11.983)	(0.860, 13.486)	(2.217, 13.428)
Wednesday day	(90.286, 101.881)	(90.611, 103.142)	(90.727, 106.277)
Wednesday night	(-2.867, 8.727)	(-2.334, 10.078)	(-1.043, 9.901)
Thursday day	(89.011, 100.577)	(88.355, 100.533)	(88.250, 103.409)
Thursday night	(-0.989, 10.577)	(-1.515, 10.850)	(-0.382, 10.772)
Friday day	(78.765, 90.382)	(77.643, 90.024)	(77.419, 93.023)
Friday night	(0.285, 11.902)	(0.115, 13.058)	(0.735, 12.967)
Saturday day	(50.904, 62.516)	(51.835, 64.912)	(51.470, 67.691)
Saturday night	(-1.961, 9.651)	(-0.265, 12.516)	(0.045, 12.018)
Sunday day	(45.866, 57.365)	(47.924, 60.746)	(47.012, 63.448)
January	(0.888, 11.453)	NA	NA
February	(4.473, 15.292)	NA	NA
March	(-8.061, 2.530)	NA	NA
April	(-7.621, 3.061)	NA	NA
May	(-2.615, 7.933)	NA	NA
June	(-5.389, 5.289)	NA	NA
July	(1.364, 11.908)	NA	NA
August	(-1.765, 8.832)	NA	NA
September	(-2.706, 7.923)	NA	NA
October	(0.292, 10.838)	NA	NA
November	(-8.843, 1.806)	NA	NA
Fall	NA	(-6.185, 1.684)	(-6.102, 1.843)
Summer	NA	(-5.770, 3.182)	(-5.806, 3.237)
Winter	NA	(-2.920, 7.158)	(-2.978, 7.531)
Holiday	(-29.459, -15.608)	(-30.387, -16.367)	(-30.600, -17.402)
Holiday – 1 day	(-17.293, -3.456)	(-17.416, -3.808)	(-17.879, -4.844)
Holiday + 1 day	(8.760, 22.594)	(8.709, 22.486)	(8.496, 21.584)
Min temperature	NA	(0.267, 0.701)	(0.274, 0.702)
Precipitation	NA	(-0.257, -0.043)	(-0.247, -0.049)
Snow	NA	(-0.231, -0.100)	(-0.230, -0.109)
Wind	NA	(0.003, 0.149)	(0.008, 0.145)
Max temperature $\geq$ 86 °F	NA	(-9.508, -1.155)	(-8.879, -0.749)
Recent arrival count 1-day prior	NA	(-0.039, 0.065)	NA
Recent arrival count 7-day prior	NA	(-0.006, 0.090)	(-0.010, 0.089)
30-day moving average	NA	(-0.749, 0.217)	(-0.772, 0.210)
Google trend “abuse”	NA	(-0.295, 0.007)	(-0.315, -0.007)
Google trend “depression”	NA	(-0.230, 0.114)	(-0.234, 0.113)
Google trend “flu”	NA	(0.142, 0.531)	(0.153, 0.547)
Google trend “heart attack”	NA	(-0.198, 0.062)	(-0.197, 0.065)
Google trend “hospital”	NA	(-0.045, 0.711)	(-0.038, 0.726)
Google trend “respiratory infection”	NA	(-0.061, 0.198)	(-0.059, 0.203)
Google trend “weather”	NA	(-0.149, 0.132)	(-0.151, 0.131)

Table 1. Continued.

Covariate	LR1	LR2	ARIMAX
Total # patients with comorbidity CANC over the last 3 days	NA	(-0.001, 1.976)	(0.043, 1.895)
Total # patients with comorbidity HP over the last 3 days	NA	(-7.601, 2.989)	(-7.049, 2.764)
Total # patients with comorbidity REND over the last 3 days	NA	(-1.659, 0.044)	(-1.629, -0.043)
AR1 (ϕ_1)	NA	NA	(-0.758, 0.144)
AR2 (ϕ_2)	NA	NA	(-0.320, 0.363)
AR3 (ϕ_3)	NA	NA	(-0.988, -0.266)
MA1 (θ_1)	NA	NA	(-1.139, -0.304)
MA2 (θ_2)	NA	NA	(-0.557, 0.175)
MA3 (θ_3)	NA	NA	(0.235, 0.956)
MA4 (θ_4)	NA	NA	(-0.986, -0.379)

