BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Forecasting emergency department visits in the city of Milan to predict high demand: a 2-day warning system

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-056017
Article Type:	Original research
Date Submitted by the Author:	02-Aug-2021
Complete List of Authors:	Murtas, Rossella; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Tunesi, Sara; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit; Andreano, Anita; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Russo, Antonio ; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit
Keywords:	ACCIDENT & EMERGENCY MEDICINE, EPIDEMIOLOGY, QUALITATIVE RESEARCH, STATISTICS & RESEARCH METHODS

Original Paper

Forecasting emergency department visits in the city of Milan to predict high demand: a 2-day warning system

Author List: Rossella Murtas, PhD[1], Sara Tunesi, PhD[1], Anita Andreano, MD[1], Antonio G. Russo, MD PhD[1]

Author Affiliations: Epidemiology Unit, Agency for Health Protection of Milan, 20123 Milan, Italy

Corresponding Author: Antonio Giampiero Russo, Via Conca del Naviglio 45, 20122 Milan, Italy, agrusso@ats-milano.it, +390285782111, +390285782128

ABSTRACT

Objectives The emergency department (ED) is one of the most critical areas in any hospitals. Recently, many countries have seen a rise in the number of ED visits, with an increase in length of stay and a detrimental effect on quality of care. Being able to forecast future demands would be a valuable support for hospitals to prevent high demands, in particular in a system with limited resources where inappropriate accesses are still a problem.

Design Time series cohort study.

Setting We collected all ED visits between January 2014 and December 2019 in the five major hospitals located in Milan. To predict daily volumes, we used a regression model with ARIMA errors. Predictors included were weekly and yearly periodicity, meteorological and environmental variables, information on influenza epidemics and festivities. Accuracy of prediction has been evaluated with the Mean Absolute Percentage Error (MAPE).

Primary outcome measures: Daily all-cause visits in EDs.

Results In the study period, we observed 2,223,479 visits. Children tended to visit ED most likely on weekends while adults and senior people on Mondays. The results confirmed the role of meteorological and environmental variables and the presence of yearly and weekly patterns. We found high correlation between observed and predicted values with a MAPE globally smaller than 8.1%.

Conclusions Results were used to establish an ED warning system based on past observations and indicators of high demands. This is important in every health system that usually deals with scarcity of resources, and it is crucial in a system where inappropriate emergency admissions are still high.

Strengths and limitations of this study

- This study is one of the few studies linking temporal periodicity, occurrence of festivities, local weather conditions and pollution together to ED visits
- Here we estimated an ARIMA model for each hospital, thus taking into consideration each specific characteristics and incorporating weekly and annual seasonality with Fourier terms
- Results were used to establish an ED warning system based on past observations and indicators of high demands
- We cannot exclude the possible presence of unmeasured variables that may better predict ED visits and overcrowding
- This study was intended to estimate ED demands and does not include information on staff rosters, two mechanisms unavoidably linked in every emergency department but that should be described separately

INTRODUCTION

The emergency department (ED) represents the gateway (an open door) and the most critical area of the Hospital, moving the different activities and determining the management problems of the elective activities, when increases the number of patients who knocking. In the last decade, many countries have seen a substantial rise in the number of ED visits, with an increase in length of stay,[1] and associated detrimental effects on quality of care. Accesses to ED are unavoidably subjected to some fluctuations and several models to predict high demands have been developed in the last decade, aiming at effectively managing hospital beds and staff rosters.[2] In Italy, even if the number of ED visits has been decreasing since 2016, the mean waiting time in EDs was high with 3.5% of accesses in 2017 between 12 and 24h, and 2.1% over 24h.[3] There are several factors that may lead to ED overcrowding, including high demand and inappropriate accesses. In Italy in 2017, only 23% of ED visits were at a high level of emergency (i.e. classified as red or yellow at triage) while 13% resulted to be inappropriate, and could have instead been managed by general practitioners.[3]

Several factors potentially affect the daily number of ED visits. Among these: annual[4], seasonal,[5] and weekly[4] periodicity as well as occurrence of festivities.[4] The effect of local weather conditions and pollution on ED visits volumes is still in debate: while some studies confirmed a significant association with temperature,[6] precipitation,[10] humidity,[12] and weather conditions,[13] other authors found these variables to be only mediocre predictors of the number of ED visits,[5] and founding air pollution mostly impacting cardiac and respiratory diseases.[12] An additional factor that has been studied in relation with ED visits volumes is the flu, with around 7% of total accesses attributable to Influenza-like Illness (ILI) during the epidemic season.[14] To our knowledge, there is no study linking all this information together to ED visits.

The present study aims to develop a model for forecasting ED arrivals, using regression-based time series analysis with Auto-regressive integrated moving average (ARIMA) errors, accounting simultaneously for the effect of meteorological and environmental variables, as well as information on flu epidemics and festivities, on the number of ED visits in the city of Milan. The model is used to establish an innovative ED warning system providing a planning instrument for hospitals, based on past observations and indicators of high demands.

METHODS

Study Design

This is a retrospective study conducted in the territory of the Milan Agency for Health Protection (AHP) using current health care databases of the emergency department admissions aggregated at hospital level. No individual level data were used, and patients cannot be identified from aggregated data which do not contain low counts (i.e. cells with ≤ 5 counts). For this reason, and according to the Italian legislation, this study was not submitted for ethics approval.[15]

Study setting and population

We collected all ED visits between the 1th of January 2014 and the 31th of December 2019 in the five major hospitals located in the city of Milan (figure 1). The five ED represent 49% of the total emergency rooms access of the city of Milan, which has a total of 17 ED.

Study protocol

Aggregated data on daily ED visits volumes, by age and gender, were extracted from the regional health database. Meteorological and environmental information was extracted from the Regional Environmental Protection Agency (ARPA).[16] Daily mean temperature, relative humidity (RH), cumulative precipitation, Nitrogen dioxide (NO2), and Particulate Matter with a diameter ≤ 10 μm (PM10) were collected from 2 monitoring stations (one measuring meteorological indicators and one measuring air pollution) located in the centre of Milan (figure 1). Missing values on a specific day were imputed with the average of all measurements of that exposure for that day collected from the other monitoring stations, weighted by the ratio of the yearly average of that monitoring station over the yearly average of the other monitoring stations, for the same environmental exposure.[17] Weekly data on ILI notifications were taken from the National Health Service Sentinel System (InfluNet).[18] Weekly incidence rates of ILI were expressed as the number of cases per 1,000 inhabitants per week. All available information has been linked to daily ED visits volumes on each of the five hospitals included in the study. Datasets were divided into training (from the 1th of January 2014 to the 31th of December 2018) and validation sets (from the 1th of January 2019 to the 31th of December 2019). For each hospital, we first estimated model parameters on the training dataset and evaluated post sample accuracy in the validation set. Multicollinearity was evaluated calculating Pearson pairwise correlation between variables and variance inflation criterion (VIF)[19].

Patient and public involvement

Patients were not involved in this research.

Data analyses

Development of the predictive model

To predict the daily volume of visits in each ED, we used a Time Series approach consisting in a regression model with ARIMA errors.[20] The statistical units were the days, 1,826 days in the training set and 365 in the validation set. This model is capable to combine two powerful statistical methods: the linear regression and the ARIMA. Linear regression of Y on X is usually described by the equation $Y_t = \alpha + \beta x_t + \epsilon_t$, where Y_t and x_t are the values of Y and X at day t, α and β are the intercept and the slope of the regression line, and ϵ_t is the error of the model at day t (the deviations from the fitted line to the observed values) assumed to be independent from other values. The ARIMA model deals with auto-correlation between errors through two components: the auto-regressive and the moving average process. The auto-regressive component assumes that previous observations are good predictors for future values, while the moving average component permits the model to update the predictions if the level of a constant time series changes. ARIMA specification is described by 3 parameters (p, d, q), where p is the order of auto-regression (AR) that is the number of time lags, d is the degree of differencing (the number of times the data have had past values subtracted to make the time series stationary), and q is the order of the moving average process (MA). For each hospital, these parameters were identified examining total and partial autocorrelation function (ACF and PACF respectively), as well as the statistical significance (p-value<0.05), and minimal Akaike Information Criteria (AIC). Weekly and annual seasonality were controlled for by including Fourier terms, a series of sine-cosine functions able to approximate periodicity.[8,20] For each seasonal period, the number of Fourier terms was chosen to minimise the AIC. Therefore, meteorological and environmental variables, as well as information on flu epidemics and festivities, were retained in the final model only if statistically significant. Diagnostics of the finally selected models were the Jarque-Bera test of normality, and correlation among the residuals according to the Portmanteau test. Variables and tests were considered statistically significant if p-value was < 0.05.

Forecasting Accuracy

Predicted values on validation sets were estimated using one-step forecast.[20] We estimated parameters only on training sets. However, we calculated forecasts on validation sets using all of the data preceding each observation. The accuracy of predictions was evaluated with the Mean Absolute Percentage Error (MAPE), which expresses, as percentages, a unit-free measure of performance:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} * 100$$

with y_t and $\hat{y}_t$ respectively the observed and the predicted number of visits at day t, and n the number of days in the validation set (n=365 in this study).

High demand definition

We proposed a definition of ED high demand as those days where the number of visits exceeded the median of the preceding 31 days. The days were defined as green (level 1) if the number of visits exceeded the median by less than 5%, yellow (level 2) if between 5% and 10%, red (level 3) if higher or equal than 10%. High demands were calculated on the observed and on the predicted ED visits in validation sets, we thus calculated the proportion of observed ED high demand that are correctly classified by predicted ED high demand (called sensitivity or recall metrics for multiclass classification problems).[21] In addition, we calculated the accuracy of predictions as the number of correct classifications over the total number of observations. All the statistical analyses were performed with R (version 3.6.3).[22]

ED warning system

In the month of January 2020, we established an ED warning system (WS), which was currently used by the selected hospitals in Milan as a planning instrument for EDs and consists in a transmission of daily reports. This WS continued until February when the COVID-19 outbreak started in Italy. A hypothetical daily report received from a hospital on the 5th of January 2020 can be found in figure 2. The report included forecasts of the number of visits for the following two days, with 95% margin errors and the indicator of high demand (green, yellow or red). The forecasts were made incorporating in the model past meteorological and environmental information via an Application Programming Interface (API), and forecast meteorological and environmental information provided daily by ARPA Lombardia. Previous week information on ILI was downloaded every week from

InfluNet, and included in the predictive models. Daily reports were constructed and dispatched automatically using R and R Markdown. During the WS campaign, we established a monitoring service capable of estimating daily sensitivity, accuracy of predictions and MAPE separately for the prediction at one and two days.

All analyses were performed with R software (V.4.0.2; R Core Team, Vienna, Austria), models and Fourier terms were estimated respectively, using the Arima and the Fourier functions in the R package forecast[23] using the parameter xreg for covariate specification. VIF were calculated using the VIF function in the package car.[24]

RESULTS

ED visits volumes

Between the 1th of January 2014 and the 31th of December 2019 (training set of 1,826 and validation set of 365 days) we observed 2,223,479 visits, 370,633 on average every year. Daily mean number of visits by hospital, temporal, climatic and patient characteristics in the training sets are summarized in table 1. Description of training and validation sets are summarized in the Supplementary Material. The Pearson correlation between predictors varied from weak (absolute correlation<0.3) to moderate (absolute correlation between 0.3 and 0.7), with a maximum of -0.67 between temperature and ILI and 0.61 between NO2 and PM10. VIF were smaller than 5 for all variables, with a maximum of 2.8 for temperature and 1.9 for ILI. We therefore included all the variables in the models, selecting the final model according to the statistical significance of predictors and minimal AIC.

Table 1

Total number of visits and mean number of daily visits by hospital, temporal and climatic factors, and patient characteristics between the 1th of January 2014 and the 31th of December 2019 in five Emergency Department of the City of Milan, Italy.

	N (%)[a]	Mean (Min-Max)[b]		N (%)[c]	Mean (Min-Max)[d]
Hospitals			**Cumulative Precipitation (mm)**		
A	421741 (19)	192 (107-301)	**≤ 0.6**	1678953 (75.5)	1018 (563-1295)
B	457021 (20.6)	209 (65-302)	**0.6+**	544526 (24.5)	1005 (627-1392)
C	272308 (12.2)	124 (61-197)	**NO2 (µg/m³)**		
D	530519 (23.9)	242 (125-337)	**≤ 32**	564957 (25.4)	974 (563-1295)
E	541890 (24.4)	247 (133-346)	**32-44**	570284 (25.6)	1022 (723-1292)
Total	2223479	1015 (563-1392)	**44-57**	536484 (24.1)	1032 (698-1392)
Gender			**57+**	551754 (24.8)	1035 (693-1272)
F	1113405 (50.6)	508 (277-782)	**PM10 (µg/m³)**		
M	1087903 (49.4)	497 (277-661)	**≤ 20**	571339 (25.7)	990 (563-1261)
Age			**20-29**	575530 (25.9)	1017 (688-1295)
≤ 14	360600 (16.4)	165 (55-443)	**29-44**	544147 (24.5)	1023 (710-1392)
14-65	1307139 (59.4)	597 (317-860)	**44+**	532463 (23.9)	1032 (693-1272)
65+	533569 (24.2)	244 (141-385)	**ILI (n. of weekly new cases per 1,000 inhabitants)**		
	N (%)[c]	**Mean (Min-Max)[d]**			
Temperature (°C)			**≤ 1.2**	1310096 (58.9)	1001 (563-1295)
≤ 9.2	564744 (25.4)	1021 (693-1392)	**1.2-2.5**	303072 (13.6)	1031 (799-1256)
9.2-15.6	563764 (25.4)	1033 (813-1261)	**2.5-5.6**	303102 (13.6)	1031 (698-1261)
15.6-22.3	563757 (25.4)	1025 (656-1295)	**5.6+**	307209 (13.8)	1045 (693-1392)
22.3+	531214 (23.9)	980 (563-1292)	**Day-before-after festivity**		
Relative Humidity (%)			**No**	2096838 (94.3)	1012 (563-1392)
≤ 50	560870 (25.2)	1018 (563-1295)	**Yes**	126641 (5.7)	1055 (688-1295)
50-62	552865 (24.9)	1009 (637-1292)	**Festivity**		

62-76	554041 (24.9)	1017 (627-1392)	**No**	2144726 (96.5)	1018 (677-1392)
76+	555703 (25)	1016 (786-1278)	**Yes**	78753 (3.5)	938 (563-1253)

ILI=Influenza-like illness

[a]Total number of visits by hospital, gender and age. In parenthesis the percentage of the number of visits of the total (2,223,479 total number of visits, 2,201,308 with information on age and gender);
[b]Mean, minimum and maximum number of daily visits by hospital, gender and age;
[c]Total number of visits by temporal and climatic factors (i.e. total number of visits in days with a particular value of temperature, Humidity etc.). In parenthesis the percentage of the number of visits of the total (2,223,479 total number of visits);
[d]Mean, minimum and maximum number of daily visits by temporal and climatic factors (i.e. mean number of daily visits in the days with a particular value of temperature, Humidity etc.).

Model specification and ARIMA results

All models showed a very strong weekly and yearly pattern, according to ACF and PACF plots. To normalize residuals, outliers (in the training sets only) were replaced by the mean of the observations of the same day in the other years, consequently all models showed residual normally distributed according to the Jarque-Bera test. All models showed a lack of fitting on New Year's Eve and/or August 15th, for this reason we chose to define specific dichotomous variable ("1" for the peculiar festivity, "0" for the other days) capable of detecting this extra variation. Table 2 displays the ARIMA parameters fitted for each model, and the number of Fourier terms that minimized AIC. All models were not stationary in mean and needed one differencing to make the time series stationary (d=1). ARIMA parameters and Fourier terms were different across hospitals, showing that each time series needed different model specification. Table 2 also displays, for each hospital, the factors that significantly influenced the number of ED visits, and that were included in the models. High temperatures were always associated with a statistically significant increase in ED visits volumes with a maximum increase of 1.84 daily visits every 1°C increase (the hospital E, s.e. 0.18). Relative Humidity was significantly associated with a limited decrease of total ED visits (-0.08, s.e. 0.04) for a 1% increment of RH only at the hospital D. High levels of cumulative precipitation were associated (except for the hospital C) with a statistically significant decrease in ED visits, with a maximum decrease of 0.31 daily visits every 1 mm of precipitation (the hospital E, s.e. 0.06). Concerning air pollution, we found an opposite effect of NO2 and PM10 on ED visits, with a mild significant negative effect for NO2 in two hospitals (-0.08 and -0.09) and an even milder positive association with PM10 in one (0.03). Except for the hospital C, the effect of ILI was always associated with the number of ED visits, showing an increase of daily visits between 0.73 and 1.74 (s.e. 0.29 and 0.41 respectively) at every unit increase in weekly ILI rates. Festivities were associated with a decrease of ED visits between 13 and 28 (s.e. 1.45 and 1.98) while special festivities were associated with the greatest decrease of at least 42 ED visits (s.e. 4.94).

Table 2

Auto-regressive integrated moving average (ARIMA) specifications and covariates effects on the number of ED accesses between the 1^{th} of January 2014 and the 31^{th} of December 2018 (training sets).

		Hospitals				
		A	B	C	D	E
Model specification	ARIMA parameters (p,d,q)	(0,1,2)	(1,1,1)	(1,1,2)	(1,1,1)	(1,1,1)
	Fourier terms†	3,13	3,14	3,13	3,16	3,15
Covariates Effects1	Temperature (°C)	1.29(0.15)	1.23(0.14)	0.68(0.11)	1.16(0.18)	1.84(0.18)
	Humidity (%)				-0.08(0.04)	
	Precipitation (mm)	-0.2(0.05)	-0.12(0.05)		-0.13(0.07)	-0.31(0.06)
	NO2 (μg/m3)	-0.08(0.03)			-0.09(0.04)	
	PM10 (μg/m3)			0.03(0.02)		
	ILI (weekly new cases per 1,000 inhabitants)	1.74(0.41)	1.05(0.37)	0.73(0.29)		0.97(0.46)
	Festivity		-28.23(1.98)	-12.96(1.45)	-25.42(2.23)	-14.56(2.39)
	Special Festivity*	-43.16(6.31)	-57.64(6.36); -62.61(6.29)	-42.06(4.92)	-59.86(7.58)	-63.24(7.92)
	Day-before-after festivity	7.14(1.5)	9.06(1.58)	3.75(1.22)		13.89(1.96)

1 Parameter estimates and standard errors in parentheses. Predictors were retained in the final model only if statistically significant (p-value<0.05)
† Number of sine and cosine terms used to approximate weekly and yearly periodicity
***New Year's Eve for hospitals A, C-D and New Year's Eve and August 15th for the hospital B**
ARIMA =Auto-regressive integrated moving average
ILI=Influenza-Like-Illness

Forecasting Accuracy and High demand definition

The accuracy of predictions (MAPE) in the validation sets, sensitivity and accuracy between observed and predicted ED high demand are displayed in table 3. Model performance was good with small MAPEs in validation sets, going from a minimum of 5.5% for the hospital D to a maximum of 8.1% for the hospital C. The models showed high sensitivity on days with green high demand, almost 90% of days with green predicted high demand were confirmed from observed values. On days with yellow high demand, sensitivity between predicted and observed was scarce ranging from 0.04 for Hospital B to 0.28 for the hospital A. Sensitivity of red high demand varied between hospitals with a minimum of 0.25 for the hospital A to a maximum of 0.57 for the hospital D. From Table 3 we can suggest that, for each hospital, at least 54% of the observed red high demands were classified as being, from predictions, at least yellow. Accuracy was high with at least 67% of the days with exactly the same predicted and observed high demand level (green, yellow or red).

Table 3

Indicators of performance of the developed models: Accuracy of predictions (MAPE) in the validation sets, and accuracy and sensitivity of high demand classification.

	MAPE	Accuracy (%)	Observed ED high demand	Predicted ED high demand (%, Sensitivity)		
				Green	Yellow	Red
Hospital A	5.9	72	Green	93	6	1
			Yellow	64	28	8
			Red	46	29	25
Hospital B	5.7	72	Green	92	8	0
			Yellow	85	4	11
			Red	35	15	50
Hospital C	8.1	67	Green	88	8	4
			Yellow	78	10	12
			Red	45	20	35
Hospital D	5.5	76	Green	91	6	3
			Yellow	65	27	8
			Red	35	9	56
Hospital E	6.1	74	Green	90	8	2
			Yellow	59	24	17
			Red	34	28	38

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

ED warning system

The accuracy of predictions (MAPE), sensitivity and accuracy between observed and predicted ED high demand in January (the operating period of the WS) at one and two days are displayed in table 4a and 4b. Errors of prediction (MAPE) were slightly higher than in the validation set with MAPE for one day always smaller than MAPE at two days. Accuracy between observed and predicted ED high demand was never smaller than 0.45 and generally smaller than in the validation set.

Table 4

The accuracy of predictions (MAPE), sensitivity and accuracy between observed and predicted ED high demand in January 2020 (the operating period of the WS) at one (4a) and two days (4b).

4a

	MAPE	Accuracy (%)	Observed ED high demand	Predicted ED high demand (%, Sensitivity)		
				Green	Yellow	Red
Hospital A	7.8	52	Green	94	6	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	7.8	81	Green	87	13	0
			Yellow	0	100	0
			Red	17	17	67
Hospital C	8.6	52	Green	100	0	0
			Yellow	67	33	0
			Red	73	27	0
Hospital D	6.6	45	Green	55	36	9
			Yellow	0	33	67
			Red	50	33	17
Hospital E	11	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

4b

	MAPE	Accuracy (%)	Observed ED high demand	Predicted ED high demand (%, Sensitivity)		
				Green	Yellow	Red
Hospital A	8.1	55	Green	100	0	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	8.6	71	Green	73	27	0
			Yellow	0	100	0
			Red	25	17	58
Hospital C	9	45	Green	93	7	0
			Yellow	83	17	0
			Red	82	18	0
Hospital D	7.6	48	Green	50	18	32
			Yellow	0	0	100
			Red	33	0	67
Hospital E	11.2	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

DISCUSSION

In this work we proposed and implemented in daily practice, a system to predict the number of ED visits in five hospitals of the city of Milan. The system is based on regression models with ARIMA errors, where ARIMA parameters were allowed to vary between hospitals, according to their specific characteristics, and it provides daily reports on the number of visits of the two subsequent days to the five hospitals participating in the study. The models showed a good overall performance with the MAPEs always smaller than 8.1%. Marcilio and colleagues[8] forecasted daily ED visits with Generalized Linear Models, finding MAPEs between 5.4% and 11.5%, according to different forecasting horizon and controlling for temperature effect.

Jones and colleagues,[5] using similar models, found MAPEs that varied between 8.5% and 15.5%. Although the number of predicted ED visits was close to the observed values, and there was good sensitivity in predicting mild high demand (green), there was moderate sensitivity in predicting the spike of ED visits volumes (red-high demand) for some hospitals and acceptable sensitivity for the hospital D. However, we found good sensitivity in classifying observed red high demands as being at least yellow from predictions, and accuracy among observed and predicted high demand levels always close to 70%. The definition of ED high demand is not straightforward as it relies on the specific hospital's characteristics. It is a natural cause of ED overcrowding, which is the most problematic issue in EDs, thus deserving the effort in trying to predict it. In Italy, the Ministry of Health suggested to define high demand as those days where the number of visits exceeded the 91° percentile of the preceding year.[25] In this study, we proposed a definition based on percentage increases compared to the median of the preceding month, to advice ED departments of requests increasing over what they have managed in the preceding month.

During the operating period of the warning system, January 2020, we found a worse adaptation of the models than in the validation year 2019. This could be due to the ongoing outbreak of COVID-19 were ED visits for non-critical problems were discouraged.[26]

Concerning potential predictors, we found a strong weekly and yearly pattern, adequately captured by the terms used to approximate periodicity (Fourier terms). There were statistically significant effects of meteorological factors on ED visits. The temperature was always positively associated with the outcome, with an increase in the number of visits for each 1-degree increase in temperature across hospitals, in accordance with previous results.[10,11,13] As reported in another study,[27] high temperatures are associated with ED visits, especially for most susceptible population, as persons with diabetes or cancer, so is important for public health officials to implement adaptation measures to manage the impact of high temperatures on population health. The role of precipitations has not yet been well established. To our knowledge only one study measured an indirect effect in reducing ED visits volumes.[11] In accordance with these results, rainy days were found to be mild associated with reduced numbers of ED visits. NO2 and PM10 had a mild significant effect only in two hospitals and in one hospital respectively, and were discordant, with a negative effect of NO2 and a positive effect of PM10 on the number of ED visits. This may be explained considering that the effect of pollution on ED visits is generally exerted and measured on respiratory, especially asthma, and/or cardiac rather than with total accesses and it may be diluted when analysing all accesses. Only a few studies found a positive association of Total Suspended Particles with all accesses, but trauma, going in the same direction of the small significant increase of the number of visits related to PM10 we found.[28] ILI were found to significantly increase the number of ED visits, as found by other researchers.[29]

This study indicated a moderate to good sensitivity in predicting high demands, showing some difficulties in anticipating the exact days of red days. In the future we aim to investigate models capable to predict directly the ED peaks instead of predicting the number of ED visits, using time series based on the number of ED visits and possibly ambulance diversion status. Finally, when interpreting these results, it is necessary to be aware of the possible multicollinearity problem between variables, which may alter the magnitude and the statistical significance of coefficients. However, according to Vatcheva 2016,[19] only high correlations between variables would result in a change of sign of the coefficients and furthermore VIFs were always smaller than 5. Correlated factors were the pollution variables (NO2 and PM10), which were never considered in the same model together. Given that the highest correlation was found among temperature and ILI, the effect of these variables on the number ED visits may potentially be biased due to multicollinearity. However, we included both terms in the models given the fact that they have an effect on ED visits independently from one another.