LR1, linear regression model without real-time information; *LR2,* linear regression model with real-time information; *ARIMAX*, Autoregressive Integrated Moving Average with eXogenous factors); *CANC*, cancer; HP, hemiplegia or paraplegia; REND, renal disease; AR,;MA,; $\boldsymbol{\phi}$; $\boldsymbol{\theta}$

"CANC" (cancer), and "REND" (renal disease) over the last 3 days. The LR2 achieved a test RMSE of 15.366 and a test MAPE of 9.109%. Table 1 lists the estimated coefficients for the covariates in LR2.

Regression tree. We referred to the tree model with real-time information as TR2, which had hyperparameters cp = 0.01 and maxdepth = 7. Note that the model trained without versus with real-time predictors (TR1 versus TR2 [Figure 1]) were identical.

XGBoost. The XGBoost model had the following hyperparameters: (i) a number of boosting rounds (num_round) equal to 180, (ii) maximum tree depth for base learners (max_depth) equal to 3, (iii) boosting learning rate (eta) equal to 0.1, (iv) L1 regularization term on weights (alpha) equal to 0.2, and (v) L2 regularization term on weights (lambda) equal to 0.8. Figure 2 illustrates the top 20 most informative predictors identified by the selected model, including day versus night, day of the week, month, holidays, weather, Google trends for "respiratory infection," "disorder," and "weather," the daily average number of patients with comorbidity "AIDS" (acquired immunodeficiency syndrome) over the last 3 days, and the percentages of patients with comorbidity "CEVD" (cerebrovascular disease) over the last 3 days. The final model achieved a test RMSE of 16.315 and a test MAPE of 9.582%.

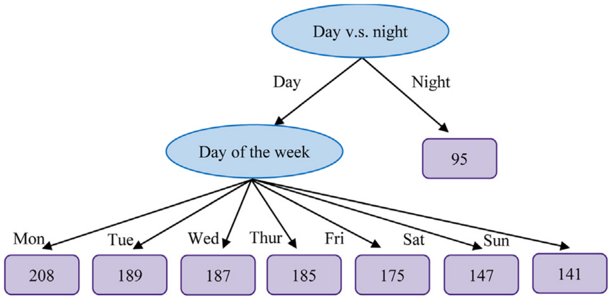

Figure 1. Visualization of TR1 and TR2. *TR1* and *TR2*, are regression trees that can be interpreted from the visualization as follows: 1) Start from the root node ("Day versus night"). 2) Go to the next node if the covariate at the root node is equal to the value specified by the edge. 3) The predicted value is given at the leaf node. For example, the predicted arrival count during a Monday day shift is 208.

SARIMA and SARIMAX. Among all SARIMA models, SARIMA (6,0,7) $(7,1,3)_{14}$ was selected, achieving a test RMSE of 15.501 and a test MAPE of 8.817%. After incorporating the external regressors and setting the seasonal term to 0, the final ARIMAX(3,1,4) model achieved a test RMSE of 14.656 and a test MAPE of 8.703%. Table 1 lists the estimated coefficients in the ARIMAX(3,1,4) model. As expected, the coefficients for the exogenous covariates had the same signs (ie, directional trends) as those for the final LR2. Moreover, as explicitly derived in Appendix E2, the coefficients suggested a positive correlation between the arrival count during the current shift and the arrival counts during the previous 2 days.

Comparison of different prediction models. For each prediction model examined, Table 2 summarizes the RMSE and MAPE on the training and test sets, and Table 3 lists the RMSE and MAPE associated with overprediction and underprediction instances. Among models that did not use real-time information, the LR1 performed the best on the

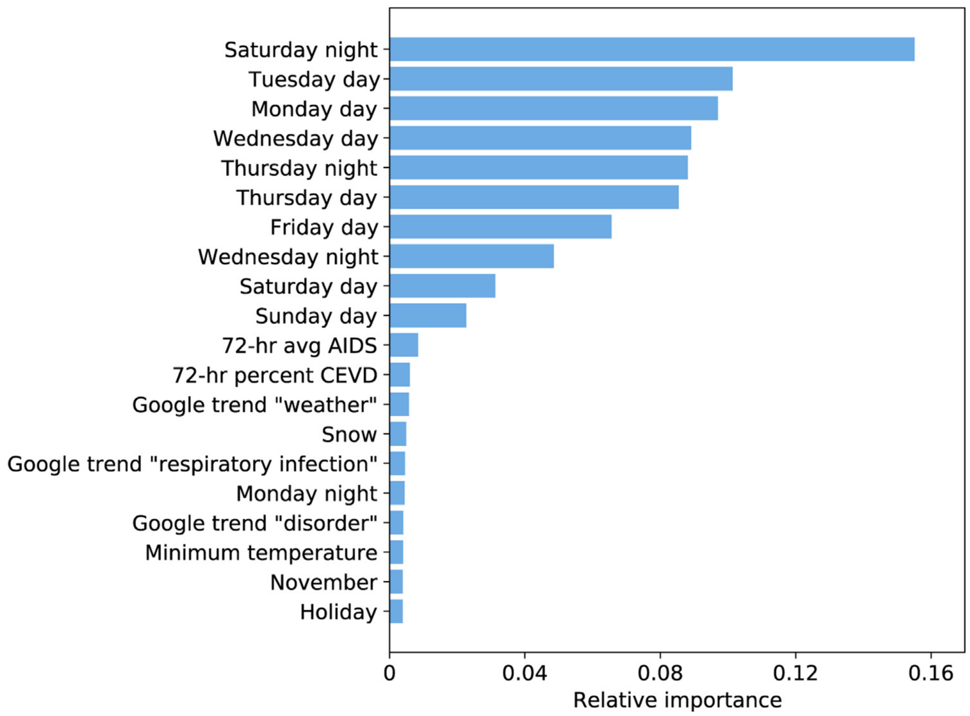

Figure 2. Top 20 informative predictors in the final extreme gradient boosting (XGBoost) model.

test set. After incorporating real-time information, the prediction accuracy of the test set can be improved. The ARIMAX achieved the best performance among models that used real-time information, improving prediction accuracy from LR1 by 10.770% (in test RMSE) and 9.598% (in test MAPE). The LR2 achieved the second-best performance, with a 6.630% reduction in test RMSE and a 5.381% reduction in test MAPE compared with LR1.

LIMITATIONS

The limitations of the study include the limited amount of training data. The training set only contained 1 year of data with 730 observations, which limited the performance of more sophisticated models that required substantial hyperparameter tuning, such as XGBoost. In addition, our study was performed for a single quaternary care facility in New York City. A meaningful extension is to apply our analysis to multiple ED sites and compare the prediction accuracy and trends. That said, the directional and structural insights (eg, procedures to develop prediction models and the value of real-time information) should be valid across facilities.

DISCUSSION

Our work employed rich real-time information to build prediction models for ED demand which can be an integrated part of the 2-stage nurse staffing framework. Existing studies have applied different prediction techniques to forecast ED arrivals but have not explored comprehensive real-time information as done in our study.[36] By exploring a

Table 2. Comparison of the selected models.

Model	Utilize real-time information	Training RMSE	Training MAPE (%)	Test RMSE	Test MAPE (%)
LR1	No	14.643	9.253	16.425	9.627
TR1/TR2	No	15.979	9.590	16.644	9.353
LR2	Yes	13.892	8.884	15.336	9.109
XGBoost	Yes	8.051	5.500	16.254	9.455
SARIMA	Yes	13.902	7.797	15.501	8.817
ARIMAX	Yes	13.604	8.618	14.656	8.703

XGBoost, extreme gradient boosting. See the Model Evaluation section for detailed definition.

Table 3. Overprediction and underprediction error.