Another limitation is the choice of the hospitals considered for this work that are the major hospitals located in the city of Milan. This methodology might not be the feasible for small hospitals as they might suffer for low counts or even no visits at all in particular days. This can be done by implementing different statistical models, for example, negative binomial or zero-inflated Poisson models, and would be one of our aims in the next years. In conclusion, we proposed a hospital specific ED warning system based on predictive models developed on previous attendances that can be used as a planning instrument in hospitals to increase resources, and to prevent patient high demand when a higher number of attendances is expected. This is important in every health system that usually deals with scarcity of resources, and it is crucial in a system where inappropriate emergency admissions are still high.

List of abbreviations

ACF: autocorrelation function

AHP: Milan agency for health protection

AIC: minimal Akaike information criteria

API: application programming interface

AR: auto-regression

ARIMA: Auto-regressive integrated moving average

ED: emergency department

ILI: Influenza-like-illness

MA: moving average process

MAPE: mean absolute percentage error

NO2: nitrogen dioxide

PM10: particulate matter with a diameter ≤ 10 µm

PACF: partial autocorrelation function

RH: relative humidity

WS: warning system

Ethics approval and consent to participate This study does not involve human participants or animal subjects. Ethics approval and consent to participate were not required, as this is an observational study based on data routinely collected by the Agency for Health Protection (ATS) of Milan, a public body of the Regional Health Service-Lombardy Region. The ATS has among its institutional functions, established by the Lombardy Region legislation (R.L. 23/2015), the government of the care pathway at the individual level in the regional social and healthcare system, the evaluation of the services provided to, and the outcomes of, patients residing in the covered area. This study is also ethically compliant with the National Law (D.Lgs. 101/2018) and the "General Authorisation to Process Personal Data for Scientific Research Purposes" (n.8 and 9/2016, referred to in the Data Protection Authority action of 13/12/2018). Data were anonymized with a unique identifier in the different datasets before being used for the analyses.

Contributors RM and AGR conceptualised the study and defined the methodology, accessed and verified the data. RM analysed the data set, ST and AA contributed to the literature search, data interpretation, and writing of the manuscript. ST, AA, have made substantial contributions to the revision of the paper. AGR supervised and administered the project.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Data availability statement Data are not publicly available due to property restrictions.

Word count 3396

REFERENCES

1 Pines JM, Hilton JA, Weber EJ, *et al.* International perspectives on emergency department crowding. *Acad Emerg Med* 2011; **18**: 1358–70.

2 Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. *Emergency medicine journal : EMJ* 2009. DOI:10.1136/emj.2008.062380.

3 Programma Nazionale Esiti - PNE 2018 - Home. https://pne.agenas.it/index.php (accessed Feb 18, 2020).

4 Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med* 2010; **17**: 970–8.

5 Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008; **15**: 159–70.

6 Duwalage KI, Burkett E, White G, Wong A, Thompson MH. Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas* 2020; published online Feb 17. DOI:10.1111/1742-6723.13481.

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**: 237–9.

8 Marcilio I, Hajat S, Gouveia N. Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine* 2013; **20**: 769–77.

9 McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med* 2008; **15**: 337–46.

10 Rising WR, O'Daniel JA, Roberts CS. Correlating Weather and Trauma Admissions at a Level I Trauma Center. *Journal of Trauma and Acute Care Surgery* 2006; **60**: 1096–1100.

11 Diehl AK, Morris MD, Mannis SA. Use of calendar and weather data to predict walk-in attendance. *South Med J* 1981; **74**: 709–12.

12 Díaz J, Alberdi JC, Pajares MS, *et al.* A model for forecasting emergency hospital admissions: effect of environmental variables. *J Environ Health* 2001; **64**: 9–15.

13 Macgregor DM. Effect of weather on attendance with injury at a paediatric emergency department. *Emergency Medicine Journal* 2003; **20**: 204–5.

14 Wargon M, Brun-Ney D, Beaujouan L, Casalino E. No more winter crisis? Forecasting daily bed requirements for emergency department admissions to hospital. *Eur J Emerg Med* 2018; **25**: 250–6.

15 Italia. Provvedimento del Garante n. 2 del 16 giugno 2004. Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici. .

16 Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. ARPA Lombardia. https://www.arpalombardia.it:443/Pages/ARPA_Home_Page.aspx (accessed March 10, 2020).

17 Faustini A, Stafoggia M, Berti G, *et al.* The relationship between ambient particulate matter and respiratory mortality: a multi-city study in Italy. *Eur Respir J* 2011; **38**: 538–47.

18 Ministero della Salute. Influenza. http://www.salute.gov.it/portale/influenza/homeInfluenza.jsp (accessed March 10, 2020).

19 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* 2016; **6**. DOI:10.4172/2161-1165.1000227.

20 Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018.

21 M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *IJDKP* 2015; **5**: 01–11.

22 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020 https://www.R-project.org/.

23 Hyndman [aut R, cre, cph, *et al.* forecast: Forecasting Functions for Time Series and Linear Models. 2020 https://CRAN.R-project.org/package=forecast (accessed Nov 14, 2020).

24 Fox J, Weisberg S. An R Companion to Applied Regression. SAGE Publications, 2018.

25 Salute M della. Pronto soccorso, via libera della Conferenza Stato-Regioni alle nuove linee di indirizzo. http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=3849 (accessed April 24, 2020).

26 Jeffery MM, D'Onofrio G, Paek H, *et al.* Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. *JAMA Intern Med* 2020; published online Aug 3. DOI:10.1001/jamainternmed.2020.3288.

27 Lavigne E, Gasparrini A, Wang X, *et al.* Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environ Health* 2014; **13**: 5.

28 Samet JM, Speizer FE, Bishop Y, Spengler JD, Ferris BG. The Relationship between Air Pollution and Emergency Room Visits in an Industrial Community. *Journal of the Air Pollution Control Association* 1981; **31**: 236–40.

29 Schanzer DL, Schwartz B. Impact of seasonal and pandemic influenza on emergency department visits, 2003-2010, Ontario, Canada. *Acad Emerg Med* 2013; **20**: 388–97.

Figure 1

Location of the five considered hospitals and of meteorological and air pollution monitoring station in the city of Milan.

Figure 2

Hypothetical daily report received from a hospital on the 5th of January 2020.

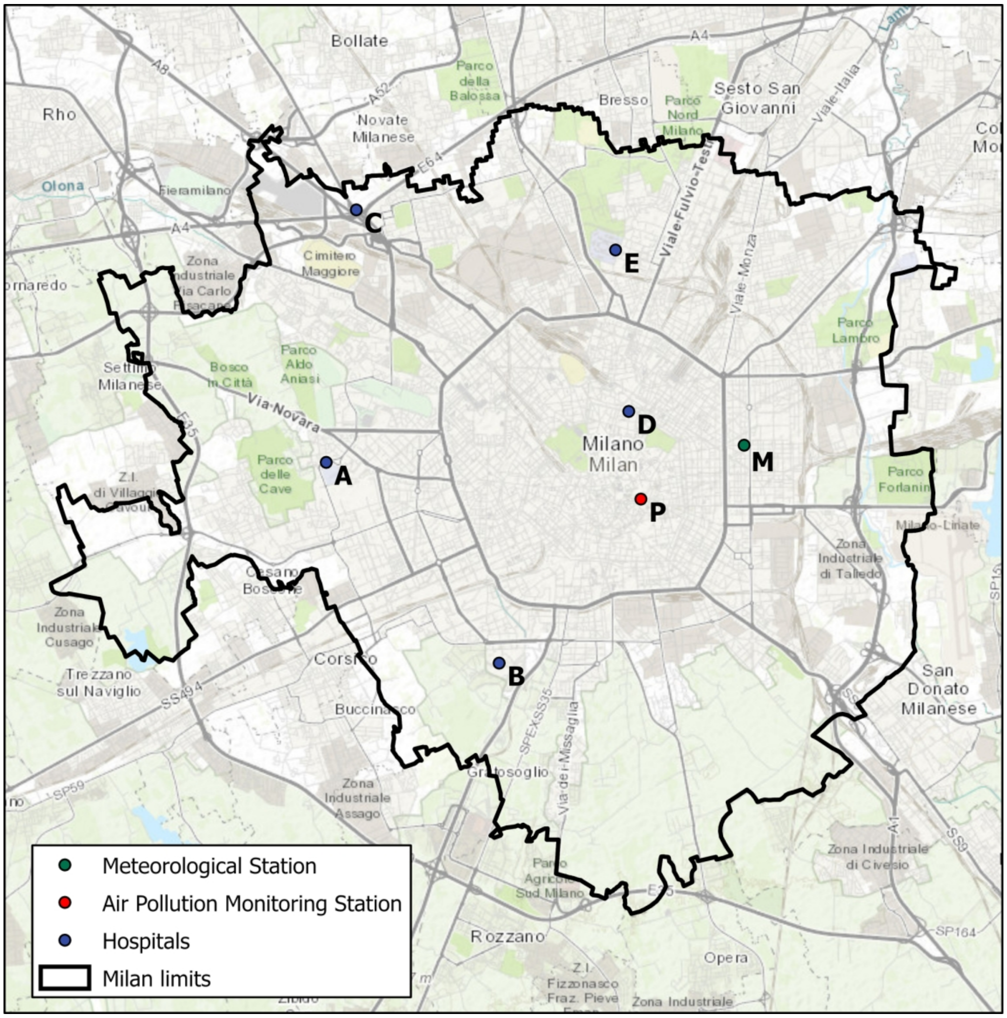


Figure 1. Location of the five considered hospitals and of meteorological and air pollution monitoring station in the city of Milan.

160x161mm (300 x 300 DPI)

EMERGENCY ADMISSION 2-DAY WARNING SYSTEM

Agency for Health Protection of Milan

Hospital A. Prediction for the day:

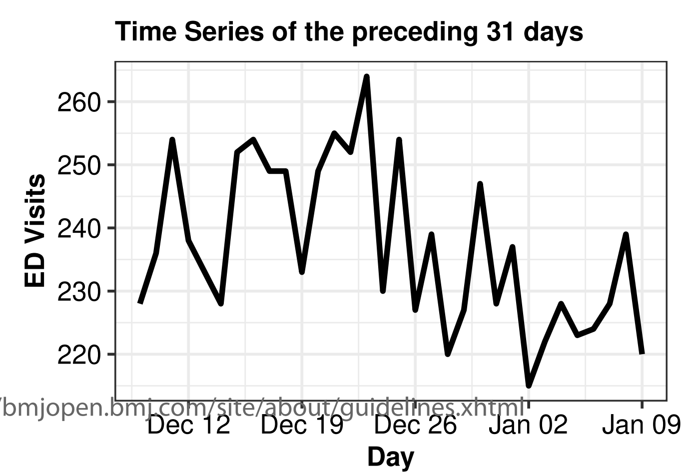
11 Jan 2020	12 Jan 2020
Level 1	**Level 2**

	11 Jan 2020	12 Jan 2020
Predicted ED visits $\pm$ 95% Margin Errors	227 ± 12	254 ± 14
Covariates		
NO2	$46\ \mu g/m^3$	$38.4\ \mu g/m^3$
PM10	$6\ \mu g/m^3$	$3.9\ \mu g/m^3$
TEMPERATURE	$5°$	$6°$
UMIDITY	65%	57%
PRECIPITATION	$0\ mm$	$0\ mm$
ILI RATE	10.87	10.87

Methodology

Level 1	Number of visits exceeded the median by less than 5%.
Level 2	Number of visits exceeded the median between 5% and 10%.
Level 3	Number of visits exceeded the median by more than 10%.
Prediction	Prediction based on a regression model with ARIMA errors (Hyndman 2018).
Environment	Meteo and Pollution Forecast from ARPA Lombardia.
ILI rates	Influence Like Illness rate, number of cases per 1,000 inhabitants per week (InfluNet).

Sistema Socio Sanitario
Regione Lombardia
ATS Milano
Città Metropolitana



Description of Training and validation sets

Patients were mostly female (50.6%), with a mean age of 44 years (standard deviation s.d. of 26 years). The mean number of daily visits was similar in genders. Higher ED accesses volumes were found among people aged over 65 years than the others ages. During the study period, mean temperature was 16°C, mean RH was 63%, and there was precipitation on 31% of the days. Mean level of NO2 exceeded the European limit (40 µg/m^3), with a mean of 46 µg/m^3, while mean PM10 was lower than the European limit, with a mean of 35 µg/m^3, during the observed period. On days-before and after festivities, we measured a higher number of visits, while on festivity days there was a lower number of visits compare to normal days. Training and validation sets were similar according to meteorological factors, but there were mild differences in air pollution and ILI rates (supplementary Table 1). The year 2019 was in fact characterized by significantly higher levels of pollution (t-test p-values<0.001) and lower ILI rates (t-test p-value<0.01) compared to the previous years. Patients were slightly younger in the training set than in the validation set, the mean age being 43 years in the former and 45 in the latter. The number of ED accesses was statistically different across age groups between days: children (0-14 years) tended to visit ED more likely on weekends (20% higher on Sundays compared to other days) while adults and senior people (15-65 and >65 years) on Mondays (14% and 11% higher than other days) (Anova test for mean differences p-values<0.001). August was the months with smaller ED accesses volumes, with a 14% decrease compared to the average of the other months.

Supplementary Table 1

Mean, standard deviation and t-test for mean difference between training and validation sets by covariates.

	Mean (s.d.)		t-test p-value
	Training	Validation	
Female (%)	50.6	50.4	-
Age (years)	43 (26)	45 (26)	<0.001
Temperature (°C)	15.7 (8)	16.4 (8)	0.1257
Relative Humidity (%)	63.4 (17)	62.3 (17)	0.2765
Cumulative Precipitation (mm)	2.3 (6.6)	2.2 (5.8)	0.7155
NO2 (µg/m3)	47 (19)	39 (17)	<0.001
PM10 (µg/m3)	36 (21)	30 (18)	<0.001
ILI (new cases per 1,000 inhabitants)	1.9 (3)	2.6 (3.8)	<0.001
S.d.=standard deviation ILI=Influenza-Like-Illness			

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	-
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2/3
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2/3
Bias	9	Describe any efforts to address potential sources of bias	3
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2/3
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	3/4
		(*b*) Describe any methods used to examine subgroups and interactions	-
		(*c*) Explain how missing data were addressed	3
		(*d*) If applicable, explain how loss to follow-up was addressed	-
		(*e*) Describe any sensitivity analyses	5
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	7
		(c) Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	Report numbers of outcome events or summary measures over time	7

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8/9/10
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	-
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

A time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: a 2-day warning system

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-056017.R1
Article Type:	Original research
Date Submitted by the Author:	14-Dec-2021
Complete List of Authors:	Murtas, Rossella; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Tunesi, Sara; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit; Andreano, Anita; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Russo, Antonio ; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	ACCIDENT & EMERGENCY MEDICINE, EPIDEMIOLOGY, QUALITATIVE RESEARCH, STATISTICS & RESEARCH METHODS

SCHOLARONE™ Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Original Paper

A time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: a 2-day warning system

Author List: Rossella Murtas, PhD[1], Sara Tunesi, PhD[1], Anita Andreano, MD[1], Antonio G. Russo, MD PhD[1]

Author Affiliations: Epidemiology Unit, Agency for Health Protection of Milan, 20123 Milan, Italy

Corresponding Author: Antonio Giampiero Russo, Via Conca del Naviglio 45, 20122 Milan, Italy, agrusso@ats-milano.it, +390285782111, +390285782128

ABSTRACT

Objectives The emergency department (ED) is one of the most critical areas in any hospital. Recently, many countries have seen a rise in the number of ED visits, with an increase in length of stay and a detrimental effect on quality of care. Being able to forecast future demands would be a valuable support for hospitals to prevent high demand, particularly in a system with limited resources where use of ED services for non-urgent visits is an important issue.

Design Time series cohort study.

Setting We collected all ED visits between January 2014 and December 2019 in the five larger hospitals in Milan. To predict daily volumes, we used a regression model with ARIMA errors. Predictors included were day of the week and year-round seasonality, meteorological and environmental variables, information on influenza epidemics and festivities. Accuracy of prediction was evaluated with the Mean Absolute Percentage Error (MAPE).

Primary outcome measures Daily all-cause EDs visits.

Results In the study period, we observed 2,223,479 visits. ED visits were most likely to occur on weekends for children and on Mondays for adults and seniors. Results confirmed the role of meteorological and environmental variables and the presence of day of the week and year-round seasonality effects. We found high correlation between observed and predicted values with a MAPE globally smaller than 8.1%.

Conclusions Results were used to establish an ED warning system based on past observations and indicators of high demand. This is important in any health system that regularly faces with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits is still high.

Strengths and limitations of this study

- This study is one of the few studies linking temporal periodicity, occurrence of festivities, local weather conditions, and pollution to ED visits
- We estimated an ARIMA model for each hospital, thus taking into consideration each specific characteristic and incorporating weekly and annual seasonality with Fourier terms
- Results were used to establish an ED warning system based on past observations and indicators of high demand
- We cannot exclude the possible presence of unmeasured variables that may better predict ED visits and overcrowding
- This study was intended to estimate ED demand and does not include information on staff rosters, two components unavoidably linked in any emergency department but that should be described separately

INTRODUCTION

The emergency department (ED) is the gateway (an open door) and the most critical area of a hospital, moving many activities and causing problems in the management of elective procedures when the number of patients who come knocking increases. In the last decade, many countries have seen a substantial rise in the number of ED visits, with an increase in length of stay,[1] and associated detrimental effects on quality of care. ED visits are unavoidably subject to fluctuation, and several models to predict high demand have been developed in the last decade, aiming at effectively managing hospital beds and staff rosters.[2] In Italy, even though the number of ED visits has been decreasing since 2016, the mean waiting time in EDs was high, between 12h and 24h in 3.5% of cases in 2017, and over 24 h in 2.1% of cases.[3] The definition of overcrowding in the ED literature is not consistent, nor are the measures used to assess overcrowding, which vary from clinician perception of overcrowding, to input measures (e.g., waiting times, number of patients arrived), throughput measures (e.g., ED capacities, patient care time), output measures (e.g. percentages of hospital admissions, hospital beds), or multidimensional indices such as the Emergency Department Work Index (EDWIN). This variety of measures corresponds to the different type of factors studied as causes of ED crowding. We concentrate here on predicting the number of visits from input factors, i.e., determinants and modalities of patient inflow, such as non-urgent visits and Influenza season. In this case it is better to speak of overflow (ref). We did not investigate throughput factors, describing organizational issues in the ED, such as inadequate staffing, nor output factors. The latter include one of the major reasons for ED overcrowding, which is the shortage of acute care bed capacity.[4–7] . Among the most investigated input factors are non-urgent visits, meaning "patients who could have been assessed and treated in other facilities that treat less urgent cases" (Howard e t al.). In Italy in 2017, only 23% of ED visits were classified as red or yellow at triage, while 13%[3] had a low level of priority, coded white triage in Italy. This use of emergency department services is a signal of lack of continuity of primary care and difficulty of access to both primary and specialist care. It is also not cost-effective and leads to an increase in waiting times in the EDs.[8,9]

Several factors potentially affect the daily number of ED visits. Among these: annual[10,11], seasonal,[12–15] and weekly[10–15] periodicity, as well as festivities.[10,12,16,17] The effect of local weather conditions and pollution on ED visit volumes is still in debated: while some studies confirmed a significant association with temperature,[11,13–15,18,19] precipitation,[13,15] humidity,[18] and weather conditions,[19] other authors found these variables to be only mediocre predictors of the number of ED visits,[12] and found air pollution mostly impacting cardiac and respiratory diseases.[18] An additional factor that has been studied in relation with ED visit volumes is the flu, with around 7% of total accesses attributable to Influenza-like Illness (ILI) during the epidemic season.[20] Murtas and colleagues[21] evaluated the hypothesis of the early presence of the COVID-19 epidemic in Italy by analysing data on trends of access to EDs using a Poisson regression model adjusted for seasonality and influenza outbreaks. In this work they found that predicting ED visits by considering both seasonality and ILI rates, compared to a model tacking into account only seasonality, notably increased the fitting of the model. Therefore, syndromic surveillance (such as ILI rates which in Italy are provided weekly by the National Health Service Sentinel System) may be able to provide early warning of hospital bed capacity strain caused by seasonal respiratory disease.[22] To our knowledge, there is no study linking all this information together to ED visits.

The present study aims to develop a model for forecasting ED arrivals, using regression-based time series analysis with Auto-regressive integrated moving average (ARIMA) errors, accounting simultaneously for the effect of meteorological and environmental variables, as well as information on flu epidemics and festivities, on the number of ED visits in the city of Milan. The model is used to establish an innovative ED warning system providing a planning instrument for hospitals, based on past observations and indicators of high demand.

METHODS

Study Design

This is a retrospective study conducted in the area served by the Milan Agency for Health Protection (AHP) using current health care databases of daily ED visits aggregated at hospital level. No individual level data were used, and patients cannot be identified from aggregated data which do not contain low counts (i.e. cells with ≤ 5 counts). For this reason, and in accordance with Italian legislation, this study was not submitted for ethics approval.[23]

Study setting and population

We collected all ED visits, including patients registered at triage that voluntarily left the ED premises before being evaluated by a physician, between the 1st of January 2014 and the 31st of December 2019 in the five largest hospitals located in the city of Milan (figure 1). All five hospitals are public hospitals and received 49% of all emergency room access of the city of Milan, which has a total of 17 EDs, with a mean number of daily ED visits during 2014-2019 ranging from 124 for hospital C to 247 for hospital E.

Study protocol

Aggregated data on daily ED visit volumes, by age and gender, were extracted from the regional health database. Meteorological and environmental information was extracted from the Regional Environmental Protection Agency (ARPA).[24] Daily mean temperature, relative humidity (RH), cumulative precipitation, nitrogen dioxide (NO2), and particulate matter with a diameter ≤ 10 μm (PM10) were collected from 2 monitoring stations (one measuring meteorological indicators and one measuring air pollution) located in the centre of Milan (figure 1). As sensitivity analysis we also investigated the effect of minimum, maximum and apparent temperature on daily ED visits.[25] Missing values on a specific day were imputed with the average of the measure in that specific year. Weekly data on ILI notifications were taken from the National Health Service Sentinel System (InfluNet).[26] Weekly incidence rates of ILI were expressed as the number of cases per 1,000 inhabitants per week. All available information was linked to daily ED visit volumes for each of the five hospitals included in the study. Datasets were divided into training (from the 1st of January 2014 to the 31st of December 2018) and validation sets (from the 1st of January 2019 to the 31st of December 2019). For each hospital, we first estimated model parameters on the training dataset and evaluated post-sample accuracy in the validation set. We included, in each model, only factors that significantly influenced the number of ED visits. Multicollinearity was evaluated calculating Pearson pairwise correlation between variables and variance inflation criterion (VIF)[27].

Patient and public involvement

Patients were not involved in this research.

Data analyses

Development of the predictive model

To predict the daily volume of visits in each ED, we used a time series approach consisting in a regression model with ARIMA errors.[28] The statistical units were days, 1,826 days in the training set and 365 in the validation set. This model is able to combine two powerful statistical methods: linear regression and ARIMA. Linear regression of Y on X is usually described by the equation $Y_t = \alpha + \beta x_t + \epsilon_t$, where Y_t and x_t are the values of Y and X at day t, α and β are the intercept and the slope of the regression line, and ϵ_t is the error of the model at day t (the deviations from the fitted line to the observed values) assumed to be independent from other values. The ARIMA model deals with auto-correlation between errors through two components: the auto-regressive and the moving average process. The auto-regressive component assumes that previous observations are good predictors for future values, while the moving average component allows the model to update the predictions if the level of a constant time series changes. ARIMA specification is described by 3 parameters (p, d, q), where p is the order of auto-regression (AR) that is the number of time lags, d is the degree of differencing (the number of times the data have had past values subtracted to make the time series stationary), and q is the order of the moving average process (MA). For each hospital, these parameters were identified examining total and partial autocorrelation function (ACF and PACF, respectively), as well as statistical significance (p-value<0.05), and minimal Akaike Information Criteria (AIC). Day of the week and year-round seasonality were controlled for by including Fourier terms, a series of sine-cosine functions capable of approximating periodicity.[16,28] The number of Fourier terms was chosen to minimise the AIC for each seasonal period (up to 7 for day of the week seasonality and up to 365 for year-round seasonality). Each seasonal component can be written in the model equation as

$$\sum_{j=1}^{n}\left[\alpha_j \sin\left(\frac{2\pi jt}{m}\right) + \beta_j \cos\left(\frac{2\pi jt}{m}\right)\right]$$

where n is the number of Fourier terms chosen to minimise the AIC (up to 7 for day of the week seasonality and up to 365 for year-round seasonality) and m is the seasonal period (7 for day of the week and 365 for year-round seasonality).

Therefore, meteorological and environmental variables, as well as information on flu epidemics and festivities, were retained in the final model only if statistically significant. As festivities, we considered Italian public holidays with school and office closures: New Year's Day, Epiphany, Easter Sunday and Monday, Italian Liberation Day, Labour Day, Foundation of the Italian Republic, assumption day, All Saints' Day, Saint Ambrose's Day (local patron saint), Feast of the Immaculate Conception, Christmas Day, Saint Stephen's day and New Year's Eve. In addition, we created dummies for specific festivities that were responsible for a significant variation in the number of ED visits: New Year's Eve and Assumption Day (August 15th). Diagnostics of the finally selected models were the Jarque-Bera test of normality, and correlation among the residuals according to the Ljung-Box test. Variables and tests were considered statistically significant if p-value was < 0.05.

The ARIMA model was compared with a simple regression model (M1) including only meteorological, environmental, and festivity covariates and with a generalized linear model (M2) also including the Fourier terms to control for seasonality. P-values were calculated by comparing the full model (ARIMA) to M1 and M2 using the likelihood ratio test.

Forecasting Accuracy

Predicted values on validation sets were estimated using one-step forecast.[28] We estimated parameters only on training sets. However, we calculated forecasts on validation sets using all of the data preceding each observation. The accuracy of predictions was evaluated with the Mean Absolute Percentage Error (MAPE), which expresses, as percentages, a unit-free measure of performance:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} * 100$$

with y_t and $\hat{y}_t$ respectively the observed and the predicted number of visits at day t, and n the number of days in the validation set (n=365 in this study).

High demand definition

We proposed a definition of high ED demand as days where the number of visits exceeded the median of the preceding 31 days. The days were defined as green (level 1) if the number of visits exceeded the median by less than 5%, yellow (level 2) if between 5% and 10%, red (level 3) if higher than or equal to 10%. High demand was calculated on the observed and predicted ED visits in validation sets, we thus calculated the proportion of observed high ED demand that is correctly classified by predicted high ED demand (called sensitivity or recall metrics for multiclass classification problems).[29] In addition, we calculated the accuracy of predictions as the number of correct classifications over the total number of observations. All statistical analyses were performed with R (version 3.6.3).[30]

To evaluate the proposed definition, we further calculated high demand as: the number of visits exceeding the median of the preceding 7, 14, and 21 days and the number of visits exceeding the mean of the preceding 7, 14, 21, and 31 days, defining green, yellow, and red levels of high demand as above. We chose 7, 14, and 21 lag days in order to adjust for weekly variation in the number of ED visits by design. We further calculated high demand as defined by the Lombardy Region[31]: when the number of visits exceeded the 91st percentile of the previous year time-series. Low demand days were defined as those with a number of visits smaller than 25th percentile, medium demand days as those with a number of visits between 25th percentile and 75th percentile, high demand days if between 75th percentile and 90th percentile, and finally very high demand days if over 91th percentile.

ED warning system

In the month of January 2020, we established an ED warning system (WS), which was used by the selected hospitals in Milan as a planning instrument for EDs and consists in a transmission of daily reports. This WS continued until February when the COVID-19 outbreak started in Italy. According to the model choices highlighted by the above methodology (validation and calibration of the model were performed with data from 2014 to 2019), parameters were updated weekly and used to establish the WS which operated in January 2020. A hypothetical daily report received from a hospital on the 5th of January 2020 can be found in figure 2. The report included forecasts of the number of visits for the following two days, with 95% margin errors and a high demand indicator (green, yellow or red). The forecasts were made incorporating in the model past meteorological and environmental information via an Application Programming Interface (API) where 2-day future forecasts of meteorological and environmental information were provided by ARPA Lombardia. Weekly information on ILI was downloaded every week from InfluNet, and included in the predictive models. Daily reports were constructed and dispatched automatically using R and R Markdown. During the WS campaign, we established a monitoring service capable of estimating daily sensitivity, accuracy of predictions and MAPE separately for prediction one and two days ahead.

All analyses were performed with R software (V.4.0.2; R Core Team, Vienna, Austria), models and Fourier terms were estimated respectively, using the Arima and the Fourier functions in the R package forecast[32] using the parameter xreg for covariate specification. VIF was calculated using the VIF function in the car package.[33]

RESULTS

ED visit volumes

Between the 1st of January 2014 and the 31st of December 2019 (training set of 1,826 and validation set of 365 days) we observed 2,223,479 visits, 370,633 on average every year. Daily mean number of visits by hospital, temporal, meteorological, and patient characteristics in the training sets are summarized in table 1. Missingness, over the whole period 2014-2019, in meteorological and environmental variables were found in 8 days for temperature, 7 days for precipitation, and 37 days for humidity. Description of training and validation sets, and plots of each hospital's time series are summarized in the Supplementary Material (Supplementary Table 1). The Pearson correlation between predictors varied from weak (absolute correlation<0.3) to moderate (absolute correlation between 0.3 and 0.7), with a maximum of -0.67 between temperature and ILI and 0.61 between NO2 and PM10. VIF was smaller than 5 for all variables, with a maximum of 2.8 for temperature and 1.9 for ILI. We therefore included all the variables in the models, selecting the final model according to the statistical significance of predictors and minimal AIC.

Table 1

Total number of visits and mean number of daily visits by hospital, temporal and meteorological factors, and patient characteristics between the 1st of January 2014 and the 31st of December 2019 in five emergency departments of the city of Milan, Italy.

	N (%)[a]	**Mean (Min-Max)[b]**		**N (%)[c]**	**Mean (Min-Max)[d]**
Hospitals			**Cumulative Precipitation (mm)**		
A	421741 (19)	192 (107-301)	**≤ 0.6**	1678953 (75.5)	1018 (563-1295)
B	457021 (20.6)	209 (65-302)	**0.7+**	544526 (24.5)	1005 (627-1392)
C	272308 (12.2)	124 (61-197)	**NO_2 ($\mu g/m^3$)**		
D	530519 (23.9)	242 (125-337)	**≤ 32**	564957 (25.4)	974 (563-1295)
E	541890 (24.4)	247 (133-346)	**33-44**	570284 (25.6)	1022 (723-1292)
Total	2223479	1015 (563-1392)	**45-57**	536484 (24.1)	1032 (698-1392)
Gender			**58+**	551754 (24.8)	1035 (693-1272)
F	1113405 (50.6)	508 (277-782)	**PM10 ($\mu g/m^3$)**		
M	1087903 (49.4)	497 (277-661)	**≤ 20**	571339 (25.7)	990 (563-1261)
Age			**21-29**	575530 (25.9)	1017 (688-1295)
≤ 14	360600 (16.4)	165 (55-443)	**30-44**	544147 (24.5)	1023 (710-1392)
15-65	1307139 (59.4)	597 (317-860)	**45+**	532463 (23.9)	1032 (693-1272)
66+	533569 (24.2)	244 (141-385)	**ILI (n. of weekly new cases per 1,000 inhabitants)**		
	N (%)[c]	**Mean (Min-Max)[d]**			
Temperature (°C)			**≤ 1.2**	1310096 (58.9)	1001 (563-1295)
≤ 9.2	564744 (25.4)	1021 (693-1392)	**1.3-2.5**	303072 (13.6)	1031 (799-1256)
9.3-15.6	563764 (25.4)	1033 (813-1261)	**2.6-5.6**	303102 (13.6)	1031 (698-1261)
15.7-22.3	563757 (25.4)	1025 (656-1295)	**5.7+**	307209 (13.8)	1045 (693-1392)
22.4+	531214 (23.9)	980 (563-1292)	**Day before/after festivity**		
Relative Humidity (%)			**No**	2096838 (94.3)	1012 (563-1392)
≤ 50	560870 (25.2)	1018 (563-1295)	**Yes**	126641 (5.7)	1055 (688-1295)
51-62	552865 (24.9)	1009 (637-1292)	**Festivity**		

63-76	554041 (24.9)	1017 (627-1392)	**No**	2144726 (96.5)	1018 (677-1392)
77+	555703 (25)	1016 (786-1278)	**Yes**	78753 (3.5)	938 (563-1253)
ILI=Influenza-like illness					

[a]Total number of visits by hospital, gender and age. In parenthesis the percentage of the number of visits out of the total (2,223,479 total number of visits, 2,201,308 with information on age and gender);
[b]Mean, minimum and maximum number of daily visits by hospital, gender and age;
[c]Total number of visits by temporal and meteorological factors (i.e. total number of visits in days with a particular value of temperature, humidity, etc.). In parenthesis the percentage of the number of visits of the total (2,223,479 total number of visits);
[d]Mean, minimum and maximum number of daily visits by temporal and meteorological factors (i.e. mean number of daily visits in the days with a particular value of temperature, Humidity etc.).

Model specification and ARIMA results

All models showed a very strong day of the week and year-round seasonality effect, according to ACF and PACF plots. To normalize residuals, outliers (in the training sets only) were replaced by the mean of the observations of the same day in the other years, consequently all models showed residual normally distributed according to the Jarque-Bera test (number of replaced outliers are presented in Supplementary Table 2). All models showed a lack of fitting on New Year's Eve and/or August 15th, for this reason we chose to define a specific dichotomous variable ("1" for the peculiar festivity, "0" for the other days) capable of detecting this extra variation. Table 2 displays the ARIMA parameters fitted for each model, and the number of Fourier terms that minimized AIC. All models were non-stationary in mean and needed one differencing to make the time series stationary (d=1). ARIMA parameters and Fourier terms were different across hospitals, showing that each time series needed different model specification. Table 2 also displays, for each hospital, the factors that significantly influenced the number of ED visits, and that were included in the models. High temperatures were always associated with a statistically significant increase in ED visit volumes, with a maximum increase of 1.84 daily visits every 1°C increase (hospital E, s.e. 0.18). Relative humidity was significantly associated with a limited decrease of total ED visits (-0.08, s.e. 0.04) for a 1% increment of RH only at hospital D. High levels of cumulative precipitation were associated (except for hospital C) with a statistically significant decrease in ED visits, with a maximum decrease of 0.31 daily visits every 1 mm of precipitation (hospital E, s.e. 0.06). Concerning air pollution, we found an opposite effect of NO_2 and PM10 on ED visits, with a mild significant negative effect for NO_2 in two hospitals (-0.08 and -0.09) and an even milder positive association with PM10 in one (0.03). Except for hospital C, the effect of ILI was always associated with the number of ED visits, showing an increase of daily visits between 0.73 and 1.74 (s.e. 0.29 and 0.41 respectively) at every

unit increase in weekly ILI rates. Festivities were associated with a decrease in ED visits of between 13 and 28 (s.e. 1.45 and 1.98), while special festivities were associated with the greatest decrease of at least 42 ED visits (s.e. 4.94). Autocorrelation function and correlation among residuals according to the Ljung-Box test by hospital and up to 30 and 366 lags can be found in Supplementary figure 1. ACF plots of residuals were overall in significance limits and the Ljung-Box test showed overall no significant correlation between residuals at different lags, except Hospital E which showed residual autocorrelation up to lag 366.

Table 2

Auto-regressive integrated moving average (ARIMA) specifications and covariate effects on the number of ED visits between the 1st of January 2014 and the 31st of December 2018 (training sets).

		Hospitals				
		A	B	C	D	E
Model specification	ARIMA parameters (p,d,q)	(0,1,2)	(1,1,1)	(1,1,2)	(1,1,1)	(1,1,1)
	Fourier terms†	3,13	3,14	3,13	3,16	3,15
Covariate Effects (se)[1]	Temperature (°C)	1.29 (0.15)	1.23 (0.14)	0.68 (0.11)	1.16 (0.18)	1.84 (0.18)
	Humidity (%)				-0.08 (0.04)	
	Precipitation (mm)	-0.2 (0.05)	-0.12 (0.05)		-0.13 (0.07)	-0.31 (0.06)
	NO2 (µg/m3)	-0.08 (0.03)			-0.09 (0.04)	
	PM10 (µg/m3)			0.03 (0.02)		
	ILI (weekly new cases per 1,000 inhabitants)	1.74 (0.41)	1.05 (0.37)	0.73 (0.29)		0.97 (0.46)
	Festivity		-28.23 (1.98)	-12.96 (1.45)	-25.42 (2.23)	-14.56 (2.39)
	Special Festivity*	-43.16 (6.31)	-57.64 (6.36); -62.61 (6.29)	-42.06 (4.92)	-59.86 (7.58)	-63.24 (7.92)
	Day before/after festivity	7.14 (1.5)	9.06 (1.58)	3.75 (1.22)		13.89 (1.96)

[1]Parameter estimates and standard errors in parentheses. Predictors were retained in the final model only if statistically significant (p-value<0.05)
† Number of sine and cosine terms used to approximate day of the week and year-round seasonality
***New Year's Eve for hospitals A, C-D and New Year's Eve and August 15th for hospital B**
ARIMA =Auto-regressive integrated moving average
ILI=Influenza-Like-Illness

Forecasting Accuracy and High demand definition

The accuracy of predictions (MAPE) in the validation sets, sensitivity and accuracy between observed and predicted high ED demand are displayed in table 3. Model performance was good, with small MAPEs in validation sets, ranging from a minimum of 5.5% for hospital D to a maximum of 8.1% for hospital C. The models showed high sensitivity on days with green-level high demand, almost 90% of days with predicted green-level high demand were confirmed from observed values.

On days with yellow-level high demand, sensitivity between predicted and observed demand was scarce, ranging from 0.04 for hospital B to 0.28 for hospital A. Sensitivity of red-level high demand varied between hospitals, with a minimum of 0.25 for hospital A to a maximum of 0.57 for hospital D. Observing Table 3 we can suggest that, for each hospital, at least 54% of the observed red-level high demand days were classified, from predictions, as being at least yellow-level. Accuracy was high, with at least 67% of the days with exactly the same predicted and observed high demand level (green, yellow or red).

All ARIMA models fitted the data significantly better than a simple regression model (M1) and a generalized linear model (M2), with MAPE for M1 and M2 above 13.5% and 9.8%, respectively (Supplementary Table 3). The scatter plot of observed vs predicted values in the validation set can be found in Supplementary Figure 2.

In Supplementary Table 2 we compared ARIMA results for different temperature specifications: mean, minimum, maximum and apparent temperature. The greatest effect on ED visits was attributed to mean temperature while indicators of performance and AIC where generally superior for mean temperature compared with minimum, maximum and apparent temperature. In Supplementary Table 2 we also calculated, only for outlier days, the relative error mean of observed vs predicted values in order to evaluate if extreme temperatures were better outlier predictors than mean temperature. Number of outliers replaced ranged from 2 for hospital A to 7 for hospital D, results suggested an overall better fit of outliers using minimum temperature (3 out 5 hospitals with smaller relative errors).

In supplementary table 4 we compared the high demand definition used in the ED warning system with similar definitions. There was slight improvement in percentage accuracy between the definition used and the other algorithms and there was no favourite algorithm for all hospitals: hospital B had a maximum improvement of 4% using the mean of the preceding 31 days or the median of the proceeding 21 days, hospitals A and C had an improvement of 2% using the mean of the preceding 31 days, hospital D had an improvement of 2% using the mean of the preceding 21 days, and finally hospital E had an improvement of 1% using the mean of the preceding 21 or 31 days. Using the high demand definition used by the Lombardy Region we did not find any improvement in accuracy, with an overall percentage of matched classification between 50% and 64%. High demand was always predicted less well compared to the definition used in our ED warning system. However, results showed good prediction of very high demand days with a sensitivity between 38% and 67%.

Table 3

Indicators of performance of the developed models: accuracy of predictions (MAPE) in the validation sets, and accuracy and sensitivity of high demand classification.

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity)		
				Green	Yellow	Red
Hospital A	5.9	72	Green	93	6	1
			Yellow	64	28	8
			Red	46	29	25
Hospital B	5.7	72	Green	92	8	0
			Yellow	85	4	11
			Red	35	15	50
Hospital C	8.1	67	Green	88	8	4
			Yellow	78	10	12
			Red	45	20	35
Hospital D	5.5	76	Green	91	6	3
			Yellow	65	27	8
			Red	35	9	56
Hospital E	6.1	74	Green	90	8	2
			Yellow	59	24	17
			Red	34	28	38

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

ED warning system

In Table 4a and 4b we provided the accuracy of predictions (MAPE), sensitivity and accuracy between observed and predicted high ED demand in January (the operating period of the WS) for one- and two-days horizons. Errors of prediction (MAPE) were slightly higher than in the validation set, with MAPE for one-day horizon always smaller than MAPE for two-days horizons. Accuracy between observed and predicted high ED demand was never smaller than 0.45 and generally smaller than in the validation set.

Table 4

Accuracy of predictions (MAPE), sensitivity, and accuracy between observed and predicted high ED demand in January 2020 (the operating period of the WS) with a one- (4a) and two-day (4b) horizon.

4a

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	7.8	52	Green	94	6	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	7.8	81	Green	87	13	0
			Yellow	0	100	0
			Red	17	17	67
Hospital C	8.6	52	Green	100	0	0
			Yellow	67	33	0
			Red	73	27	0
Hospital D	6.6	45	Green	55	36	9
			Yellow	0	33	67
			Red	50	33	17
Hospital E	11	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

4b

	MAPE	Accuracy (%)	Observed high EDdemand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	8.1	55	Green	100	0	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	8.6	71	Green	73	27	0
			Yellow	0	100	0
			Red	25	17	58
Hospital C	9	45	Green	93	7	0
			Yellow	83	17	0
			Red	82	18	0
Hospital D	7.6	48	Green	50	18	32
			Yellow	0	0	100
			Red	33	0	67
Hospital E	11.2	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

DISCUSSION

In this work we proposed and implemented in daily practice, a system to predict the number of ED visits in five hospitals of the city of Milan. The system is based on regression models with ARIMA errors, where ARIMA parameters were allowed to vary between hospitals, according to their specific characteristics, and it provides daily reports on the number of visits predicted for the two subsequent days at the five hospitals participating in the study. The models showed a good overall performance with the MAPEs always smaller than 5.5% and 8.1%. Our results are slightly better than other studies: Marcilio and colleagues[16] forecast daily ED visits with Generalized Linear Models, finding MAPEs between 5.4% and 11.5%, according to different forecasting horizons and controlling for temperature effect.

Jones and colleagues,[12] using similar models, found MAPEs that varied between 8.5% and 15.5%. However, Duwalage et al.[11] using a Generalized Additive Model found MAPEs consistently lower than 5% for 14-day forecasts, which significantly improved including temperature in the model. Although the number of predicted ED visits was close to the observed values, and there was good sensitivity in predicting mild (green) high demand, there was moderate sensitivity in predicting the spike of ED visit volumes (red-level high demand) for some hospitals and acceptable sensitivity for hospital D. This is particularly important for the scope of this study, which aimed to forecast emergency department visits in order to develop a 2-day warning system. For this reason, a better predictive performance of the red-level forecast would be desired. In fact, one of the major reasons for ED overcrowding is the shortage of acute care bed capacity compared with the huge number of visiting patients. Comparing our definition with similar definitions, we found a slight improvement in percentage accuracy, around 1% and 4%, but there was no a favourite algorithm for all hospitals. Furthermore, using the definition of very high demand for ED visits defined by the Lombardy Region, we found sensitivity was better compared to our models, and we plan to implement this in further evolutions of our warning system. However, we found good sensitivity in classifying observed red-level demand as at least yellow from predictions, and accuracy among observed and predicted high demand levels was always close to 70%. The definition of high ED demand is not straightforward as it relies on the specific hospital's characteristics. It is one of the main causes of ED overcrowding, which is the most problematic issue in EDs, thus deserving the effort in trying to predict it. In this study, we proposed a definition based on percentage increases compared to the median of the preceding month, to warn EDs of requests rising over the levels they managed in the preceding month.

During the operating period of the warning system, January 2020, we found a worse adaptation of the models than in the validation year 2019. This could be due to the ongoing outbreak of COVID-19, as ED visits for non-critical problems were discouraged.[34]

Concerning potential predictors, we found a strong day of the week and year-round seasonality effect, adequately captured by the terms used to approximate periodicity (Fourier terms). Even if the aim of this work was to develop a forecasting model and not an explanatory model, here we found statistically significant effects of meteorological factors on ED visits. Temperature was always positively associated with outcome, with an increase in the number of visits for each 1-degree increase in temperature across hospitals, in accordance with previous results.[13,15,19] As reported in another study,[35] high temperatures are associated with ED visits, especially for the most susceptible population, as persons with diabetes or cancer, so it is important for public health officials to implement adaptation measures to manage the impact of high temperatures on population health. Here we found a slightly better fit for outliers using minimum temperature instead of mean temperature. Nonetheless, we decided to include mean temperature in the ED warning system because it showed the greatest effect on ED visits. Further work has to be done in order to investigate the role of extreme temperature on ED visit fluctuations. The role of precipitations has not yet been well established. To our knowledge only one study measured an indirect effect in reducing ED visit volumes.[15] In accordance with these results, rainy days were found to be mildly associated with reduced numbers of ED visits. NO2 and PM10 had a mild significant effect only in two hospitals and in one hospital respectively, and were discordant, with a negative effect of NO2 and a positive effect of PM10 on the number of ED visits. This may be explained considering that the effect of pollution on ED visits is generally exerted and measured on respiratory conditions, especially asthma, and/or cardiac rather than with total visits and it may be diluted when analysing all ED visits. Only a few studies found a positive association of Total Suspended Particles with all visits but trauma, going in the same direction as the small significant increase in the number of visits related to PM10 we found.[36] In addition, pollution estimated from the monitoring station (classified as from urban traffic) used in the analysis might be of a greater magnitude than that really observed in each hospital. However, even though the hospitals were mostly located on the outskirts of the city of Milan, they are all located in urban areas characterized by a similar air pollution pattern. ILI were found to significantly increase the number of ED visits, as found by other researchers.[37]

This study indicated a moderate to good sensitivity in predicting high demand, showing some difficulties in anticipating the exact red-level days. In the future we aim to investigate models capable to directly predicting ED peaks instead of predicting the number of ED visits such as copulas used for detecting spikes in signal processing in brain circuits[38] or machine learning models. Finally, when interpreting these results, it is necessary to be aware of the possible multicollinearity problem between variables, which may alter the magnitude and statistical significance of coefficients. However, according to Vatcheva 2016,[27] only high correlations between variables would result in a change of sign of the coefficients and furthermore VIFs were always smaller than 5. Correlated factors were the pollution variables (NO2 and PM10), which were never considered in the same model together. Given that the highest correlation was found among temperature and ILI, the effect of these variables on the number of ED visits may potentially be biased due to multicollinearity. However, we included both terms in the models given the fact that they have an effect on ED visits independently from one another.

Another limitation is the choice of the hospitals considered for this work, that is, major hospitals located in the city of Milan. This methodology might not be the feasible for use by small hospitals as they might have low counts or even no visits at all on particular days. A solution can be provided by implementing different statistical models, for example, negative binomial or zero-inflated Poisson models, and would be one of our aims in the next years. In conclusion, we proposed a hospital specific ED warning system based on predictive models developed on previous attendances that can be used as a planning instrument in hospitals to increase resources, and to prevent high patient demand when a higher number of attendances is expected. This is important in any health system that usually deals with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits are still high.

List of abbreviations

ACF: autocorrelation function

AHP: Milan agency for health protection

AIC: minimal Akaike information criteria

API: application programming interface

AR: auto-regression

ARIMA: Auto-regressive integrated moving average

ED: emergency department

ILI: Influenza-like-illness

MA: moving average process

MAPE: mean absolute percentage error

NO2: nitrogen dioxide

PM10: particulate matter with a diameter ≤ 10 µm

PACF: partial autocorrelation function

RH: relative humidity

WS: warning system

Ethics approval and consent to participate This study does not involve human participants or animal subjects. Ethics approval and consent to participate were not required, as this is an observational study based on data routinely collected by the Agency for Health Protection (ATS) of Milan, a public body of the Regional Health Service-Lombardy Region. Among the institutional functions of ATS, established by Lombardy regional legislation (R.L. 23/2015), is management of the care pathway at the individual level in the regional healthcare system and evaluation of the services provided to, and the outcomes of, patients residing in the covered area. This study is also ethically compliant with Italian National Law (D.Lgs. 101/2018) and the "General Authorisation to Process Personal Data for Scientific Research Purposes" (nos. 8 and 9/2016, referred to in the Data Protection Authority action of 13/12/2018). Data were anonymized with a unique identifier in the different datasets before being used for the analyses.