	Training set				Test set			
	RMSE		MAPE (%)		RMSE		MAPE (%)	
Model	Overprediction	Under-	Over-(%)	Under	Over-	Under-	Over-	Under-
LR1	15.242	14.052	10.839	7.752	13.763	18.215	10.769	8.746
TR1/TR2	16.811	15.155	11.336	7.930	14.153	18.423	10.385	8.546
LR2	14.209	13.574	10.114	7.682	13.267	16.961	10.253	8.128
XGBoost	8.132	7.969	6.669	4.331	13.972	17.973	11.005	8.132
SARIMA	15.626	14.759	10.886	7.826	14.681	16.235	9.804	7.887
ARIMAX	13.989	13.241	9.877	7.465	13.528	15.597	9.634	7.868

See the Model Evaluation section for detailed definition.

novel large set of real-time predictors from the concurrent patient electronic health records, weather, and Google trends, we demonstrated that this real-time information was able to improve demand forecasts compared with base prediction models. The improvement in prediction accuracy can be used to develop prediction-driven 2-stage staffing policies to improve operational efficiency.

Noninferiority of the "Tried-And-True" Prediction Models

As illustrated by Tables 2 and 3, LR2 and ARIMAX achieved the best performance among all prediction models that used real-time information, improving prediction accuracy by 5% to 1% in RMSE and MAPE than models without real-time information (LR1). The worse performance of the regression tree and SARIMA models was well expected because of their relatively simple structure, eg, the SARIMA models only took previous arrival counts into account. On the other hand, the performance of the more advanced XGBoost model could be impeded by overfitting, eg, the XGBoost model was trained with 128 features on 730 observations (shifts) only. The XGBoost model also had the disadvantage of lacking interpretability, which was especially concerning in health care settings because of the high-stakes decisionmaking. Hence, by establishing the noninferiority of the "tried-and-true" linear regression and time series models (embedded with exogenous variables), we provided the foundation for ED managers to deploy more interpretable models.

Relevant Real-Time Information in Predicting ED Demand

Among the extensive amount of real-time information examined, only a few real-time predictors had predictive power and were coherently identified by different prediction models. According to the estimated coefficients by LR2 and ARIMAX (Table 1), ED arrivals were positively correlated with the patient volume 1 day and 7 days prior. Severe weather, such as snow, precipitation, and extremely cold or hot temperature, could reduce ED arrivals. Nevertheless, the ED tended to see more patients on days with strong wind. In addition, ED arrivals increased during the weeks when there were more Google search records for "flu." Intuitively, the search volume for "flu" could be seen as the concurrent flu trend information. Moreover, the total number of patients with a history of cancer (CANC) over the last 72 hours was positively correlated with ED arrivals. This trend could be corroborated by the findings that patients with a higher weighted sum of Charlson comorbidity indexes were more likely to return to the ED within 72 hours.[33,34] The selected XGBoost model identified similar significant predictors (Figure 2), with several new features such as the Google trends for "disorder," the percentages of patients with comorbidities of cerebrovascular disease (CEVD) and AIDS over the last 3 days.

Implication for Prediction-Driven Staffing

The development of accurate prediction models for ED demand was an integrated part of our efforts in using predictive analytics to facilitate better medical resource planning. As mentioned before, ED staffing generally involves 2 stages: a base stage, which takes place weeks to months ahead of the actual shift, and the surge stage, which happens days to hours before the shift starts. The base prediction model without real-time information can be used to guide the base staffing decision, whereas the more sophisticated prediction model with real-time information can be used to guide surge staffing decisions. At the base stage, the staffing cost is lower and more preferable by nurses based on consistency and predictability of work hours. However, the accuracy of the prediction model may be low. On the other hand, at the surge stage, the staffing cost is higher, but a more accurate prediction of patients' demand is available. How to optimally balance the tradeoff depends on how much real-time predictors improve prediction accuracy over the base prediction. Our results provide important quantification of this, which can be incorporated into the 2-stage staffing framework developed by Hu et al[27] to reduce the staffing cost and ED waiting times. We note that even relatively small prediction accuracy improvement, ie, 5% to 11% as found in our study, can lead to significant cost savings, 11% to 16% as demonstrated in Hu et al.[27] Lastly, we remark that alternative prediction targets other than shift-level arrival counts could be used in the prediction-driven staffing framework. In Appendix E3 (available at http://www.annemergmed.com), we constructed logistic regression models to predict "outlier" shifts that would have demanded surges and obtained similar insights on the value of real-time information. That said, predicting shift-level arrival counts (compared to a binary indicator on whether there would be demand surge) led to more actionable staffing implications.