Contributors RM and AGR conceptualised the study and defined the methodology, accessed and verified the data. RM analysed the data set, ST and AA contributed to the literature search, data interpretation, and writing of the manuscript. ST and AA, have made substantial contributions to the revision of the paper. AGR supervised and managed the project.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Data availability statement Data are not publicly available due to property restrictions.

Word count 4862

Acknowledgement Authors thank ARPA Lombardia, in particular Orietta Cazzuli and Gian Paolo Minardi, for their valuable contribution.

REFERENCES

1 Pines JM, Hilton JA, Weber EJ, *et al.* International perspectives on emergency department crowding. *Acad Emerg Med* 2011; **18**: 1358–70.

2 Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. *Emergency medicine journal : EMJ* 2009. DOI:10.1136/emj.2008.062380.

3 Programma Nazionale Esiti - PNE 2018 - Home. https://pne.agenas.it/index.php (accessed Feb 18, 2020).

4 Wang Z, Xiong X, Wang S, Yan J, Springer M, Dellinger RP. Causes of Emergency Department Overcrowding and Blockage of Access to Critical Services in Beijing: A 2-Year Study. *The Journal of Emergency Medicine* 2018; **54**: 665–73.

5 Jeyaraman MM, Copstein L, Al-Yousif N, *et al.* Interventions and strategies involving primary healthcare professionals to manage emergency department overcrowding: a scoping review. *BMJ Open* 2021; **11**: e048613.

6 Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emergency Medicine Journal* 2003; **20**: 402–5.

7 McKenna P, Heslin SM, Viccellio P, Mallon WK, Hernandez C, Morley EJ. Emergency department and hospital crowding: causes, consequences, and cures. *Clin Exp Emerg Med* 2019; **6**: 189–95.

8 Angelillo I, Ricciardi G, Nante N, *et al.* Appropriateness of hospital utilisation in Italy. *Public Health* 2000; **114**: 9–14.

9 Hakkoymaz H. How Appropriate are The Emergency Department Admissions? A Study From a Tertiary Hospital Emergency Department. *EJMI* 2020. DOI:10.14744/ejmi.2020.47376.

10 Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med* 2010; **17**: 970–8.

11 Duwalage KI, Burkett E, White G, Wong A, Thompson MH. Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas* 2020; published online Feb 17. DOI:10.1111/1742-6723.13481.

12 Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008; **15**: 159–70.

13 Rising WR, O'Daniel JA, Roberts CS. Correlating Weather and Trauma Admissions at a Level I Trauma Center. *Journal of Trauma and Acute Care Surgery* 2006; **60**: 1096–100.

14 Holleman DR, Bowling RL, Gathy C. Predicting daily visits to a waik-in clinic and emergency department using calendar and weather data. *J Gen Intern Med* 1996; **11**: 237–9.

15 Diehl AK, Morris MD, Mannis SA. Use of calendar and weather data to predict walk-in attendance. *South Med J* 1981; **74**: 709–12.

16 Marcilio I, Hajat S, Gouveia N. Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine* 2013; **20**: 769–77.

17 McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med* 2008; **15**: 337–46.

18 Díaz J, Alberdi JC, Pajares MS, *et al.* A model for forecasting emergency hospital admissions: effect of environmental variables. *J Environ Health* 2001; **64**: 9–15.

19 Macgregor DM. Effect of weather on attendance with injury at a paediatric emergency department. *Emergency Medicine Journal* 2003; **20**: 204–5.

20 Wargon M, Brun-Ney D, Beaujouan L, Casalino E. No more winter crisis? Forecasting daily bed requirements for emergency department admissions to hospital. *Eur J Emerg Med* 2018; **25**: 250–6.

21 Murtas R, Decarli A, Russo AG. Trend of pneumonia diagnosis in emergency departments as a COVID-19 surveillance system: a time series study. *BMJ Open* 2021; **11**: e044388.

22 Morbey RA, Charlett A, Lake I, *et al.* Can syndromic surveillance help forecast winter hospital bed pressures in England? *PLoS ONE* 2020; **15**: e0228804.

23 Italia. Provvedimento del Garante n. 2 del 16 giugno 2004. Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici. .

24 Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. ARPA Lombardia. https://www.arpalombardia.it:443/Pages/ARPA_Home_Page.aspx (accessed March 10, 2020).

25 Steadman RG. Norms of apparent temperature in Australia. *Aust Met Mag* 1994; **43**: 1–16.

26 Ministero della Salute. Influenza. http://www.salute.gov.it/portale/influenza/homeInfluenza.jsp (accessed March 10, 2020).

27 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* 2016; **6**. DOI:10.4172/2161-1165.1000227.

28 Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018.

29 M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *IJDKP* 2015; **5**: 01–11.

30 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020 https://www.R-project.org/.

31 Salute M della. Pronto soccorso, via libera della Conferenza Stato-Regioni alle nuove linee di indirizzo. http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=3849 (accessed April 24, 2020).

32 Hyndman [aut R, cre, cph, *et al.* forecast: Forecasting Functions for Time Series and Linear Models. 2020 https://CRAN.R-project.org/package=forecast (accessed Nov 14, 2020).

33 Fox J, Weisberg S. An R Companion to Applied Regression. SAGE Publications, 2018.

34 Jeffery MM, D'Onofrio G, Paek H, *et al.* Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. *JAMA Intern Med* 2020; published online Aug 3. DOI:10.1001/jamainternmed.2020.3288.

35 Lavigne E, Gasparrini A, Wang X, *et al.* Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environ Health* 2014; **13**: 5.

36 Samet JM, Speizer FE, Bishop Y, Spengler JD, Ferris BG. The Relationship between Air Pollution and Emergency Room Visits in an Industrial Community. *Journal of the Air Pollution Control Association* 1981; **31**: 236–40.

37 Schanzer DL, Schwartz B. Impact of seasonal and pandemic influenza on emergency department visits, 2003-2010, Ontario, Canada. *Acad Emerg Med* 2013; **20**: 388–97.

38 Hu M, Li M, Li W, Liang H. Joint analysis of spikes and local field potentials using copula. *NeuroImage* 2016; **133**: 457–67.

Figure 1

Location of the five partecipating hospitals and of meteorological and air pollution monitoring stations in the city of Milan.

Figure 2

Hypothetical daily report received from a hospital on the 5th of January 2020.

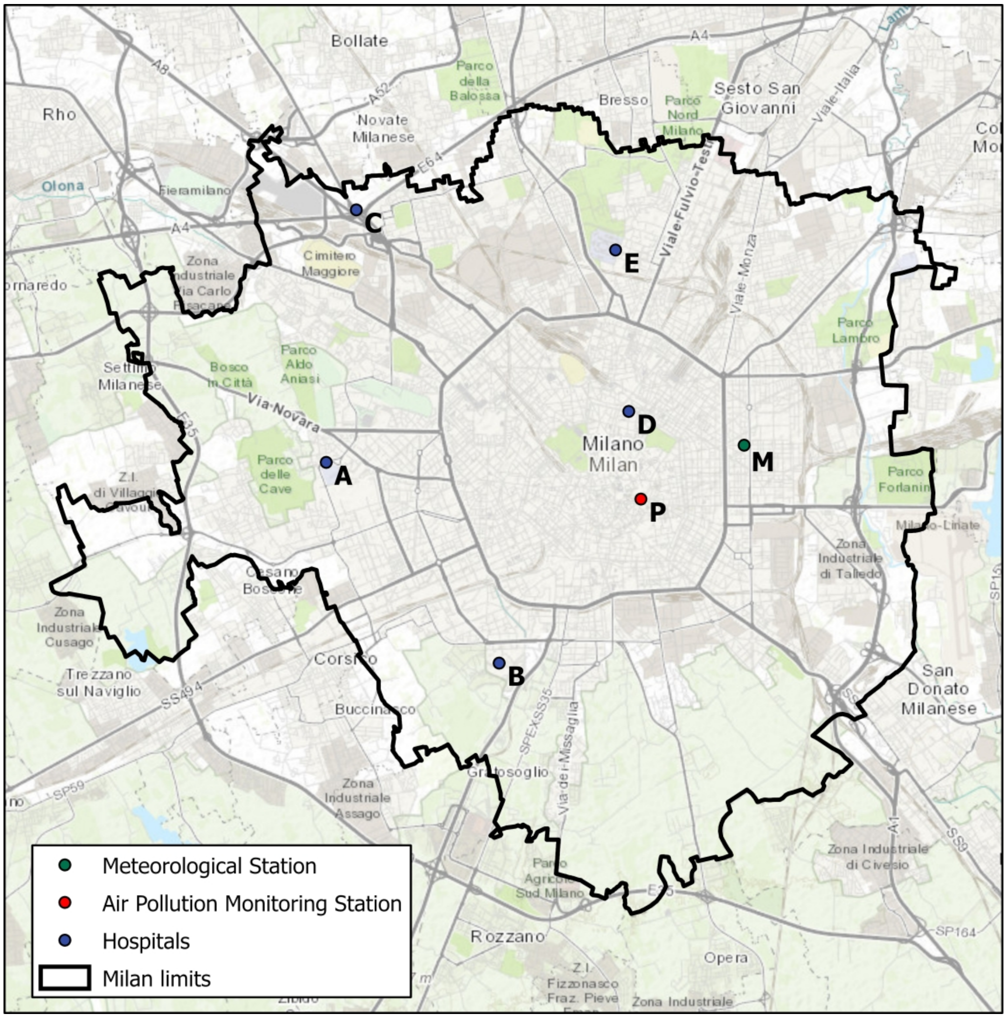


Figure 1. Location of the five considered hospitals and of meteorological and air pollution monitoring station in the city of Milan.

160x161mm (600 x 600 DPI)

EMERGENCY ADMISSION 2-DAY WARNING SYSTEM

Agency for Health Protection of Milan

Hospital A. Prediction for the day:

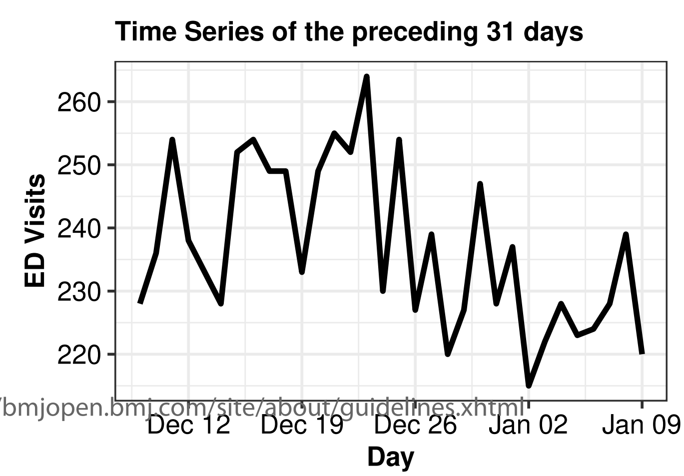
11 Jan 2020	12 Jan 2020
Level 1	Level 2

Predicted ED visits ± 95% Margin Errors	227 ± 12	254 ± 14
Covariates		
NO2	$46\ \mu g/m^3$	$38.4\ \mu g/m^3$
PM10	$6\ \mu g/m^3$	$3.9\ \mu g/m^3$
TEMPERATURE	5°	6°
UMIDITY	65%	57%
PRECIPITATION	$0\ mm$	$0\ mm$
ILI RATE	10.87	10.87

Methodology

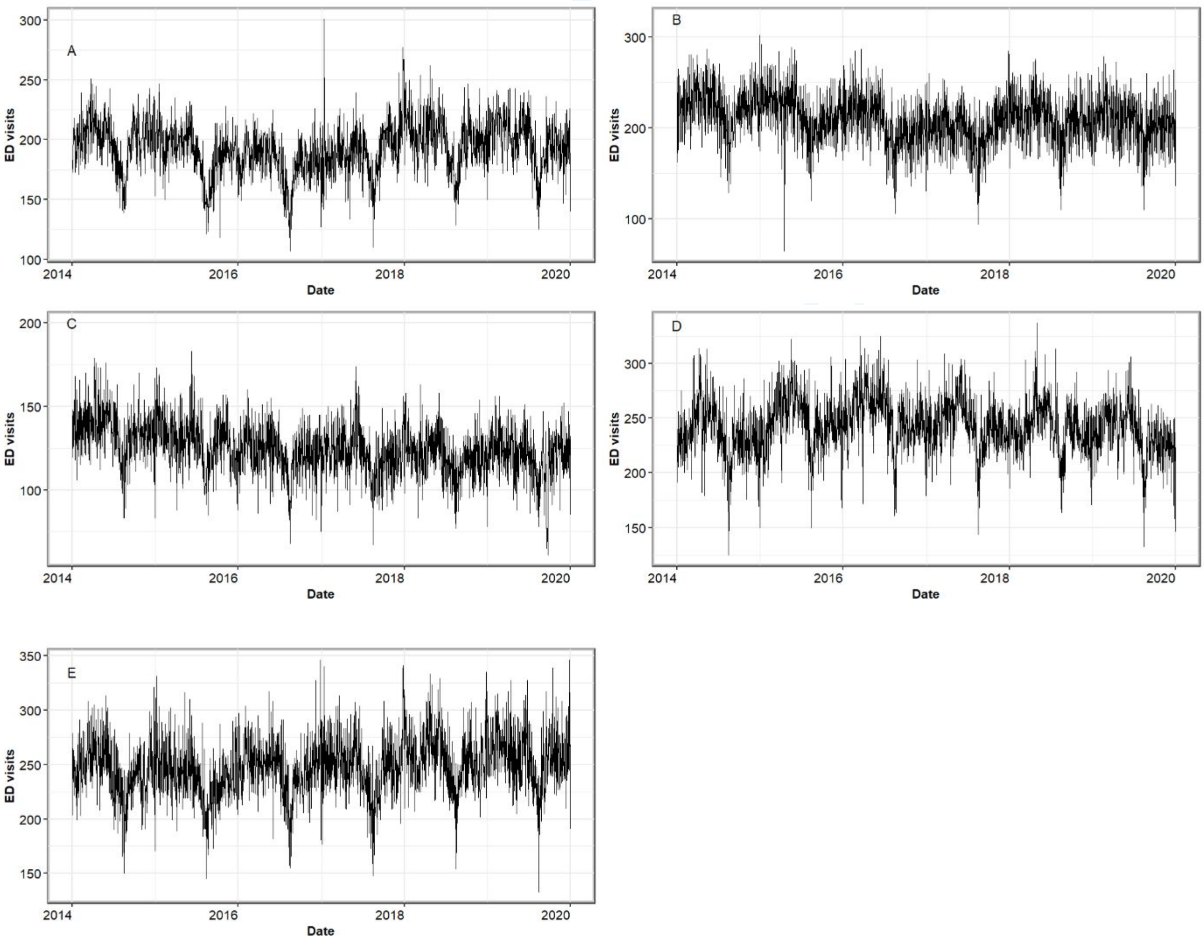
Level 1	Number of visits exceeded the median by less than 5%.
Level 2	Number of visits exceeded the median between 5% and 10%.
Level 3	Number of visits exceeded the median by more than 10%.
Prediction	Prediction based on a regression model with ARIMA errors (Hyndman 2018).
Environment	Meteo and Pollution Forecast from ARPA Lombardia.
ILI rates	Influence Like Illness rate, number of cases per 1,000 inhabitants per week (InfluNet).

Sistema Socio Sanitario
Regione Lombardia
ATS Milano
Città Metropolitana



Description of Training and validation sets

Patients were mostly female (50.6%), with a mean age of 44 years (standard deviation s.d. of 26 years). The mean number of daily visits was similar in genders. Higher ED accesses volumes were found among people aged over 65 years than the others ages. During the study period, mean temperature was 16°C, mean RH was 63%, and there was precipitation on 31% of the days. Mean level of NO2 exceeded the European limit (40 µg/m^3), with a mean of 46 µg/m^3, while mean PM10 was lower than the European limit, with a mean of 35 µg/m^3, during the observed period. On days-before and after festivities, we measured a higher number of visits, while on festivity days there was a lower number of visits compare to normal days. Training and validation sets were similar according to meteorological factors, but there were mild differences in air pollution and ILI rates (supplementary Table 1). The year 2019 was in fact characterized by significantly higher levels of pollution (t-test p-values<0.001) and lower ILI rates (t-test p-value<0.01) compared to the previous years. Patients were slightly younger in the training set than in the validation set, the mean age being 43 years in the former and 45 in the latter. The number of ED accesses was statistically different across age groups between days: children (0-14 years) tended to visit ED more likely on weekends (20% higher on Sundays compared to other days) while adults and senior people (15-65 and >65 years) on Mondays (14% and 11% higher than other days) (Anova test for mean differences p-values<0.001). August was the months with smaller ED accesses volumes, with a 14% decrease compared to the average of the other months.



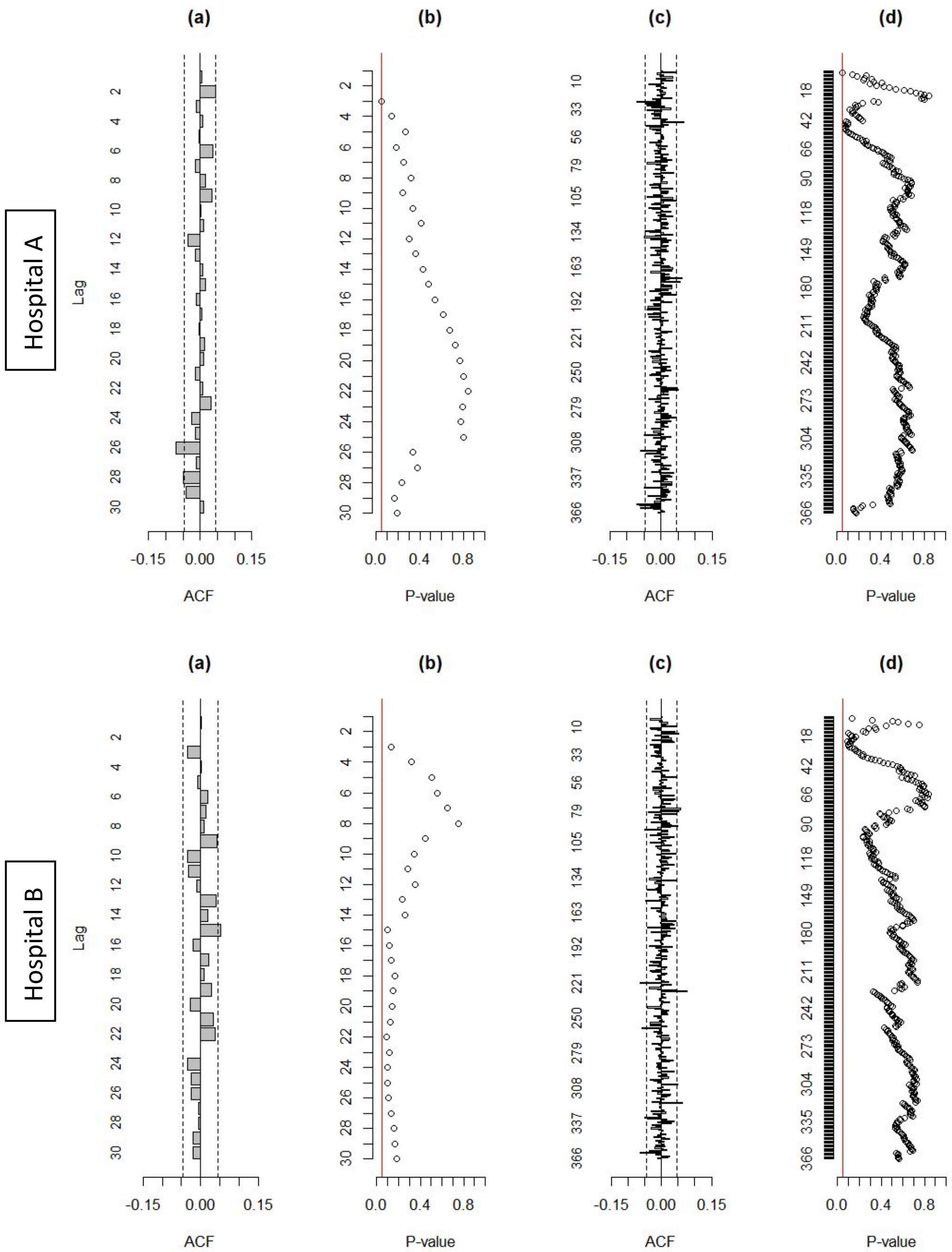
Supplementary Table 1

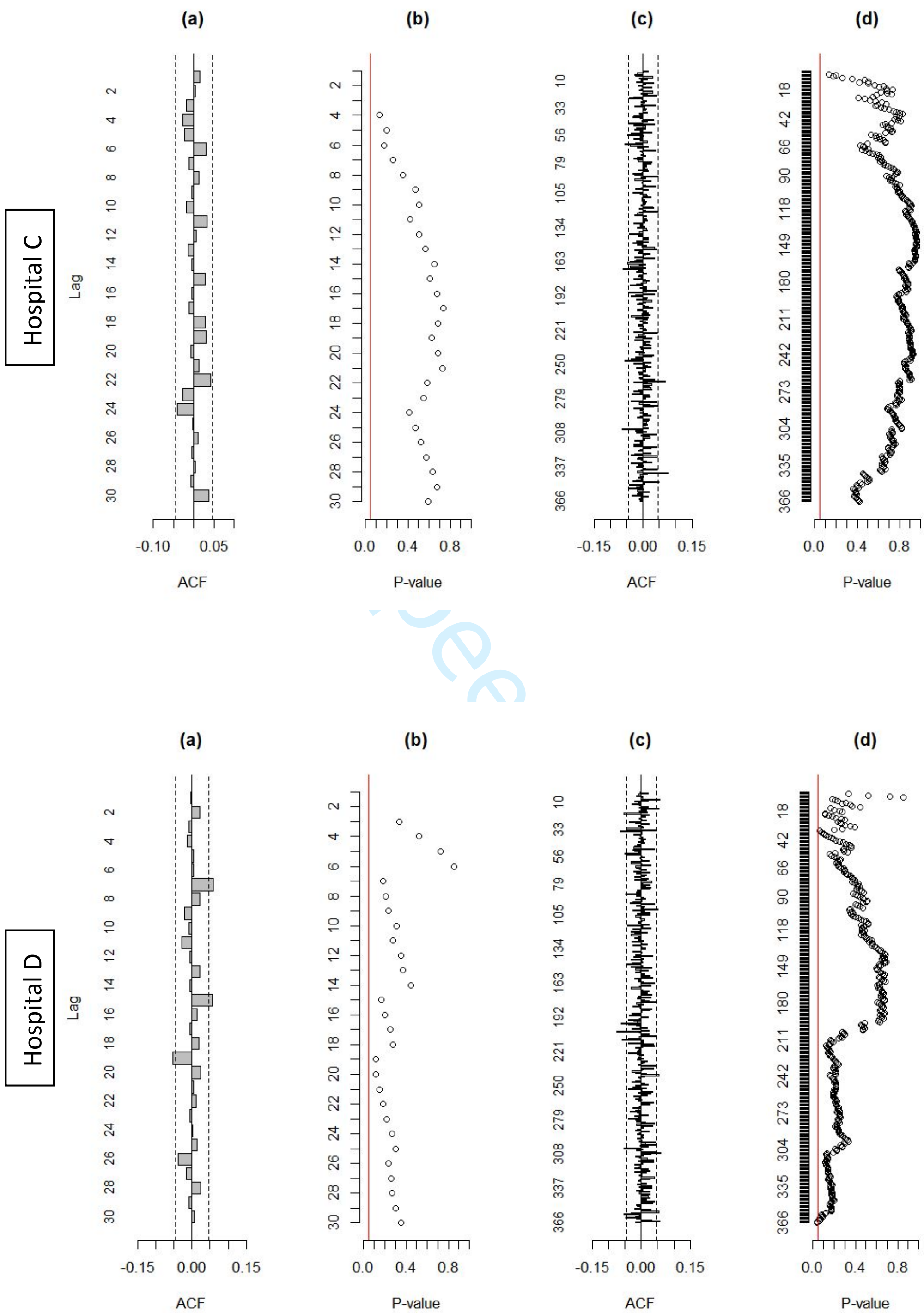
Mean, standard deviation and t-test for mean difference between training and validation sets by covariates.

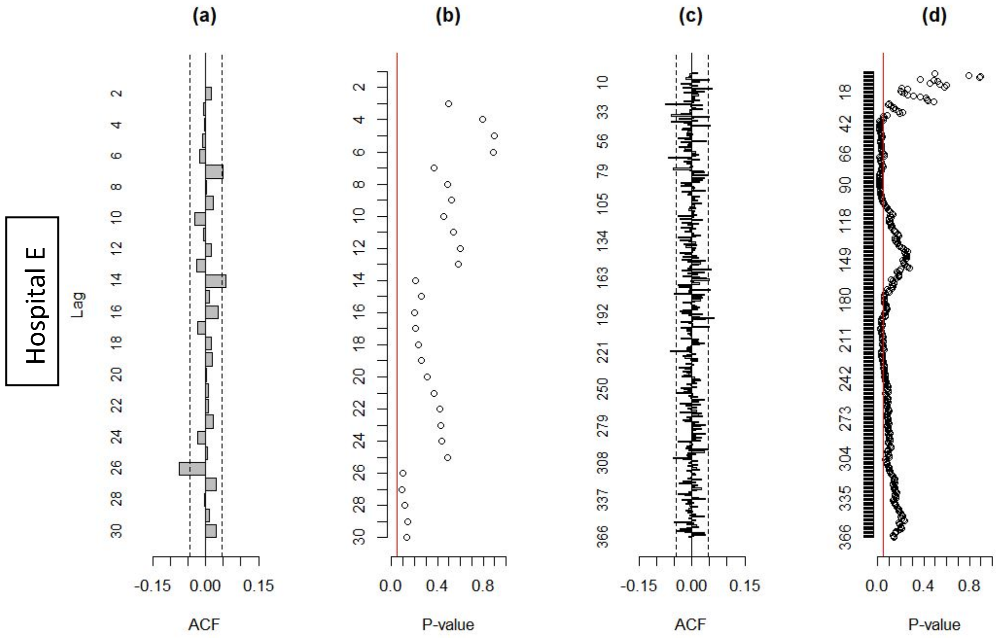
	Mean (s.d.)		t-test p-value
	Training	Validation	
Female (%)	50.6	50.4	-
Age (years)	43 (26)	45 (26)	<0.001
Temperature (°C)	15.7 (8)	16.4 (8)	0.1257
Relative Humidity (%)	63.4 (17)	62.3 (17)	0.2765
Cumulative Precipitation (mm)	2.3 (6.6)	2.2 (5.8)	0.7155
NO2 (µg/m3)	47 (19)	39 (17)	<0.001
PM10 (µg/m3)	36 (21)	30 (18)	<0.001
ILI (new cases per 1,000 inhabitants)	1.9 (3)	2.6 (3.8)	<0.001
S.d.=standard deviation ILI=Influenza-Like-Illness			

Supplementary Figure 1

Autocorrelation function (ACF) and correlation among residuals according to the Ljung-Box test (LB) by hospital: (a) ACF up to lag 30; (b) LB test up to lag 30; (c) ACF up to lag 366; (d) LB test up to lag 366.




(a)
(b)
(c)
(d)
Hospital C
Hospital D
Lag
ACF
P-value


Hospital E
(a)
(b)
(c)
(d)
Lag
ACF
P-value
ACF
P-value

Supplementary Table 2

Auto-regressive integrated moving average (ARIMA) parameters, indicators of performance (accuracy of predictions in the validation sets, and accuracy of high demand classification), Akaike Information Criteria (AIC) and relative error mean for outliers.

		Mean temperature					Minimum Temperature					Maximum Temp					Apparent Temperature				
	Number of outliers' days*	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative Error mean*	AIC
Hospital A	2	1.29 (0.15)	5.9	72	55.3	14861	1.13 (0.15)	5.9	73	55.1	14882	1.03 (0.12)	5.9	72	55.1	14864	1.0[illegible] (0.1[illegible])	5.9	72	55.5	14870
Hospital B	6	1.23 (0.14)	5.7	72	44.7	14847	1.02 (0.15)	5.7	74	44.1	14874	0.98 (0.11)	5.7	73	45.6	14850	1.04 (0.12)	5.7	73	44.9	14846
Hospital C	5	0.68 (0.11)	8.1	67	36.5	13928	0.57 (0.11)	8.1	67	36.8	13941	0.55 (0.09)	8.1	66	36.5	13925	0.54 (0.09)	8.1	66	36.8	13934
Hospital D	7	1.16 (0.18)	5.5	76	22.3	15506	1.00 (0.18)	5.6	76	22.6	15516	0.96 (0.15)	5.6	75	22.4	15508	0.88 (0.15)	5.5	76	22.3	15511
Hospital E	6	1.84 (0.18)	6.1	74	24.4	15624	1.68 (0.19)	6.1	73	24	15649	1.5 (0.14)	6.1	73	24.9	15622	1.53 (0.16)	6.3	71	24.2	15811

*Number of outliers' days replaced by the mean of the observations of the same day in the other years for normalization of results

**Relative error mean of observed vs predicted values calculated for outliers (in the training sets only) which were replaced by the mean of the observations of the same day in the other years for normalization of residuals

Supplementary Table 3

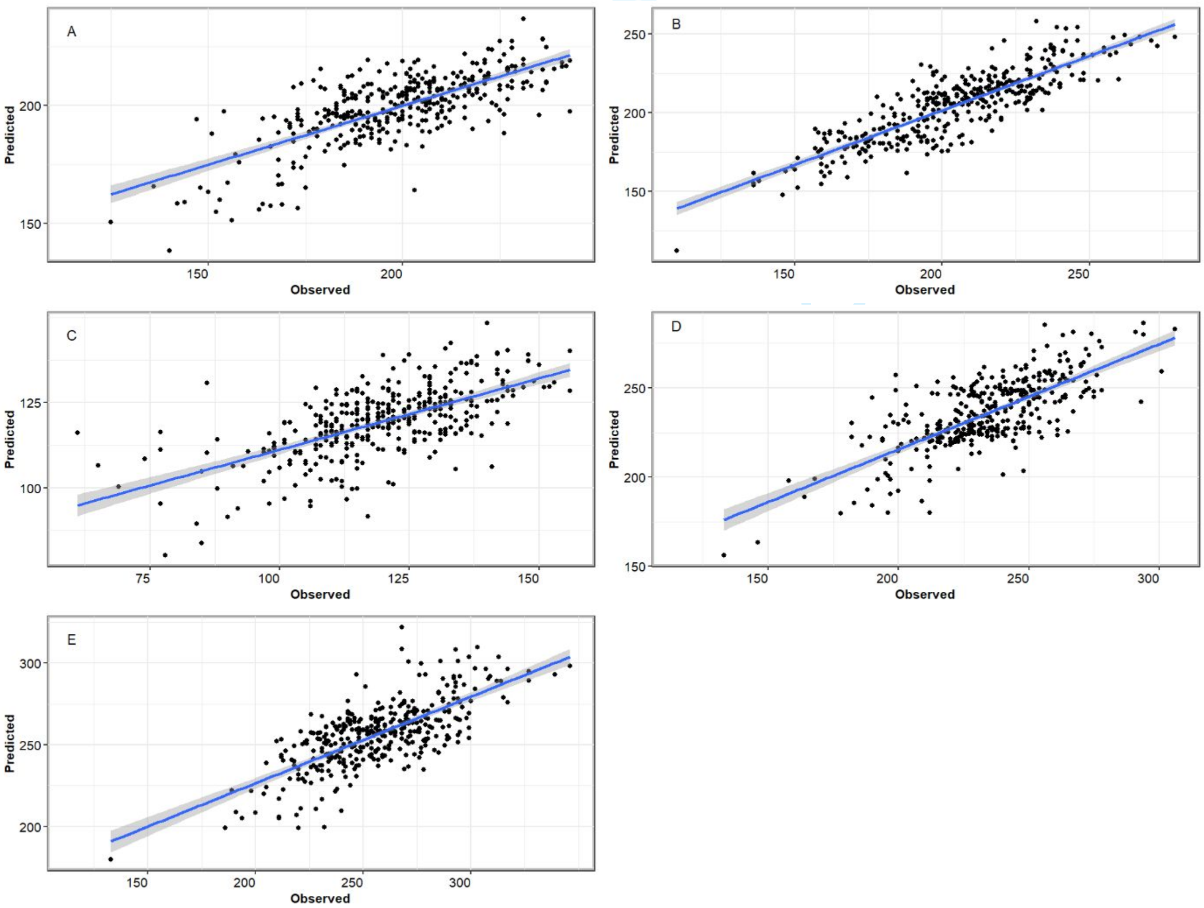
Model comparisons between the regression model with ARIMA errors and: a simple regression model (M1) and a generalized linear model (M2).

	Linear model (M1[a])		Likelihood ratio test*		Generalized linear model (M2[b])		Likelihood ratio test**	
	MAPE	Accuracy (%)	LRT (df)	p-value	MAPE	Accuracy (%)	LRT (df)	p-value
Hospital A	17.4	52	3957 (34)	<0.001	11.9	56	2618 (2)	<0.001
Hospital B	25.3	49	5610 (36)	<0.001	13.2	60	3318 (2)	<0.001
Hospital C	20.8	51	3975 (35)	<0.001	14.4	58	2662 (3)	<0.001
Hospital D	13.5	51	3289 (40)	<0.001	9.8	57	2164 (2)	<0.001
Hospital E	23.9	55	5222 (38)	<0.001	12.3	57	2909 (2)	<0.001

[a]M1: linear model with only meteorological, environmental and festivities covariates
[b]M2: generalized linear model with meteorological, environmental, festivities covariates and Fourier terms to control for seasonality
*Likelihood ratio test (LRT) comparing the regression model with ARIMA errors with M1
**Likelihood ratio test (LRT) comparing the regression model with ARIMA errors with M2

Supplementary Figure 2

Observed vs predicted ED visits in the validation sets (from the 1th of January 2019 to the 31th of December 2019) by hospital.



Supplementary Table 4

Indicators of performance (accuracy and sensitivity of high demand classification) by different definition of high demand: the number of visits exceeded the median of the preceding 7, 14 and 21 days (4a), the number of visits exceeded the mean of the preceding 7, 14, 21 and 31 days (4b), and high demand defined by the Lombardy Region as exceeding thresholds based on previous year percentiles (4c).

4a	Median of the preceding 7 days					Median of the preceding 14 days					Median of the preceding 21 days				
			Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)		
	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red
Hospital A	72	Green	95	5	0	73	Green	95	5	0	73	Green	95	5	0
		Yellow	84	16	0		Yellow	71	21	8		Yellow	72	21	7
		Red	58	32	10		Red	59	31	10		Red	46	34	20
Hospital B	75	Green	97	2	1	75	Green	97	3	0	76	Green	96	4	0
		Yellow	80	9	11		Yellow	80	8	12		Yellow	75	16	9
		Red	38	14	48		Red	32	19	49		Red	41	10	49
Hospital C	64	Green	91	6	3	66	Green	90	6	4	67	Green	91	6	3
		Yellow	66	16	18		Yellow	63	23	14		Yellow	72	15	13
		Red	59	22	19		Red	56	22	22		Red	47	23	30
Hospital D	75	Green	92	6	2	74	Green	90	7	3	76	Green	92	6	2
		Yellow	68	18	14		Yellow	81	11	8		Yellow	73	16	11
		Red	47	14	39		Red	32	16	52		Red	29	10	61
Hospital E	70	Green	91	5	4	72	Green	92	6	2	74	Green	91	6	3
		Yellow	77	4	19		Yellow	69	9	22		Yellow	64	16	20
		Red	40	23	37		Red	39	20	41		Red	38	17	45

ED=Emergency Department

4b	Mean of the preceding 7 days					Mean of the preceding 14 days					Mean of the preceding 21 days					Mean of the preceding 31 days				
			Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)		
	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red
Hospital A		Green	96	4	0		Green	96	4	0		Green	94	5	1		Green	94	4	2
	70	Yellow	87	10	3	72	Yellow	78	15	7	73	Yellow	76	21	3	74	Yellow	65	27	8
		Red	63	30	7		Red	60	34	6		Red	49	27	24		Red	44	31	25
Hospital B		Green	92	6	2		Green	91	9	0		Green	90	10	0		Green	91	8	1
	73	Yellow	74	13	13	72	Yellow	74	16	100	73	Yellow	71	21	8	76	Yellow	68	23	9
		Red	29	19	52		Red	26	20	54		Red	30	17	53		Red	24	15	61
Hospital C		Green	90	7	3		Green	93	4	3		Green	91	5	4		Green	91	5	4
	65	Yellow	74	15	11	67	Yellow	68	21	11	66	Yellow	71	18	11	69	Yellow	72	17	11
		Red	55	27	18		Red	53	23	24		Red	54	20	26		Red	49	19	32
Hospital D		Green	92	6	2		Green	92	5	3		Green	93	6	1		Green	93	6	1
	73	Yellow	74	10	16	77	Yellow	76	14	10	78	Yellow	72	15	13	76	Yellow	63	26	11
		Red	51	17	32		Red	35	16	49		Red	32	12	56		Red	36	18	46
Hospital E		Green	94	4	2		Green	93	5	2		Green	92	6	2		Green	93	5	2
	76	Yellow	70	9	21	75	Yellow	66	17	17	75	Yellow	64	21	15	75	Yellow	68	15	17
		Red	46	22	32		Red	44	18	38		Red	39	23	38		Red	30	30	40

ED=Emergency Department

4c	Mean of the preceding 7 days					
			Predicted ED high demand (%, Sensitivity)			
	Accuracy (%)	Observed ED high demand	Low	Middle	High	Very high
Hospital A	53	Low	62%	36%	1%	1%
		Middle	26%	56%	14%	4%
		High	8%	37%	33%	22%
		Very high	0%	20%	23%	57%
Hospital B	64	Low	79%	21%	0%	0%
		Middle	14%	73%	12%	1%
		High	0%	62%	24%	14%
		Very high	0%	15%	18%	67%
Hospital C	50	Low	66%	31%	3%	0%
		Middle	21%	60%	14%	5%
		High	6%	56%	19%	19%
		Very high	4%	20%	38%	38%
Hospital D	55	Low	67%	30%	3%	0%
		Middle	33%	53%	11%	3%
		High	11%	28%	44%	17%
		Very high	0%	41%	9%	50%
Hospital E	54	Low	56%	42%	2%	0%
		Middle	17%	66%	15%	3%
		High	3%	54%	26%	17%
		Very high	0%	20%	20%	60%

ED=Emergency Department

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	-
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2/3
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2/3
Bias	9	Describe any efforts to address potential sources of bias	3
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2/3
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	3/4
		(*b*) Describe any methods used to examine subgroups and interactions	-
		(*c*) Explain how missing data were addressed	3
		(*d*) If applicable, explain how loss to follow-up was addressed	-
		(*e*) Describe any sensitivity analyses	5
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	7
		(c) Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	Report numbers of outcome events or summary measures over time	7

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8/9/10
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	-
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

A time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: a 2-day warning system

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-056017.R2
Article Type:	Original research
Date Submitted by the Author:	21-Mar-2022
Complete List of Authors:	Murtas, Rossella; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Tunesi, Sara; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit; Andreano, Anita; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit Russo, Antonio ; Agency for the Protection of Health of the Metropolitan Area of Milan, Epidemiology Unit
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	ACCIDENT & EMERGENCY MEDICINE, EPIDEMIOLOGY, QUALITATIVE RESEARCH, STATISTICS & RESEARCH METHODS

Original Paper

A time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: a 2-day warning system

Author List: Rossella Murtas, PhD[1], Sara Tunesi, PhD[1], Anita Andreano, MD[1], Antonio G. Russo, MD PhD[1]

Author Affiliations: Epidemiology Unit, Agency for Health Protection of Milan, 20123 Milan, Italy

Corresponding Author: Antonio Giampiero Russo, Via Conca del Naviglio 45, 20122 Milan, Italy, agrusso@ats-milano.it, +390285782111, +390285782128

ABSTRACT

Objectives The emergency department (ED) is one of the most critical areas in any hospital. Recently, many countries have seen a rise in the number of ED visits, with an increase in length of stay and a detrimental effect on quality of care. Being able to forecast future demands would be a valuable support for hospitals to prevent high demand, particularly in a system with limited resources where use of ED services for non-urgent visits is an important issue.

Design Time series cohort study.

Setting We collected all ED visits between January 2014 and December 2019 in the five larger hospitals in Milan. To predict daily volumes, we used a regression model with ARIMA errors. Predictors included were day of the week and year-round seasonality, meteorological and environmental variables, information on influenza epidemics and festivities. Accuracy of prediction was evaluated with the Mean Absolute Percentage Error (MAPE).

Primary outcome measures Daily all-cause EDs visits.

Results In the study period, we observed 2,223,479 visits. ED visits were most likely to occur on weekends for children and on Mondays for adults and seniors. Results confirmed the role of meteorological and environmental variables and the presence of day of the week and year-round seasonality effects. We found high correlation between observed and predicted values with a MAPE globally smaller than 8.1%.

Conclusions Results were used to establish an ED warning system based on past observations and indicators of high demand. This is important in any health system that regularly faces with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits is still high.

Strengths and limitations of this study

- This study is one of the few studies linking temporal periodicity, occurrence of festivities, local weather conditions, and pollution to ED visits
- We estimated an ARIMA model for each hospital, thus taking into consideration each specific characteristic and incorporating weekly and annual seasonality with Fourier terms
- Results were used to establish an ED warning system based on past observations and indicators of high demand
- We cannot exclude the possible presence of unmeasured variables that may better predict ED visits and overcrowding

INTRODUCTION

The emergency department (ED) is the gateway (an open door) and the most critical area of a hospital, moving many activities and causing problems in the management of elective procedures when the number of patients who come knocking increases. In the last decade, many countries have seen a substantial rise in the number of ED visits, with an increase in length of stay,[1] and associated detrimental effects on quality of care. ED visits are unavoidably subject to fluctuation, and several models to predict high demand have been developed in the last decade, aiming at effectively managing hospital beds and staff rosters.[2] In Italy, even though the number of ED visits has been decreasing since 2016, the mean waiting time in EDs was high, between 12h and 24h in 3.5% of cases in 2017, and over 24 h in 2.1% of cases.[3] The definition of overcrowding[4] in the ED literature is not consistent, nor are the measures used to assess overcrowding, which vary from clinician perception of overcrowding, to input measures (e.g., waiting times, number of patients arrived), throughput measures (e.g., ED capacities, patient care time), output measures (e.g. percentages of hospital admissions, hospital beds), or multidimensional indices such as the Emergency Department Work Index (EDWIN). This variety of measures corresponds to the different type of factors studied as causes of ED crowding. We concentrate here on predicting the number of visits from input factors, i.e., determinants and modalities of patient inflow, such as non-urgent visits and Influenza season. In this case it is better to speak of overflow.[5] We did not investigate throughput factors, describing organizational issues in the ED, such as inadequate staffing, nor output factors. The latter include one of the major reasons for ED overcrowding, which is the shortage of acute care bed capacity.[6–10] . Among the most investigated input factors are non-urgent visits, meaning "patients who could have been assessed and treated in other facilities that treat less urgent cases".[11] In Italy in 2017, only 23% of ED visits were classified as red or yellow at triage, while 13%[3] had a low level of priority, coded white triage in Italy. This use of emergency department services is a signal of lack of continuity of primary care and difficulty of access to both primary and specialist care. It is also not cost-effective and leads to an increase in waiting times in the EDs.[12,13]

Several factors potentially affect the daily number of ED visits. Among these: annual[14,15], seasonal,[16–19] and weekly[14–19] periodicity, as well as festivities.[14,16,20,21] The effect of local weather conditions and pollution on ED visit volumes is still in debated: while some studies confirmed a significant association with temperature,[15,17–19,22,23] precipitation,[17,19] humidity,[22] and weather conditions,[23] other authors found these variables to be only mediocre predictors of the number of ED visits,[16] and found air pollution mostly impacting cardiac and respiratory diseases.[22] An additional factor that has been studied in relation with ED visit volumes is the flu, with around 7% of total accesses attributable to Influenza-like Illness (ILI) during the epidemic season.[24] Murtas and colleagues[25] evaluated the hypothesis of the early presence of the COVID-19 epidemic in Italy by analysing data on trends of access to EDs using a Poisson regression model adjusted for seasonality and influenza outbreaks. In this work they found that predicting ED visits by considering both seasonality and ILI rates, compared to a model tacking into account only seasonality, notably increased the fitting of the model. Therefore, syndromic surveillance (such as ILI rates which in Italy are provided weekly by the National Health Service Sentinel System) may be able to provide early warning of hospital bed capacity strain caused by seasonal respiratory disease.[26] To our knowledge, there is no study linking all this information together to ED visits.

The present study aims to develop a model for forecasting ED arrivals, using regression-based time series analysis with Auto-regressive integrated moving average (ARIMA) errors, accounting simultaneously for the effect of meteorological and environmental variables, as well as information on flu epidemics and festivities, on the number of ED visits in the city of Milan. The model is used to establish an innovative ED warning system providing a planning instrument for hospitals, based on past observations and indicators of high demand.

METHODS

Study Design

This is a retrospective study conducted in the area served by the Milan Agency for Health Protection (AHP) using current health care databases of daily ED visits aggregated at hospital level. No individual level data were used, and patients cannot be identified from aggregated data which do not contain low counts (i.e. cells with ≤ 5 counts). For this reason, and in accordance with Italian legislation, this study was not submitted for ethics approval.[27]

Study setting and population

We collected all ED visits, including patients registered at triage that voluntarily left the ED premises before being evaluated by a physician, between the 1st of January 2014 and the 31st of December 2019 in the five largest hospitals located in the city of Milan (figure 1). All five hospitals are public hospitals and received 49% of all emergency room access of the city of Milan, which has a total of 17 EDs, with a mean number of daily ED visits during 2014-2019 ranging from 124 for hospital C to 247 for hospital E.

Study protocol

Aggregated data on daily ED visit volumes, by age and gender, were extracted from the regional health database. Meteorological and environmental information was extracted from the Regional Environmental Protection Agency (ARPA).[28] Daily mean temperature, relative humidity (RH), cumulative precipitation, nitrogen dioxide (NO2), and particulate matter with a diameter ≤ 10 μm (PM10) were collected from 2 monitoring stations (one measuring meteorological indicators and one measuring air pollution) located in the centre of Milan (figure 1). As sensitivity analysis we also investigated the effect of minimum, maximum and apparent temperature on daily ED visits.[29] Missing values on a specific day were imputed with the average of the measure in that specific year. Weekly data on ILI notifications were taken from the National Health Service Sentinel System (InfluNet).[30] Weekly incidence rates of ILI were expressed as the number of cases per 1,000 inhabitants per week. All available information was linked to daily ED visit volumes for each of the five hospitals included in the study. Datasets were divided into training (from the 1st of January 2014 to the 31st of December 2018) and validation sets (from the 1st of January 2019 to the 31st of December 2019). For each hospital, we first estimated model parameters on the training dataset and evaluated post-sample accuracy in the validation set. We included, in each model, only factors that significantly influenced the number of ED visits. Multicollinearity was evaluated calculating Pearson pairwise correlation between variables and variance inflation criterion (VIF)[31].

Patient and public involvement

Patients were not involved in this research.

Data analyses

Development of the predictive model

To predict the daily volume of visits in each ED, we used a time series approach consisting in a regression model with ARIMA errors.[32] The statistical units were days, 1,826 days in the training set and 365 in the validation set. This model is able to combine two powerful statistical methods: linear regression and ARIMA. Linear regression of Y on X is usually described by the equation $Y_t = \alpha + \beta x_t + \epsilon_t$, where Y_t and x_t are the values of Y and X at day t, α and β are the intercept and the slope of the regression line, and ϵ_t is the error of the model at day t (the deviations from the fitted line to the observed values) assumed to be independent from other values. The ARIMA model deals with auto-correlation between errors through two components: the auto-regressive and the moving average process. The auto-regressive component assumes that previous observations are good predictors for future values, while the moving average component allows the model to update the predictions if the level of a constant time series changes. ARIMA specification is described by 3 parameters (p, d, q), where p is the order of auto-regression (AR) that is the number of time lags, d is the degree of differencing (the number of times the data have had past values subtracted to make the time series stationary), and q is the order of the moving average process (MA). For each hospital, these parameters were identified examining total and partial autocorrelation function (ACF and PACF, respectively), as well as statistical significance (p-value<0.05), and minimal Akaike Information Criteria (AIC). Day of the week and year-round seasonality were controlled for by including Fourier terms, a series of sine-cosine functions capable of approximating periodicity.[20,32] The number of Fourier terms was chosen to minimise the AIC for each seasonal period (up to 7 for day of the week seasonality and up to 365 for year-round seasonality). Each seasonal component can be written in the model equation as

$$\sum_{j=1}^{n}\left[\alpha_j \sin\left(\frac{2\pi jt}{m}\right) + \beta_j \cos\left(\frac{2\pi jt}{m}\right)\right]$$

where *n* is the number of Fourier terms chosen to minimise the AIC (up to 7 for day of the week seasonality and up to 365 for year-round seasonality) and *m* is the seasonal period (7 for day of the week and 365 for year-round seasonality).