In conclusion, we constructed and evaluated prediction models with rich real-time information to forecast ED patient volume. In alignment with the nursing shift structure in an ED site at a quaternary care facility in New York City, we aimed to predict the shift-level patient arrival count. Various prediction techniques were examined, including linear regression, regression tree, XGBoost, SARIMA, and (S)ARIMAX. Based on the data from our partner ED site, linear regression and ARIMAX, when combined with real-time information, achieved the highest prediction accuracy measured by RMSE and MAPE. Compared with prediction models without real-time predictors, we found that contemporary information was able to improve prediction accuracy in near-real time. Among the extensive list of real-time predictors tested, recent patient arrival counts, weather, Google trends, and concurrent patient comorbidity information had the highest predictive power. The effectiveness of real-time information in improving demand forecasts has policy implications for staffing. The ED management can use real-time demand forecasts to make timely adjustments to staffing levels, which, in turn, can effectively mitigate ED crowding.

Supervising editor: Stephen Schenkel, MD, MPP. Specific detailed information about a possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author Affiliations: From the Decision, Risk, and Operations Division (Hu, Chan, Dong), Columbia Business School, New York, NY; School of Nursing (Cato, Rossetti), Columbia University, New York, NY; the Office of Nursing Research, EBP, and Innovation (Cato), New York-Presbyterian Hospital, New York, NY; Department of Emergency Medicine (Cato, Gavin, Chang), New York, NY; Department of Biomedical Informatics (Rossetti), Columbia University, New York, NY, USA.

Author contributions: Design of work—all authors; Data acquisition— BPC, SCR; Analysis—CWC, JD, YH; Interpretation of Results—all authors; Drafting the manuscript—CWC, JD, YH; Revision of manuscript for critically important intellectual content—all authors; Final approval of the version to be published—all authors; Agreement to be accountable for all aspects of the work—YH. YH takes responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). YH, KDC, NG, SCR, and BPC report no conflicts of interest. This project received support from the Columbia University Mid-Career Faculty Grants (CWC) and the NSF CAREER Award CMMI-1944209 (JD).

Publication dates: Received for publication March 3, 2022. Revision received October 1, 2022. Accepted for publication November 8, 2022.

REFERENCES

1. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med*. 1994;23:299-306.
2. Morzuch BJ, Allen PG. Forecasting hospital emergency department arrivals. *26th Annual Symposium on Forecasting, Santander, Spain*. 2006;11–14.
3. Schweigler LM, Desmond JS, McCarthy ML, et al. Forecasting models of emergency department crowding. *Acad Emerg Med*. 2009;16:301-308.
4. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. *Emerg Med*. 2012;29:358-365.
5. Asheim A, Bjørnsen LPBW, Næss-Pleym LE, et al. Real-time forecasting of emergency department arrivals using prehospital data. *BMC Emerg Med*. 2019;19:1-6.
6. Choudhury A. Hourly forecasting of emergency department arrivals: time series analysis. *Br J Health Care Manag*. 2020;26:34–43.
7. Holleman DR, Bowling RL, Gathy C. Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. *J Gen Intern Med*. 1996;11:237-239.
8. Batal H, Tench J, McMillian S, et al. Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med*. 2001;8:48-53.
9. Zibners LM, Bonsu BK, Hayes JR, et al. Local weather effects on emergency department visits: a time series and regression analysis. *Pediatr Emerg Care*. 2006;22:104-106.
10. Jones SS, Thomas A, Evans RS, et al. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15:159-170.
11. Marcilio I, Hajat S, Gouveia N. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Acad Emerg Med*. 2013;20:769-777.