Therefore, meteorological and environmental variables, as well as information on flu epidemics and festivities, were retained in the final model only if statistically significant. As festivities, we considered Italian public holidays with school and office closures: New Year's Day, Epiphany, Easter Sunday and Monday, Italian Liberation Day, Labour Day, Foundation of the Italian Republic, assumption day, All Saints' Day, Saint Ambrose's Day (local patron saint), Feast of the Immaculate Conception, Christmas Day, Saint Stephen's day and New Year's Eve. In addition, we created dummies for specific festivities that were responsible for a significant variation in the number of ED visits: New Year's Eve and Assumption Day (August 15th). Diagnostics of the finally selected models were the Jarque-Bera test of normality, and correlation among the residuals according to the Ljung-Box test. Variables and tests were considered statistically significant if p-value was < 0.05.

The ARIMA model was compared with a simple regression model (M1) including only meteorological, environmental, and festivity covariates and with a generalized linear model (M2) also including the Fourier terms to control for seasonality. P-values were calculated by comparing the full model (ARIMA) to M1 and M2 using the likelihood ratio test.

Forecasting Accuracy

Predicted values on validation sets were estimated using one-step forecast.[32] We estimated parameters only on training sets. However, we calculated forecasts on validation sets using all of the data preceding each observation. The accuracy of predictions was evaluated with the Mean Absolute Percentage Error (MAPE), which expresses, as percentages, a unit-free measure of performance:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} * 100$$

with y_t and $\hat{y}_t$ respectively the observed and the predicted number of visits at day t, and n the number of days in the validation set (n=365 in this study).

High demand definition

We proposed a definition of high ED demand as days where the number of visits exceeded the median of the preceding 31 days. The days were defined as green (level 1) if the number of visits exceeded the median by less than 5%, yellow (level 2) if between 5% and 10%, red (level 3) if higher than or equal to 10%. High demand was calculated on the observed and predicted ED visits in validation sets, we thus calculated the proportion of observed high ED demand that is correctly classified by predicted high ED demand (called sensitivity or recall metrics for multiclass classification problems).[33] In addition, we calculated the accuracy of predictions as the number of correct classifications over the total number of observations. All statistical analyses were performed with R (version 3.6.3).[34]

To evaluate the proposed definition, we further calculated high demand as: the number of visits exceeding the median of the preceding 7, 14, and 21 days and the number of visits exceeding the mean of the preceding 7, 14, 21, and 31 days, defining green, yellow, and red levels of high demand as above. We chose 7, 14, and 21 lag days in order to adjust for weekly variation in the number of ED visits by design. We further calculated high demand as defined by the Lombardy Region[35]: when the number of visits exceeded the 91st percentile of the previous year time-series. Low demand days were defined as those with a number of visits smaller than 25th percentile, medium demand days as those with a number of visits between 25th percentile and 75th percentile, high demand days if between 75th percentile and 90th percentile, and finally very high demand days if over 91th percentile.

ED warning system

In the month of January 2020, we established an ED warning system (WS), which was used by the selected hospitals in Milan as a planning instrument for EDs and consists in a transmission of daily reports. This WS continued until February when the COVID-19 outbreak started in Italy. According to the model choices highlighted by the above methodology (validation and calibration of the model were performed with data from 2014 to 2019), parameters were updated weekly and used to establish the WS which operated in January 2020. A hypothetical daily report received from a hospital on the 5th of January 2020 can be found in figure 2. The report included forecasts of the number of visits for the following two days, with 95% margin errors and a high demand indicator (green, yellow or red). The forecasts were made incorporating in the model past meteorological and environmental information via an Application Programming Interface (API) where 2-day future forecasts of meteorological and environmental information were provided by ARPA Lombardia. Weekly information on ILI was downloaded every week from InfluNet, and included in the predictive models. Daily reports were constructed and dispatched automatically using R and R Markdown. During the WS campaign, we established a monitoring service capable of estimating daily sensitivity, accuracy of predictions and MAPE separately for prediction one and two days ahead.

All analyses were performed with R software (V.4.0.2; R Core Team, Vienna, Austria), models and Fourier terms were estimated respectively, using the Arima and the Fourier functions in the R package forecast[36] using the parameter xreg for covariate specification. VIF was calculated using the VIF function in the car package.[37]

RESULTS

ED visit volumes

Between the 1st of January 2014 and the 31st of December 2019 (training set of 1,826 and validation set of 365 days) we observed 2,223,479 visits, 370,633 on average every year. Daily mean number of visits by hospital, temporal, meteorological, and patient characteristics in the training sets are summarized in table 1. Missingness, over the whole period 2014-2019, in meteorological and environmental variables were found in 8 days for temperature, 7 days for precipitation, and 37 days for humidity. Description of training and validation sets, and plots of each hospital's time series are summarized in the Supplementary Material (Supplementary Table 1). The Pearson correlation between predictors varied from weak (absolute correlation<0.3) to moderate (absolute correlation between 0.3 and 0.7), with a maximum of -0.67 between temperature and ILI and 0.61 between NO2 and PM10. VIF was smaller than 5 for all variables, with a maximum of 2.8 for temperature and 1.9 for ILI. We therefore included all the variables in the models, selecting the final model according to the statistical significance of predictors and minimal AIC.

Table 1

Total number of visits and mean number of daily visits by hospital, temporal and meteorological factors, and patient characteristics between the 1st of January 2014 and the 31st of December 2019 in five emergency departments of the city of Milan, Italy.

	N (%)[a]	Mean (Min-Max)[b]		N (%)[c]	Mean (Min-Max)[d]
Hospitals			**Cumulative Precipitation (mm)**		
A	421741 (19)	192 (107-301)	**≤ 0.6**	1678953 (75.5)	1018 (563-1295)
B	457021 (20.6)	209 (65-302)	**0.7+**	544526 (24.5)	1005 (627-1392)
C	272308 (12.2)	124 (61-197)	**NO2 (µg/m³)**		
D	530519 (23.9)	242 (125-337)	**≤ 32**	564957 (25.4)	974 (563-1295)
E	541890 (24.4)	247 (133-346)	**33-44**	570284 (25.6)	1022 (723-1292)
Total	2223479	1015 (563-1392)	**45-57**	536484 (24.1)	1032 (698-1392)
Gender			**58+**	551754 (24.8)	1035 (693-1272)
F	1113405 (50.6)	508 (277-782)	**PM10 (µg/m³)**		
M	1087903 (49.4)	497 (277-661)	**≤ 20**	571339 (25.7)	990 (563-1261)
Age			**21-29**	575530 (25.9)	1017 (688-1295)
≤ 14	360600 (16.4)	165 (55-443)	**30-44**	544147 (24.5)	1023 (710-1392)
15-65	1307139 (59.4)	597 (317-860)	**45+**	532463 (23.9)	1032 (693-1272)
66+	533569 (24.2)	244 (141-385)	**ILI (n. of weekly new cases per 1,000 inhabitants)**		
	N (%)[c]	**Mean (Min-Max)[d]**			
Temperature (°C)			**≤ 1.2**	1310096 (58.9)	1001 (563-1295)
≤ 9.2	564744 (25.4)	1021 (693-1392)	**1.3-2.5**	303072 (13.6)	1031 (799-1256)
9.3-15.6	563764 (25.4)	1033 (813-1261)	**2.6-5.6**	303102 (13.6)	1031 (698-1261)
15.7-22.3	563757 (25.4)	1025 (656-1295)	**5.7+**	307209 (13.8)	1045 (693-1392)
22.4+	531214 (23.9)	980 (563-1292)	**Day before/after festivity**		
Relative Humidity (%)			**No**	2096838 (94.3)	1012 (563-1392)
≤ 50	560870 (25.2)	1018 (563-1295)	**Yes**	126641 (5.7)	1055 (688-1295)
51-62	552865 (24.9)	1009 (637-1292)	**Festivity**		

63-76	554041 (24.9)	1017 (627-1392)	**No**	2144726 (96.5)	1018 (677-1392)
77+	555703 (25)	1016 (786-1278)	**Yes**	78753 (3.5)	938 (563-1253)
ILI=Influenza-like illness					

[a]Total number of visits by hospital, gender and age. In parenthesis the percentage of the number of visits out of the total (2,223,479 total number of visits, 2,201,308 with information on age and gender);
[b]Mean, minimum and maximum number of daily visits by hospital, gender and age;
[c]Total number of visits by temporal and meteorological factors (i.e. total number of visits in days with a particular value of temperature, humidity, etc.). In parenthesis the percentage of the number of visits of the total (2,223,479 total number of visits);
[d]Mean, minimum and maximum number of daily visits by temporal and meteorological factors (i.e. mean number of daily visits in the days with a particular value of temperature, Humidity etc.).

Model specification and ARIMA results

All models showed a very strong day of the week and year-round seasonality effect, according to ACF and PACF plots. To normalize residuals, outliers (in the training sets only) were replaced by the mean of the observations of the same day in the other years, consequently all models showed residual normally distributed according to the Jarque-Bera test (number of replaced outliers are presented in Supplementary Table 2). All models showed a lack of fitting on New Year's Eve and/or August 15th, for this reason we chose to define a specific dichotomous variable ("1" for the peculiar festivity, "0" for the other days) capable of detecting this extra variation. Table 2 displays the ARIMA parameters fitted for each model, and the number of Fourier terms that minimized AIC. All models were non-stationary in mean and needed one differencing to make the time series stationary (d=1). ARIMA parameters and Fourier terms were different across hospitals, showing that each time series needed different model specification. Table 2 also displays, for each hospital, the factors that significantly influenced the number of ED visits, and that were included in the models. High temperatures were always associated with a statistically significant increase in ED visit volumes, with a maximum increase of 1.84 daily visits every 1°C increase (hospital E, s.e. 0.18). Relative humidity was significantly associated with a limited decrease of total ED visits (-0.08, s.e. 0.04) for a 1% increment of RH only at hospital D. High levels of cumulative precipitation were associated (except for hospital C) with a statistically significant decrease in ED visits, with a maximum decrease of 0.31 daily visits every 1 mm of precipitation (hospital E, s.e. 0.06). Concerning air pollution, we found an opposite effect of NO_2 and PM10 on ED visits, with a mild significant negative effect for NO_2 in two hospitals (-0.08 and -0.09) and an even milder positive association with PM10 in one (0.03). Except for hospital C, the effect of ILI was always associated with the number of ED visits, showing an increase of daily visits between 0.73 and 1.74 (s.e. 0.29 and 0.41 respectively) at every

unit increase in weekly ILI rates. Festivities were associated with a decrease in ED visits of between 13 and 28 (s.e. 1.45 and 1.98), while special festivities were associated with the greatest decrease of at least 42 ED visits (s.e. 4.94). Autocorrelation function and correlation among residuals according to the Ljung-Box test by hospital and up to 30 and 366 lags can be found in Supplementary figure 1. ACF plots of residuals were overall in significance limits and the Ljung-Box test showed overall no significant correlation between residuals at different lags, except Hospital E which showed residual autocorrelation up to lag 366.

Table 2

Auto-regressive integrated moving average (ARIMA) specifications and covariate effects on the number of ED visits between the 1st of January 2014 and the 31st of December 2018 (training sets).

		Hospitals				
		A	B	C	D	E
Model specification	ARIMA parameters (p,d,q)	(0,1,2)	(1,1,1)	(1,1,2)	(1,1,1)	(1,1,1)
	Fourier terms†	3,13	3,14	3,13	3,16	3,15
Covariate Effects (se)[1]	Temperature (°C)	1.29 (0.15)	1.23 (0.14)	0.68 (0.11)	1.16 (0.18)	1.84 (0.18)
	Humidity (%)				-0.08 (0.04)	
	Precipitation (mm)	-0.2 (0.05)	-0.12 (0.05)		-0.13 (0.07)	-0.31 (0.06)
	NO2 (µg/m3)	-0.08 (0.03)			-0.09 (0.04)	
	PM10 (µg/m3)			0.03 (0.02)		
	ILI (weekly new cases per 1,000 inhabitants)	1.74 (0.41)	1.05 (0.37)	0.73 (0.29)		0.97 (0.46)
	Festivity		-28.23 (1.98)	-12.96 (1.45)	-25.42 (2.23)	-14.56 (2.39)
	Special Festivity*	-43.16 (6.31)	-57.64 (6.36); -62.61 (6.29)	-42.06 (4.92)	-59.86 (7.58)	-63.24 (7.92)
	Day before/after festivity	7.14 (1.5)	9.06 (1.58)	3.75 (1.22)		13.89 (1.96)

[1]Parameter estimates and standard errors in parentheses. Predictors were retained in the final model only if statistically significant (p-value<0.05)
† Number of sine and cosine terms used to approximate day of the week and year-round seasonality
***New Year's Eve for hospitals A, C-D and New Year's Eve and August 15th for hospital B**
ARIMA =Auto-regressive integrated moving average
ILI=Influenza-Like-Illness

Forecasting Accuracy and High demand definition

The accuracy of predictions (MAPE) in the validation sets, sensitivity and accuracy between observed and predicted high ED demand are displayed in table 3. Model performance was good, with small MAPEs in validation sets, ranging from a minimum of 5.5% for hospital D to a maximum of 8.1% for hospital C. The models showed high sensitivity on days with green-level high demand, almost 90% of days with predicted green-level high demand were confirmed from observed values.

On days with yellow-level high demand, sensitivity between predicted and observed demand was scarce, ranging from 0.04 for hospital B to 0.28 for hospital A. Sensitivity of red-level high demand varied between hospitals, with a minimum of 0.25 for hospital A to a maximum of 0.57 for hospital D. Observing Table 3 we can suggest that, for each hospital, at least 54% of the observed red-level high demand days were classified, from predictions, as being at least yellow-level. Accuracy was high, with at least 67% of the days with exactly the same predicted and observed high demand level (green, yellow or red).

All ARIMA models fitted the data significantly better than a simple regression model (M1) and a generalized linear model (M2), with MAPE for M1 and M2 above 13.5% and 9.8%, respectively (Supplementary Table 3). Observed and predicted ED visits in the validation sets (from the 1th of January 2019 to the 31th of December 2019) by date and hospital can be found in Supplementary Figure 2.In Supplementary Table 2 we compared ARIMA results for different temperature specifications: mean, minimum, maximum and apparent temperature. The greatest effect on ED visits was attributed to mean temperature while indicators of performance and AIC where generally superior for mean temperature compared with minimum, maximum and apparent temperature. In Supplementary Table 2 we also calculated, only for outlier days, the relative error mean of observed vs predicted values in order to evaluate if extreme temperatures were better outlier predictors than mean temperature. Number of outliers replaced ranged from 2 for hospital A to 7 for hospital D, results suggested an overall better fit of outliers using minimum temperature (3 out 5 hospitals with smaller relative errors).

In supplementary table 4 we compared the high demand definition used in the ED warning system with similar definitions. There was slight improvement in percentage accuracy between the definition used and the other algorithms and there was no favourite algorithm for all hospitals: hospital B had a maximum improvement of 4% using the mean of the preceding 31 days or the median of the proceeding 21 days, hospitals A and C had an improvement of 2% using the mean of the preceding 31 days, hospital D had an improvement of 2% using the mean of the preceding 21 days, and finally hospital E had an improvement of 1% using the mean of the preceding 21 or 31 days. Using the high demand definition used by the Lombardy Region we did not find any improvement in accuracy, with an overall percentage of matched classification between 50% and 64%. High demand was always predicted less well compared to the definition used in our ED warning system. However, results showed good prediction of very high demand days with a sensitivity between 38% and 67%.

Table 3

Indicators of performance of the developed models: accuracy of predictions (MAPE) in the validation sets, and accuracy and sensitivity of high demand classification.

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity)		
				Green	Yellow	Red
Hospital A	5.9	72	Green	93	6	1
			Yellow	64	28	8
			Red	46	29	25
Hospital B	5.7	72	Green	92	8	0
			Yellow	85	4	11
			Red	35	15	50
Hospital C	8.1	67	Green	88	8	4
			Yellow	78	10	12
			Red	45	20	35
Hospital D	5.5	76	Green	91	6	3
			Yellow	65	27	8
			Red	35	9	56
Hospital E	6.1	74	Green	90	8	2
			Yellow	59	24	17
			Red	34	28	38

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

ED warning system

In Table 4a and 4b we provided the accuracy of predictions (MAPE), sensitivity and accuracy between observed and predicted high ED demand in January (the operating period of the WS) for one- and two-days horizons. Errors of prediction (MAPE) were slightly higher than in the validation set, with MAPE for one-day horizon always smaller than MAPE for two-days horizons. Accuracy between observed and predicted high ED demand was never smaller than 0.45 and generally smaller than in the validation set.

Table 4

Accuracy of predictions (MAPE), sensitivity, and accuracy between observed and predicted high ED demand in January 2020 (the operating period of the WS) with a one- (4a) and two-day (4b) horizon.

4a

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	7.8	52	Green	94	6	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	7.8	81	Green	87	13	0
			Yellow	0	100	0
			Red	17	17	67
Hospital C	8.6	52	Green	100	0	0
			Yellow	67	33	0
			Red	73	27	0
Hospital D	6.6	45	Green	55	36	9
			Yellow	0	33	67
			Red	50	33	17
Hospital E	11	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

4b

	MAPE	Accuracy (%)	Observed high EDdemand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	8.1	55	Green	100	0	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	8.6	71	Green	73	27	0
			Yellow	0	100	0
			Red	25	17	58
Hospital C	9	45	Green	93	7	0
			Yellow	83	17	0
			Red	82	18	0
Hospital D	7.6	48	Green	50	18	32
			Yellow	0	0	100
			Red	33	0	67
Hospital E	11.2	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

DISCUSSION

In this work we proposed and implemented in daily practice, a system to predict the number of ED visits in five hospitals of the city of Milan. The system is based on regression models with ARIMA errors, where ARIMA parameters were allowed to vary between hospitals, according to their specific characteristics, and it provides daily reports on the number of visits predicted for the two subsequent days at the five hospitals participating in the study. The models showed a good overall performance with the MAPEs always smaller than 5.5% and 8.1%. Our results are slightly better than other studies: Marcilio and colleagues[20] forecast daily ED visits with Generalized Linear Models, finding MAPEs between 5.4% and 11.5%, according to different forecasting horizons and controlling for temperature effect.

Jones and colleagues,[16] using similar models, found MAPEs that varied between 8.5% and 15.5%. However, Duwalage et al.[15] using a Generalized Additive Model found MAPEs consistently lower than 5% for 14-day forecasts, which significantly improved including temperature in the model. Although the number of predicted ED visits was close to the observed values, and there was good sensitivity in predicting mild (green) high demand, there was moderate sensitivity in predicting the spike of ED visit volumes (red-level high demand) for some hospitals and acceptable sensitivity for hospital D. This is particularly important for the scope of this study, which aimed to forecast emergency department visits in order to develop a 2-day warning system. For this reason, a better predictive performance of the red-level forecast would be desired. In fact, one of the major reasons for ED overcrowding is the shortage of acute care bed capacity compared with the huge number of visiting patients. Comparing our definition with similar definitions, we found a slight improvement in percentage accuracy, around 1% and 4%, but there was no a favourite algorithm for all hospitals. Furthermore, using the definition of very high demand for ED visits defined by the Lombardy Region, we found sensitivity was better compared to our models, and we plan to implement this in further evolutions of our warning system. However, we found good sensitivity in classifying observed red-level demand as at least yellow from predictions, and accuracy among observed and predicted high demand levels was always close to 70%. The definition of high ED demand is not straightforward as it relies on the specific hospital’s characteristics. It is one of the main causes of ED overcrowding, which is the most problematic issue in EDs, thus deserving the effort in trying to predict it. In this study, we proposed a definition based on percentage increases compared to the median of the preceding month, to warn EDs of requests rising over the levels they managed in the preceding month.

During the operating period of the warning system, January 2020, we found a worse adaptation of the models than in the validation year 2019. This could be due to the ongoing outbreak of COVID-19, as ED visits for non-critical problems were discouraged.[38]

Concerning potential predictors, we found a strong day of the week and year-round seasonality effect, adequately captured by the terms used to approximate periodicity (Fourier terms). Even if the aim of this work was to develop a forecasting model and not an explanatory model, here we found statistically significant effects of meteorological factors on ED visits. Temperature was always positively associated with outcome, with an increase in the number of visits for each 1-degree increase in temperature across hospitals, in accordance with previous results.[17,19,23] As reported in another study,[39] high temperatures are associated with ED visits, especially for the most susceptible population, as persons with diabetes or cancer, so it is important for public health officials to implement adaptation measures to manage the impact of high temperatures on population health. Here we found a slightly better fit for outliers using minimum temperature instead of mean temperature. Nonetheless, we decided to include mean temperature in the ED warning system because it showed the greatest effect on ED visits. Further work has to be done in order to investigate the role of extreme temperature on ED visit fluctuations. The role of precipitations has not yet been well established. To our knowledge only one study measured an indirect effect in reducing ED visit volumes.[19] In accordance with these results, rainy days were found to be mildly associated with reduced numbers of ED visits. NO2 and PM10 had a mild significant effect only in two hospitals and in one hospital respectively, and were discordant, with a negative effect of NO2 and a positive effect of PM10 on the number of ED visits. This may be explained considering that the effect of pollution on ED visits is generally exerted and measured on respiratory conditions, especially asthma, and/or cardiac rather than with total visits and it may be diluted when analysing all ED visits. Only a few studies found a positive association of Total Suspended Particles with all visits but trauma, going in the same direction as the small significant increase in the number of visits related to PM10 we found.[40] In addition, pollution estimated from the monitoring station (classified as from urban traffic) used in the analysis might be of a greater magnitude than that really observed in each hospital. However, even though the hospitals were mostly located on the outskirts of the city of Milan, they are all located in urban areas characterized by a similar air pollution pattern. ILI were found to significantly increase the number of ED visits, as found by other researchers.[41]

This study indicated a moderate to good sensitivity in predicting high demand, showing some difficulties in anticipating the exact red-level days. In the future we aim to investigate models capable to directly predicting ED peaks instead of predicting the number of ED visits such as copulas used for detecting spikes in signal processing in brain circuits[42] or machine learning models. Finally, when interpreting these results, it is necessary to be aware of the possible multicollinearity problem between variables, which may alter the magnitude and statistical significance of coefficients. However, according to Vatcheva 2016,[31] only high correlations between variables would result in a change of sign of the coefficients and furthermore VIFs were always smaller than 5. Correlated factors were the pollution variables (NO_2 and PM_{10}), which were never considered in the same model together. Given that the highest correlation was found among temperature and ILI, the effect of these variables on the number of ED visits may potentially be biased due to multicollinearity. However, we included both terms in the models given the fact that they have an effect on ED visits independently from one another.

Another limitation is the choice of the hospitals considered for this work, that is, major hospitals located in the city of Milan. This methodology might not be the feasible for use by small hospitals as they might have low counts or even no visits at all on particular days. A solution can be provided by implementing different statistical models, for example, negative binomial or zero-inflated Poisson models, and would be one of our aims in the next years.

High-demand ED forecasting has a dual nature that should be addressed: first, knowing in advance the number of expected visits would allow a more reasoned choice of the hospital to which request assistance and second, forecasts should be follow immediately by an evaluation of the available beds and of the staff needed to accommodate these expected visits. These two problems were not addressed in this work given that this study was intended to estimate ED demand only and does not include information on hospital capacity, but are fundamental ingredients that should be considered in the future.

In conclusion, we proposed a hospital specific ED warning system based on predictive models developed on previous attendances that can be used as a planning instrument in hospitals to increase resources, and to prevent high patient demand when a higher number of attendances is expected. This is important in any health system that usually deals with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits are still high.

List of abbreviations

ACF: autocorrelation function

AHP: Milan agency for health protection

AIC: minimal Akaike information criteria

API: application programming interface

AR: auto-regression

ARIMA: Auto-regressive integrated moving average

ED: emergency department

ILI: Influenza-like-illness

MA: moving average process

MAPE: mean absolute percentage error

NO2: nitrogen dioxide

PM10: particulate matter with a diameter ≤ 10 µm

PACF: partial autocorrelation function

RH: relative humidity

WS: warning system

Ethics approval and consent to participate This study does not involve human participants or animal subjects. Ethics approval and consent to participate were not required, as this is an observational study based on data routinely collected by the Agency for Health Protection (ATS) of Milan, a public body of the Regional Health Service-Lombardy Region. Among the institutional functions of ATS, established by Lombardy regional legislation (R.L. 23/2015), is management of the care pathway at the individual level in the regional healthcare system and evaluation of the services provided to, and the outcomes of, patients residing in the covered area. This study is also ethically compliant with Italian National Law (D.Lgs. 101/2018) and the "General Authorisation to Process Personal Data for Scientific Research Purposes" (nos. 8 and 9/2016, referred to in the Data Protection Authority action of 13/12/2018). Data were anonymized with a unique identifier in the different datasets before being used for the analyses.

Contributors RM and AGR conceptualised the study and defined the methodology, accessed and verified the data. RM analysed the data set, ST and AA contributed to the literature search, data interpretation, and writing of the manuscript. ST and AA, have made substantial contributions to the revision of the paper. AGR supervised and managed the project.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Data availability statement Data are not publicly available due to property restrictions.

Word count 4862

Acknowledgement Authors thank ARPA Lombardia, in particular Orietta Cazzuli and Gian Paolo Minardi, for their valuable contribution.

REFERENCES

1 Pines JM, Hilton JA, Weber EJ, *et al.* International perspectives on emergency department crowding. *Acad Emerg Med* 2011; **18**: 1358–70.

2 Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. *Emergency medicine journal : EMJ* 2009. DOI:10.1136/emj.2008.062380.

3 Programma Nazionale Esiti - PNE 2018 - Home. https://pne.agenas.it/index.php (accessed Feb 18, 2020).

4 Hwang U, McCarthy ML, Aronsky D, *et al.* Measures of crowding in the emergency department: a systematic review. *Acad Emerg Med* 2011; **18**: 527–38.

5 Asplin BR. Measuring Crowding: Time for a Paradigm Shift. *Academic Emergency Medicine* 2006; **13**: 459–61.

6 Wang Z, Xiong X, Wang S, Yan J, Springer M, Dellinger RP. Causes of Emergency Department Overcrowding and Blockage of Access to Critical Services in Beijing: A 2-Year Study. *The Journal of Emergency Medicine* 2018; **54**: 665–73.

7 Jeyaraman MM, Copstein L, Al-Yousif N, *et al.* Interventions and strategies involving primary healthcare professionals to manage emergency department overcrowding: a scoping review. *BMJ Open* 2021; **11**: e048613.

8 Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emergency Medicine Journal* 2003; **20**: 402–5.

9 McKenna P, Heslin SM, Viccellio P, Mallon WK, Hernandez C, Morley EJ. Emergency department and hospital crowding: causes, consequences, and cures. *Clin Exp Emerg Med* 2019; **6**: 189–95.

10 Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. *Ann Emerg Med* 2008; **52**: 126–36.

11 Howard M, Davis B, Anderson C, Cherry D, Koller P, Shelton D. Patients' perspective on choosing the emergency department for nonurgent medical care: a qualitative study exploring one reason for overcrowding. *Journal of emergency nursing: JEN : official publication of the Emergency Department Nurses Association* 2005. DOI:10.1016/J.JEN.2005.06.023.

12 Angelillo I, Ricciardi G, Nante N, *et al.* Appropriateness of hospital utilisation in Italy. *Public Health* 2000; **114**: 9–14.

13 Hakkoymaz H. How Appropriate are The Emergency Department Admissions? A Study From a Tertiary Hospital Emergency Department. *EJMI* 2020. DOI:10.14744/ejmi.2020.47376.

14 Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med* 2010; **17**: 970–8.

15 Duwalage KI, Burkett E, White G, Wong A, Thompson MH. Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas* 2020; published online Feb 17. DOI:10.1111/1742-6723.13481.

16 Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008; **15**: 159–70.

17 Rising WR, O'Daniel JA, Roberts CS. Correlating Weather and Trauma Admissions at a Level I Trauma Center. *Journal of Trauma and Acute Care Surgery* 2006; **60**: 1096–100.

18 Holleman DR, Bowling RL, Gathy C. Predicting daily visits to a waik-in clinic and emergency department using calendar and weather data. *J Gen Intern Med* 1996; **11**: 237–9.

19 Diehl AK, Morris MD, Mannis SA. Use of calendar and weather data to predict walk-in attendance. *South Med J* 1981; **74**: 709–12.

20 Marcilio I, Hajat S, Gouveia N. Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine* 2013; **20**: 769–77.

21 McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med* 2008; **15**: 337–46.

22 Díaz J, Alberdi JC, Pajares MS, *et al.* A model for forecasting emergency hospital admissions: effect of environmental variables. *J Environ Health* 2001; **64**: 9–15.

23 Macgregor DM. Effect of weather on attendance with injury at a paediatric emergency department. *Emergency Medicine Journal* 2003; **20**: 204–5.

24 Wargon M, Brun-Ney D, Beaujouan L, Casalino E. No more winter crisis? Forecasting daily bed requirements for emergency department admissions to hospital. *Eur J Emerg Med* 2018; **25**: 250–6.

25 Murtas R, Decarli A, Russo AG. Trend of pneumonia diagnosis in emergency departments as a COVID-19 surveillance system: a time series study. *BMJ Open* 2021; **11**: e044388.

26 Morbey RA, Charlett A, Lake I, *et al.* Can syndromic surveillance help forecast winter hospital bed pressures in England? *PLoS ONE* 2020; **15**: e0228804.

27 Italia. Provvedimento del Garante n. 2 del 16 giugno 2004. Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici. .

28 Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. ARPA Lombardia. https://www.arpalombardia.it:443/Pages/ARPA_Home_Page.aspx (accessed March 10, 2020).

29 Steadman RG. Norms of apparent temperature in Australia. *Aust Met Mag* 1994; **43**: 1–16.

30 Ministero della Salute. Influenza. http://www.salute.gov.it/portale/influenza/homeInfluenza.jsp (accessed March 10, 2020).

31 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* 2016; **6**. DOI:10.4172/2161-1165.1000227.

32 Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018.

33 M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *IJDKP* 2015; **5**: 01–11.

34 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020 https://www.R-project.org/.

35 Salute M della. Pronto soccorso, via libera della Conferenza Stato-Regioni alle nuove linee di indirizzo. http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=3849 (accessed April 24, 2020).

36 Hyndman [aut R, cre, cph, *et al.* forecast: Forecasting Functions for Time Series and Linear Models. 2020 https://CRAN.R-project.org/package=forecast (accessed Nov 14, 2020).

37 Fox J, Weisberg S. An R Companion to Applied Regression. SAGE Publications, 2018.

38 Jeffery MM, D'Onofrio G, Paek H, *et al.* Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. *JAMA Intern Med* 2020; published online Aug 3. DOI:10.1001/jamainternmed.2020.3288.

39 Lavigne E, Gasparrini A, Wang X, *et al.* Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environ Health* 2014; **13**: 5.

40 Samet JM, Speizer FE, Bishop Y, Spengler JD, Ferris BG. The Relationship between Air Pollution and Emergency Room Visits in an Industrial Community. *Journal of the Air Pollution Control Association* 1981; **31**: 236–40.

41 Schanzer DL, Schwartz B. Impact of seasonal and pandemic influenza on emergency department visits, 2003-2010, Ontario, Canada. *Acad Emerg Med* 2013; **20**: 388–97.

42 Hu M, Li M, Li W, Liang H. Joint analysis of spikes and local field potentials using copula. *NeuroImage* 2016; **133**: 457–67.

Figure 1

Location of the five partecipating hospitals and of meteorological and air pollution monitoring stations in the city of Milan.

Figure 2

Hypothetical daily report received from a hospital on the 5th of January 2020.

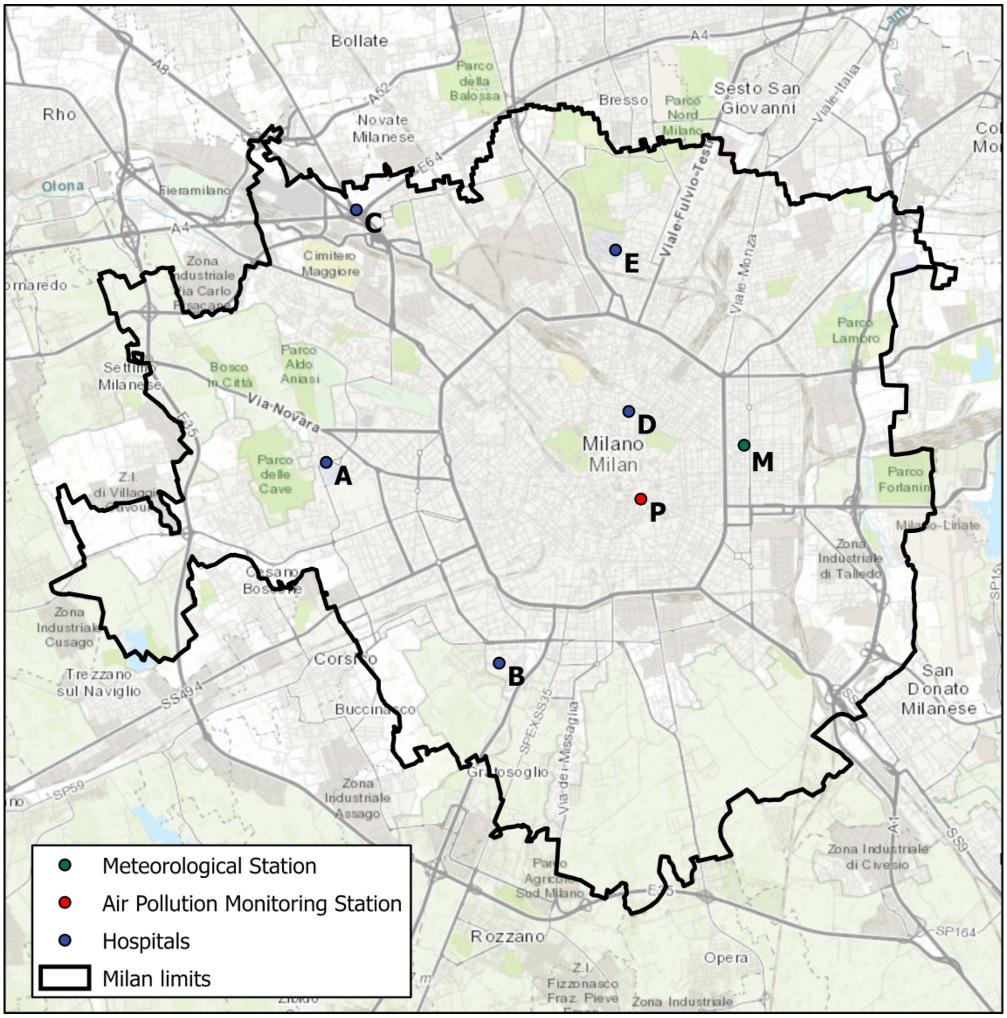


Figure 1. Location of the five considered hospitals and of meteorological and air pollution monitoring station in the city of Milan.

160x161mm (600 x 600 DPI)

EMERGENCY ADMISSION 2-DAY WARNING SYSTEM

Agency for Health Protection of Milan

Hospital A. Prediction for the day:

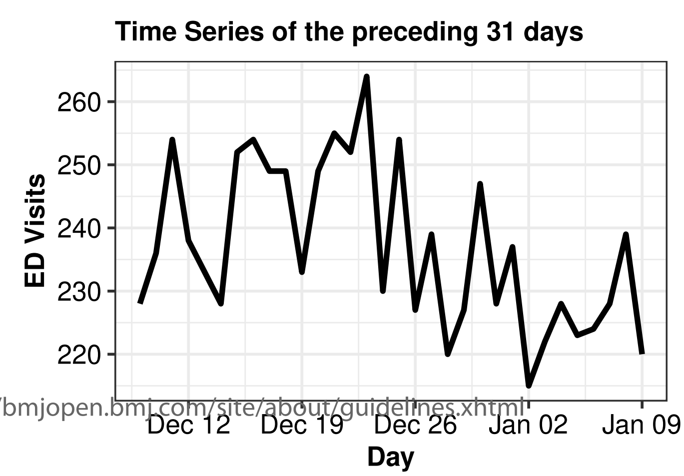
11 Jan 2020	12 Jan 2020
Level 1	Level 2

	11 Jan 2020	12 Jan 2020
Predicted ED visits ± 95% Margin Errors	227 ± 12	254 ± 14
Covariates		
NO2	$46\ \mu g/m^3$	$38.4\ \mu g/m^3$
PM10	$6\ \mu g/m^3$	$3.9\ \mu g/m^3$
TEMPERATURE	5°	6°
UMIDITY	65%	57%
PRECIPITATION	0 mm	0 mm
ILI RATE	10.87	10.87

Methodology

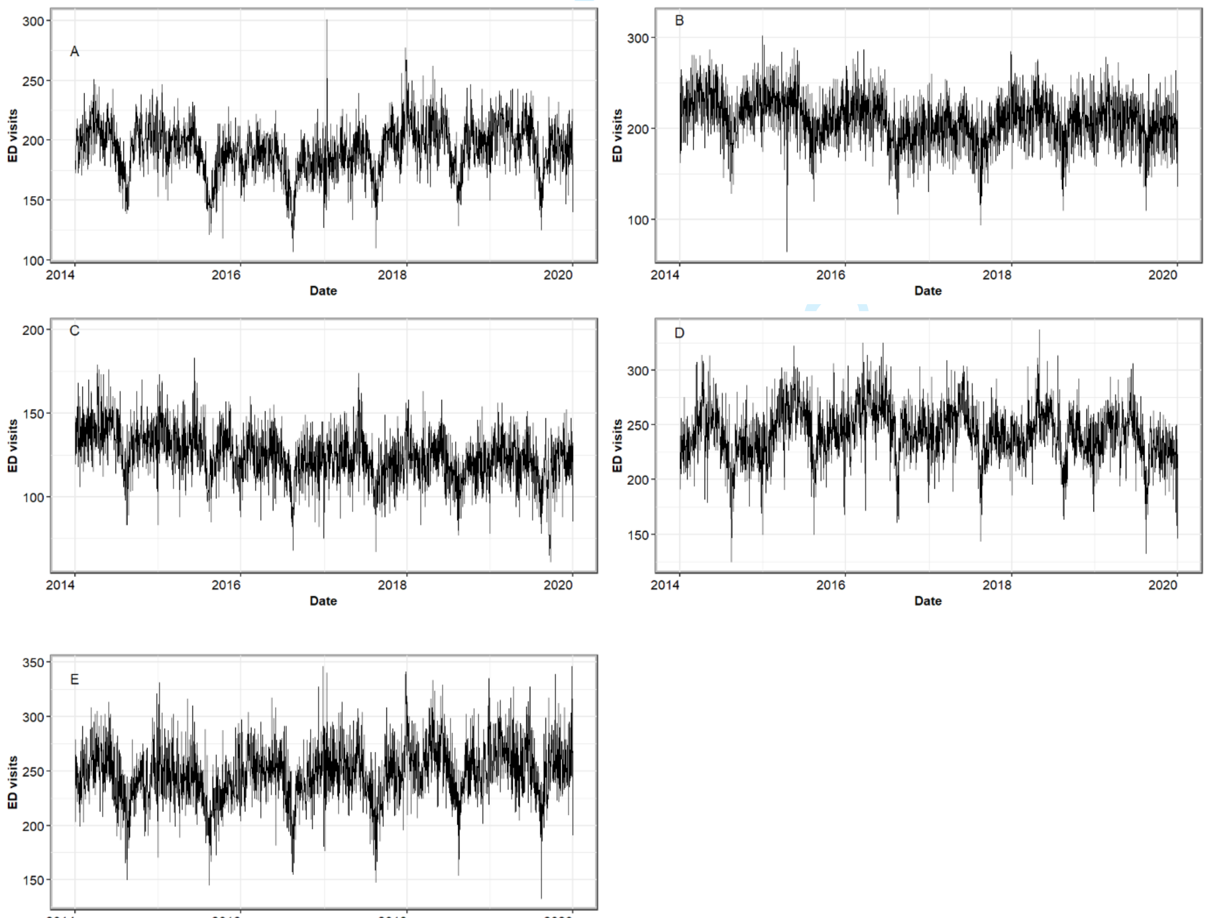
Level 1	Number of visits exceeded the median by less than 5%.
Level 2	Number of visits exceeded the median between 5% and 10%.
Level 3	Number of visits exceeded the median by more than 10%.
Prediction	Prediction based on a regression model with ARIMA errors (Hyndman 2018).
Environment	Meteo and Pollution Forecast from ARPA Lombardia.
ILI rates	Influence Like Illness rate, number of cases per 1,000 inhabitants per week (InfluNet).

Sistema Socio Sanitario
Regione Lombardia
ATS Milano
Città Metropolitana



Description of Training and validation sets

Patients were mostly female (50.6%), with a mean age of 44 years (standard deviation s.d. of 26 years). The mean number of daily visits was similar in genders. Higher ED accesses volumes were found among people aged over 65 years than the others ages. During the study period, mean temperature was 16°C, mean RH was 63%, and there was precipitation on 31% of the days. Mean level of NO2 exceeded the European limit (40 μg/m^3), with a mean of 46 μg/m^3, while mean PM10 was lower than the European limit, with a mean of 35 μg/m^3, during the observed period. On days-before and after festivities, we measured a higher number of visits, while on festivity days there was a lower number of visits compare to normal days. Training and validation sets were similar according to meteorological factors, but there were mild differences in air pollution and ILI rates (supplementary Table 1). The year 2019 was in fact characterized by significantly higher levels of pollution (t-test p-values<0.001) and lower ILI rates (t-test p-value<0.01) compared to the previous years. Patients were slightly younger in the training set than in the validation set, the mean age being 43 years in the former and 45 in the latter. The number of ED accesses was statistically different across age groups between days: children (0-14 years) tended to visit ED more likely on weekends (20% higher on Sundays compared to other days) while adults and senior people (15-65 and >65 years) on Mondays (14% and 11% higher than other days) (Anova test for mean differences p-values<0.001). August was the months with smaller ED accesses volumes, with a 14% decrease compared to the average of the other months.



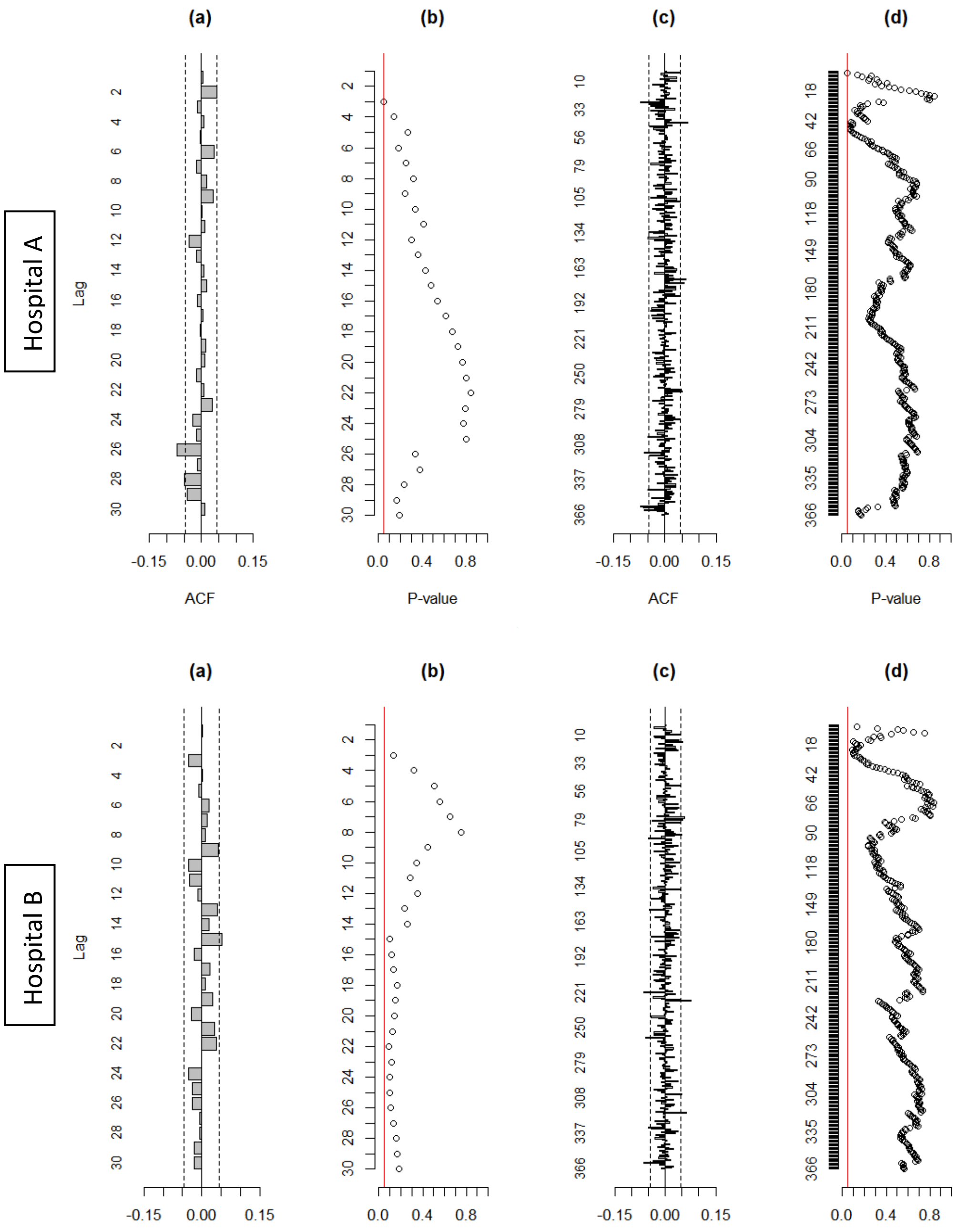
Supplementary Table 1

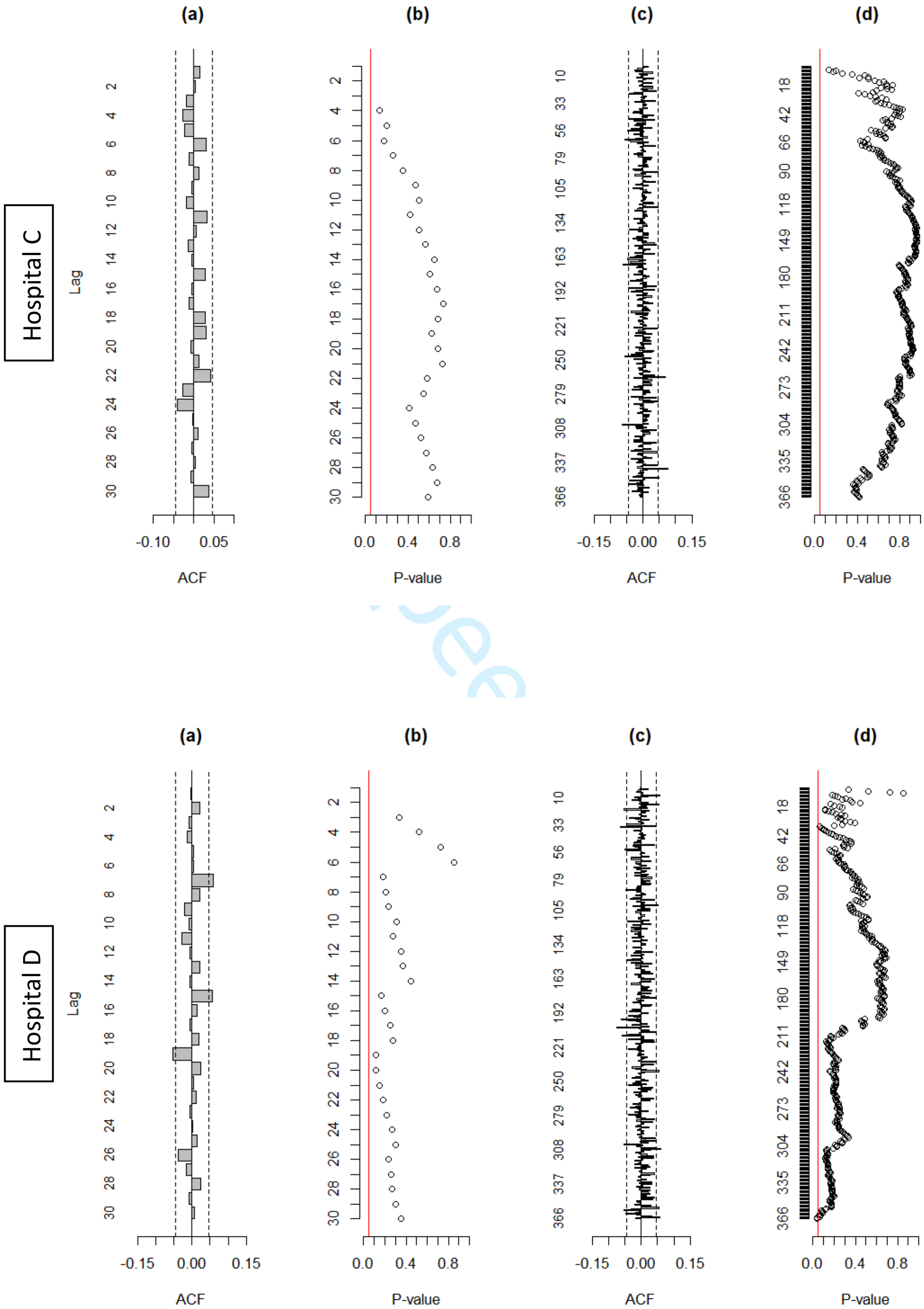
Mean, standard deviation and t-test for mean difference between training and validation sets by covariates.