12. Calegari R, Fogliatto FS, Lucini FR, et al. Forecasting daily volume and acuity of patients in the emergency department. *Comput Math Methods Med*. 2016.
13. Whitt W, Zhang X. Forecasting arrivals and occupancy levels in an emergency department. *Oper Res Health Care*. 2019;21:1-18.
14. Brillman JC, Burr T, Forslund D, et al. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance. *BMC Med Inform Decis Mak*. 2005;5:1-14.
15. McCarthy ML, Zeger SL, Ding R, et al. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15:337-346.
16. Chase VJ, Cohn AE, Peterson TA, et al. Predicting emergency department volume using forecasting methods to create a "surge response" for noncrisis events. *Acad Emerg Med*. 2012;19:569-576.
17. Joseph JW, White BA. Emergency department operations: an overview. *Emerg Med Clin North Am*. 2020;38:549-562.
18. Rodriguez RM, Montoy JCC, Hoth KF, et al. Symptoms of anxiety, burnout, and PTSD and the mitigation effect of serologic testing in emergency department personnel during the COVID-19 pandemic. *Ann Emerg Med*. 2021;78:35-43.
19. Chang BP, Gallos G, Wasson L, et al. The unique environmental influences of acute care settings on patient and physician well-being: a call to action. *J Emerg Med*. 2018;54:e19.
20. Curtis JR, Puntillo K. Is there an epidemic of burnout and post-traumatic stress in critical care clinicians? *Am J Respir Crit Care Med*. 2007;175:634-636.
21. Johnson KD, Winkelman C. The effect of emergency department crowding on patient outcomes: a literature review. *Adv Emerg Nurs J*. 2011;33:39-54.
22. Ball J, Griffiths P, Hope J. Evidence on the effect of nurse staffing levels on patient outcomes. *Nurs Times*. 2017;113:48-49.
23. Pines JM, Batt RJ, Hilton JA, et al. The financial consequences of lost demand and reducing boarding in hospital emergency departments. *Ann Emerg Med*. 2011;58:331-340.
24. Jo S, Kim K, Lee JH, et al. Emergency department crowding is associated with 28-day mortality in community-acquired pneumonia patients. *J Infect*. 2012;64:268-275.
25. McCusker J, Vadeboncoeur A, Lévesque JF, et al. Increases in emergency department occupancy are associated with adverse 30-day outcomes. *Acad Emerg Med*. 2014;21:1092-1100.
26. Lall MD, Chang BP, Park J, et al. Are emergency physicians satisfied? An analysis of operational/organization factors. *J Am Coll Emerg Phisicians Open*. 2021;2:e12546.
27. Hu Y, Chan CW, Dong J. Prediction-driven surge planning with application in the emergency department. Accessed July 1, 2022. http://www.columbia.edu/~yh2987/Files/Prediction_Driven_Surge_Planning_Nonblind.pdf
28. National Centers for Environmental Information. Global historical climatology network (GHCN)-daily dataset. 2020. Accessed December 8, 2020. https://www.ncdc.noaa.gov/cdo-web/cart
29. Google Trends. 2020. Accessed April 26, 2020. https://www.ncdc.noaa.gov/cdo-web/cart
30. Araz OM, Dan B, Robert LM. Using Google flu trends data in forecasting influenza-like–illness related ED visits in Omaha, Nebraska. *Am. J. Emerg. Med*. 2014;32:1016-1023.
31. Tuominen J, Francesco L, Niku O, et al. Forecasting daily emergency department arrivals using high-dimensional multivariate data: a feature selection approach. *BMC Med Inform. Decis. Mak*. 2022;22:1-12.
32. Weiss AJ, Jiang HJ. Most frequent reasons for emergency department visits, 2018: Statistical Brief #286. 2021. Available Online: Accessed August 22, 2022. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb286-ED-Frequent-Conditions-2018.pdf
33. Wang HY, Chew G, Kung C, et al. The use of Charlson comorbidity index for patients revisiting the emergency department within 72 hours. *Chang Gung Med J*. 2007;30:437.
34. Hong WS, Haimovich AD, Taylor RA. Predicting 72-hour and 9-day return to the emergency department using machine learning. *JAMIA Open*. 2019;2:346-352.
35. Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice. OTexts*. 2018.
36. Gul M, Celik E. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Syst*. 2020;9:263-284.

Images in Emergency Medicine

Annals of Emergency Medicine (www.annemergmed.com) features a collection of hundreds of emergency medicine-related images, complete with brief discussion and diagnosis, in 23 categories. Select "Images" from the online menu or scan the QR code below and test your diagnostic skills today.