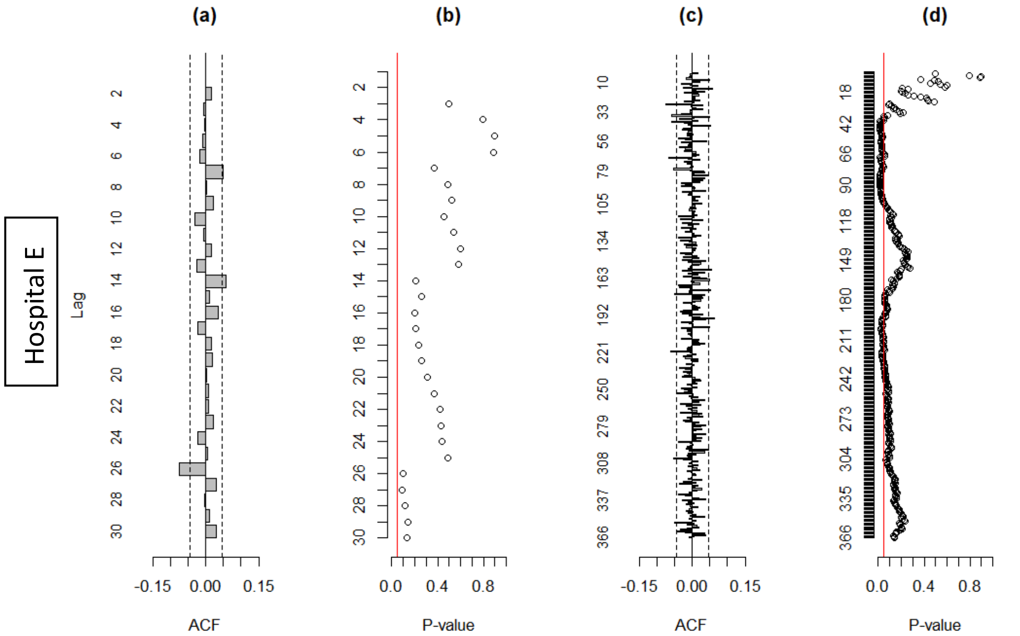
	Mean (s.d.)		t-test p-value
	Training	Validation	
Female (%)	50.6	50.4	-
Age (years)	43 (26)	45 (26)	<0.001
Temperature (°C)	15.7 (8)	16.4 (8)	0.1257
Relative Humidity (%)	63.4 (17)	62.3 (17)	0.2765
Cumulative Precipitation (mm)	2.3 (6.6)	2.2 (5.8)	0.7155
NO2 (µg/m3)	47 (19)	39 (17)	<0.001
PM10 (µg/m3)	36 (21)	30 (18)	<0.001
ILI (new cases per 1,000 inhabitants)	1.9 (3)	2.6 (3.8)	<0.001
S.d.=standard deviation ILI=Influenza-Like-Illness			

Supplementary Figure 1

Autocorrelation function (ACF) and correlation among residuals according to the Ljung-Box test (LB) by hospital: (a) ACF up to lag 30; (b) LB test up to lag 30; (c) ACF up to lag 366; (d) LB test up to lag 366.




Hospital C
(a)
(b)
(c)
(d)
Lag
ACF
P-value
ACF
P-value
Hospital D
(a)
(b)
(c)
(d)
Lag
ACF
P-value
ACF
P-value


Hospital E
(a)
(b)
(c)
(d)
Lag
ACF
P-value
ACF
P-value

Supplementary Table 2

Auto-regressive integrated moving average (ARIMA) parameters, indicators of performance (accuracy of predictions in the validation sets, and accuracy of high demand classification), Akaike Information Criteria (AIC) and relative error mean for outliers.

		Mean temperature					Minimum Temperature					Maximum Temp					Apparent Temperature				
	Number of outliers' days*	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative error mean**	AIC	Beta (se)	MAPE	Accuracy (%)	Relative Error mean*	AIC
Hospital A	2	1.29 (0.15)	5.9	72	55.3	14861	1.13 (0.15)	5.9	73	55.1	14882	1.03 (0.12)	5.9	72	55.1	14864	1.03 (0.13)	5.9	72	55.5	14870
Hospital B	6	1.23 (0.14)	5.7	72	44.7	14847	1.02 (0.15)	5.7	74	44.1	14874	0.98 (0.11)	5.7	73	45.6	14850	1.04 (0.12)	5.7	73	44.9	14846
Hospital C	5	0.68 (0.11)	8.1	67	36.5	13928	0.57 (0.11)	8.1	67	36.8	13941	0.55 (0.09)	8.1	66	36.5	13925	0.54 (0.09)	8.1	66	36.8	13934
Hospital D	7	1.16 (0.18)	5.5	76	22.3	15506	1.00 (0.18)	5.6	76	22.6	15516	0.96 (0.15)	5.6	75	22.4	15508	0.88 (0.15)	5.5	76	22.3	15511
Hospital E	6	1.84 (0.18)	6.1	74	24.4	15624	1.68 (0.19)	6.1	73	24	15649	1.5 (0.14)	6.1	73	24.9	15622	1.53 (0.16)	6.3	71	24.2	15811

*Number of outliers' days replaced by the mean of the observations of the same day in the other years for normalization of results

**Relative error mean of observed vs predicted values calculated for outliers (in the training sets only) which were replaced by the mean of the observations of the same day in the other years for normalization of residuals

Supplementary Table 3

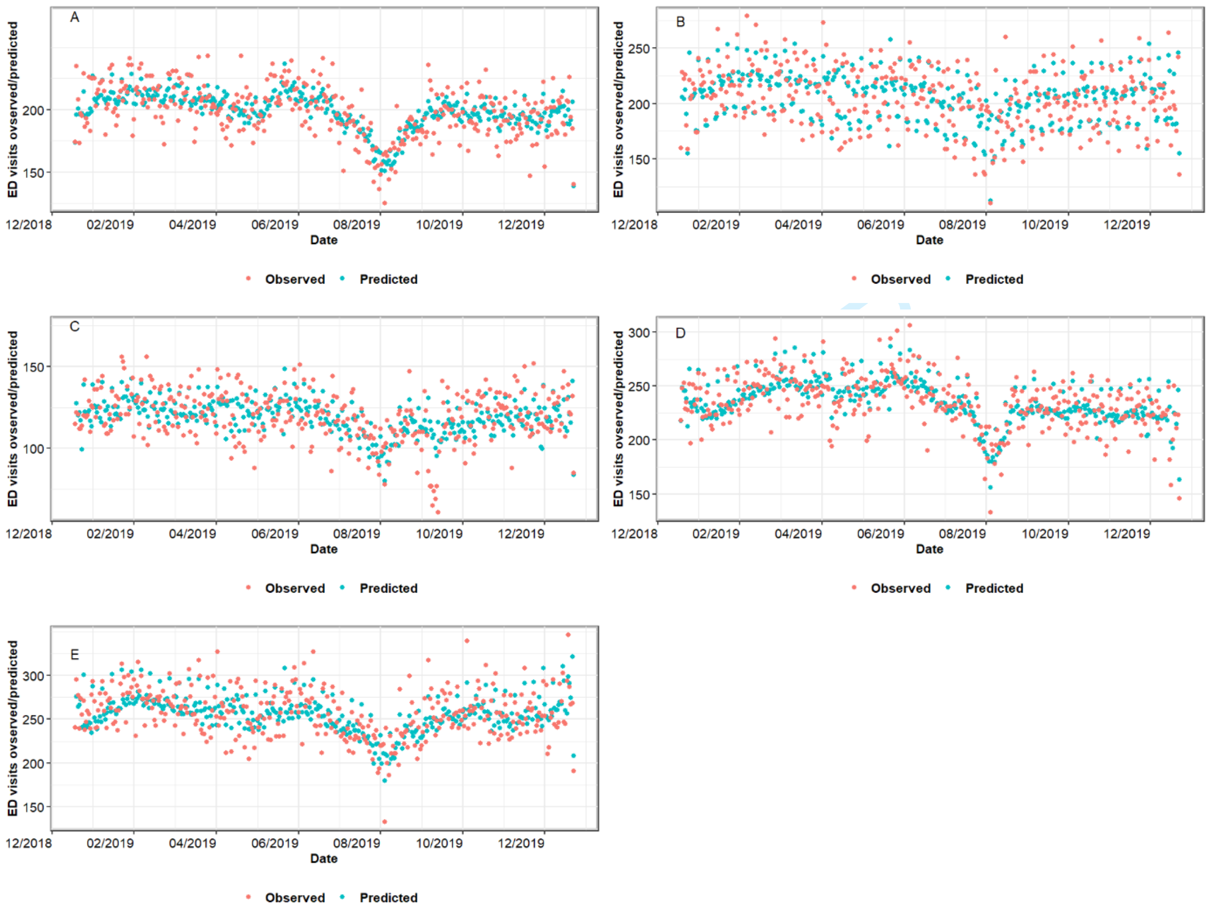
Model comparisons between the regression model with ARIMA errors and: a simple regression model (M1) and a generalized linear model (M2).

	Linear model (M1[a])		Likelihood ratio test*		Generalized linear model (M2[b])		Likelihood ratio test**	
	MAPE	Accuracy (%)	LRT (df)	p-value	MAPE	Accuracy (%)	LRT (df)	p-value
Hospital A	17.4	52	3957 (34)	<0.001	11.9	56	2618 (2)	<0.001
Hospital B	25.3	49	5610 (36)	<0.001	13.2	60	3318 (2)	<0.001
Hospital C	20.8	51	3975 (35)	<0.001	14.4	58	2662 (3)	<0.001
Hospital D	13.5	51	3289 (40)	<0.001	9.8	57	2164 (2)	<0.001
Hospital E	23.9	55	5222 (38)	<0.001	12.3	57	2909 (2)	<0.001

[a]M1: linear model with only meteorological, environmental and festivities covariates
[b]M2: generalized linear model with meteorological, environmental, festivities covariates and Fourier terms to control for seasonality
*Likelihood ratio test (LRT) comparing the regression model with ARIMA errors with M1
**Likelihood ratio test (LRT) comparing the regression model with ARIMA errors with M2

Supplementary Figure 2

Observed and predicted ED visits in the validation sets (from the 1th of January 2019 to the 31th of December 2019) by date and hospital.



Supplementary Table 4

Indicators of performance (accuracy and sensitivity of high demand classification) by different definition of high demand: the number of visits exceeded the median of the preceding 7, 14 and 21 days (4a), the number of visits exceeded the mean of the preceding 7, 14, 21 and 31 days (4b), and high demand defined by the Lombardy Region as exceeding thresholds based on previous year percentiles (4c).

4a	Median of the preceding 7 days					Median of the preceding 14 days					Median of the preceding 21 days				
			Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)		
	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red
Hospital A	72	Green	95	5	0	73	Green	95	5	0	73	Green	95	5	0
		Yellow	84	16	0		Yellow	71	21	8		Yellow	72	21	7
		Red	58	32	10		Red	59	31	10		Red	46	34	20
Hospital B	75	Green	97	2	1	75	Green	97	3	0	76	Green	96	4	0
		Yellow	80	9	11		Yellow	80	8	12		Yellow	75	16	9
		Red	38	14	48		Red	32	19	49		Red	41	10	49
Hospital C	64	Green	91	6	3	66	Green	90	6	4	67	Green	91	6	3
		Yellow	66	16	18		Yellow	63	23	14		Yellow	72	15	13
		Red	59	22	19		Red	56	22	22		Red	47	23	30
Hospital D	75	Green	92	6	2	74	Green	90	7	3	76	Green	92	6	2
		Yellow	68	18	14		Yellow	81	11	8		Yellow	73	16	11
		Red	47	14	39		Red	32	16	52		Red	29	10	61
Hospital E	70	Green	91	5	4	72	Green	92	6	2	74	Green	91	6	3
		Yellow	77	4	19		Yellow	69	9	22		Yellow	64	16	20
		Red	40	23	37		Red	39	20	41		Red	38	17	45

ED=Emergency Department

4b	Mean of the preceding 7 days					Mean of the preceding 14 days					Mean of the preceding 21 days					Mean of the preceding 31 days				
			Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)					Predicted ED high demand (%, Sensitivity)		
	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red	Accuracy (%)	Observed ED high demand	Green	Yellow	Red
Hospital A	70	Green	96	4	0	72	Green	96	4	0	73	Green	94	5	1	74	Green	94	4	2
		Yellow	87	10	3		Yellow	78	15	7		Yellow	76	21	3		Yellow	65	27	8
		Red	63	30	7		Red	60	34	6		Red	49	27	24		Red	44	31	25
Hospital B	73	Green	92	6	2	72	Green	91	9	0	73	Green	90	10	0	76	Green	91	8	1
		Yellow	74	13	13		Yellow	74	16	100		Yellow	71	21	8		Yellow	68	23	9
		Red	29	19	52		Red	26	20	54		Red	30	17	53		Red	24	15	61
Hospital C	65	Green	90	7	3	67	Green	93	4	3	66	Green	91	5	4	69	Green	91	5	4
		Yellow	74	15	11		Yellow	68	21	11		Yellow	71	18	11		Yellow	72	17	11
		Red	55	27	18		Red	53	23	24		Red	54	20	26		Red	49	19	32
Hospital D	73	Green	92	6	2	77	Green	92	5	3	78	Green	93	6	1	76	Green	93	6	1
		Yellow	74	10	16		Yellow	76	14	10		Yellow	72	15	13		Yellow	63	26	11
		Red	51	17	32		Red	35	16	49		Red	32	12	56		Red	36	18	46
Hospital E	76	Green	94	4	2	75	Green	93	5	2	75	Green	92	6	2	75	Green	93	5	2
		Yellow	70	9	21		Yellow	66	17	17		Yellow	64	21	15		Yellow	68	15	17
		Red	46	22	32		Red	44	18	38		Red	39	23	38		Red	30	30	40

ED=Emergency Department

4c	Mean of the preceding 7 days					
			Predicted ED high demand (%, Sensitivity)			
	Accuracy (%)	Observed ED high demand	Low	Middle	High	Very high
Hospital A	53	Low	62%	36%	1%	1%
		Middle	26%	56%	14%	4%
		High	8%	37%	33%	22%
		Very high	0%	20%	23%	57%
Hospital B	64	Low	79%	21%	0%	0%
		Middle	14%	73%	12%	1%
		High	0%	62%	24%	14%
		Very high	0%	15%	18%	67%
Hospital C	50	Low	66%	31%	3%	0%
		Middle	21%	60%	14%	5%
		High	6%	56%	19%	19%
		Very high	4%	20%	38%	38%
Hospital D	55	Low	67%	30%	3%	0%
		Middle	33%	53%	11%	3%
		High	11%	28%	44%	17%
		Very high	0%	41%	9%	50%
Hospital E	54	Low	56%	42%	2%	0%
		Middle	17%	66%	15%	3%
		High	3%	54%	26%	17%
		Very high	0%	20%	20%	60%

ED=Emergency Department

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	2
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	-
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	2/3
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2/3
Bias	9	Describe any efforts to address potential sources of bias	3
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	2/3
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	3/4
		(*b*) Describe any methods used to examine subgroups and interactions	-
		(*c*) Explain how missing data were addressed	3
		(*d*) If applicable, explain how loss to follow-up was addressed	-
		(*e*) Describe any sensitivity analyses	5
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6
		(b) Indicate number of participants with missing data for each variable of interest	7
		(c) Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	Report numbers of outcome events or summary measures over time	7

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8/9/10
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	-
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13
Generalisability	21	Discuss the generalisability (external validity) of the study results	13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	15

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Original Paper

A time-series cohort study to forecast emergency department visits in the city of Milan and predict high demand: a 2-day warning system

Author List: Rossella Murtas, PhD[1], Sara Tunesi, PhD[1], Anita Andreano, MD[1], Antonio G. Russo, MD PhD[1]

Author Affiliations: Epidemiology Unit, Agency for Health Protection of Milan, 20123 Milan, Italy

Corresponding Author: Antonio Giampiero Russo, Via Conca del Naviglio 45, 20122 Milan, Italy, agrusso@ats-milano.it, +390285782111, +390285782128

ABSTRACT

Objectives The emergency department (ED) is one of the most critical areas in any hospital. Recently, many countries have seen a rise in the number of ED visits, with an increase in length of stay and a detrimental effect on quality of care. Being able to forecast future demands would be a valuable support for hospitals to prevent high demand, particularly in a system with limited resources where use of ED services for non-urgent visits is an important issue.

Design Time series cohort study.

Setting We collected all ED visits between January 2014 and December 2019 in the five larger hospitals in Milan. To predict daily volumes, we used a regression model with ARIMA errors. Predictors included were day of the week and year-round seasonality, meteorological and environmental variables, information on influenza epidemics and festivities. Accuracy of prediction was evaluated with the Mean Absolute Percentage Error (MAPE).

Primary outcome measures Daily all-cause EDs visits.

Results In the study period, we observed 2,223,479 visits. ED visits were most likely to occur on weekends for children and on Mondays for adults and seniors. Results confirmed the role of meteorological and environmental variables and the presence of day of the week and year-round seasonality effects. We found high correlation between observed and predicted values with a MAPE globally smaller than 8.1%.

Conclusions Results were used to establish an ED warning system based on past observations and indicators of high demand. This is important in any health system that regularly faces with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits is still high.

Strengths and limitations of this study

- This study is one of the few studies linking temporal periodicity, occurrence of festivities, local weather conditions, and pollution to ED visits
- We estimated an ARIMA model for each hospital, thus taking into consideration each specific characteristic and incorporating weekly and annual seasonality with Fourier terms
- Results were used to establish an ED warning system based on past observations and indicators of high demand
- We cannot exclude the possible presence of unmeasured variables that may better predict ED visits and overcrowding
- ~~This study was intended to estimate ED demand and does not include information on staff rosters, two components unavoidably linked in any emergency department but that should be described separately~~

INTRODUCTION

The emergency department (ED) is the gateway (an open door) and the most critical area of a hospital, moving many activities and causing problems in the management of elective procedures when the number of patients who come knocking increases. In the last decade, many countries have seen a substantial rise in the number of ED visits, with an increase in length of stay,[1] and associated detrimental effects on quality of care. ED visits are unavoidably subject to fluctuation, and several models to predict high demand have been developed in the last decade, aiming at effectively managing hospital beds and staff rosters.[2] In Italy, even though the number of ED visits has been decreasing since 2016, the mean waiting time in EDs was high, between 12h and 24h in 3.5% of cases in 2017, and over 24 h in 2.1% of cases.[3] The definition of overcrowding[4] in the ED literature is not consistent, nor are the measures used to assess overcrowding, which vary from clinician perception of overcrowding, to input measures (e.g., waiting times, number of patients arrived), throughput measures (e.g., ED capacities, patient care time), output measures (e.g. percentages of hospital admissions, hospital beds), or multidimensional indices such as the Emergency Department Work Index (EDWIN). This variety of measures corresponds to the different type of factors studied as causes of ED crowding. We concentrate here on predicting the number of visits from input factors, i.e., determinants and modalities of patient inflow, such as non-urgent visits and Influenza season. In this case it is better to speak of overflow ~~(ref)~~.[5] We did not investigate throughput factors, describing organizational issues in the ED, such as inadequate staffing, nor output factors. The latter include one of the major reasons for ED overcrowding, which is the shortage of acute care bed capacity.[6–10] . Among the most investigated input factors are non-urgent visits, meaning “patients who could have been assessed and treated in other facilities that treat less urgent cases” ~~(Howard e t al.)~~.[11] In Italy in 2017, only 23% of ED visits were classified as red or yellow at triage, while 13%[3] had a low level of priority, coded white triage in Italy. This use of emergency department services is a signal of lack of continuity of primary care and difficulty of access to both primary and specialist care. It is also not cost-effective and leads to an increase in waiting times in the EDs.[12,13]

Several factors potentially affect the daily number of ED visits. Among these: annual[14,15], seasonal,[16–19] and weekly[14–19] periodicity, as well as festivities.[14,16,20,21] The effect of local weather conditions and pollution on ED visit volumes is still in debated: while some studies confirmed a significant association with temperature,[15,17–19,22,23] precipitation,[17,19] humidity,[22] and weather conditions,[23] other authors found these variables to be only mediocre predictors of the number of ED visits,[16] and found air pollution mostly impacting cardiac and respiratory diseases.[22] An additional factor that has been studied in relation with ED visit volumes is the flu, with around 7% of total accesses attributable to Influenza-like Illness (ILI) during the epidemic season.[24] Murtas and colleagues[25] evaluated the hypothesis of the early presence of the COVID-19 epidemic in Italy by analysing data on trends of access to EDs using a Poisson regression model adjusted for seasonality and influenza outbreaks. In this work they found that predicting ED visits by considering both seasonality and ILI rates, compared to a model tacking into account only seasonality, notably increased the fitting of the model. Therefore, syndromic surveillance (such as ILI rates which in Italy are provided weekly by the National Health Service Sentinel System) may be able to provide early warning of hospital bed capacity strain caused by seasonal respiratory disease.[26] To our knowledge, there is no study linking all this information together to ED visits.

The present study aims to develop a model for forecasting ED arrivals, using regression-based time series analysis with Auto-regressive integrated moving average (ARIMA) errors, accounting simultaneously for the effect of meteorological and environmental variables, as well as information on flu epidemics and festivities, on the number of ED visits in the city of Milan. The model is used to establish an innovative ED warning system providing a planning instrument for hospitals, based on past observations and indicators of high demand.

METHODS

Study Design

This is a retrospective study conducted in the area served by the Milan Agency for Health Protection (AHP) using current health care databases of daily ED visits aggregated at hospital level. No individual level data were used, and patients cannot be identified from aggregated data which do not contain low counts (i.e. cells with ≤ 5 counts). For this reason, and in accordance with Italian legislation, this study was not submitted for ethics approval.[27]

Study setting and population

We collected all ED visits, including patients registered at triage that voluntarily left the ED premises before being evaluated by a physician, between the 1st of January 2014 and the 31st of December 2019 in the five largest hospitals located in the city of Milan (figure 1). All five hospitals are public hospitals and received 49% of all emergency room access of the city of Milan, which has a total of 17 EDs, with a mean number of daily ED visits during 2014-2019 ranging from 124 for hospital C to 247 for hospital E.

Study protocol

Aggregated data on daily ED visit volumes, by age and gender, were extracted from the regional health database. Meteorological and environmental information was extracted from the Regional Environmental Protection Agency (ARPA).[28] Daily mean temperature, relative humidity (RH), cumulative precipitation, nitrogen dioxide (NO_2), and particulate matter with a diameter ≤ 10 μm (PM10) were collected from 2 monitoring stations (one measuring meteorological indicators and one measuring air pollution) located in the centre of Milan (figure 1). As sensitivity analysis we also investigated the effect of minimum, maximum and apparent temperature on daily ED visits.[29] Missing values on a specific day were imputed with the average of the measure in that specific year. Weekly data on ILI notifications were taken from the National Health Service Sentinel System (InfluNet).[30] Weekly incidence rates of ILI were expressed as the number of cases per 1,000 inhabitants per week. All available information was linked to daily ED visit volumes for each of the five hospitals included in the study. Datasets were divided into training (from the 1st of January 2014 to the 31st of December 2018) and validation sets (from the 1st of January 2019 to the 31st of December 2019). For each hospital, we first estimated model parameters on the training dataset and evaluated post-sample accuracy in the validation set. We included, in each model, only factors that significantly influenced the number of ED visits. Multicollinearity was evaluated calculating Pearson pairwise correlation between variables and variance inflation criterion (VIF)[31].

Patient and public involvement

Patients were not involved in this research.

Data analyses

Development of the predictive model

To predict the daily volume of visits in each ED, we used a time series approach consisting in a regression model with ARIMA errors.[32] The statistical units were days, 1,826 days in the training set and 365 in the validation set. This model is able to combine two powerful statistical methods: linear regression and ARIMA. Linear regression of Y on X is usually described by the equation $Y_t = \alpha + \beta x_t + \epsilon_t$, where Y_t and x_t are the values of Y and X at day t, α and β are the intercept and the slope of the regression line, and ϵ_t is the error of the model at day t (the deviations from the fitted line to the observed values) assumed to be independent from other values. The ARIMA model deals with auto-correlation between errors through two components: the auto-regressive and the moving average process. The auto-regressive component assumes that previous observations are good predictors for future values, while the moving average component allows the model to update the predictions if the level of a constant time series changes. ARIMA specification is described by 3 parameters (p, d, q), where p is the order of auto-regression (AR) that is the number of time lags, d is the degree of differencing (the number of times the data have had past values subtracted to make the time series stationary), and q is the order of the moving average process (MA). For each hospital, these parameters were identified examining total and partial autocorrelation function (ACF and PACF, respectively), as well as statistical significance (p-value<0.05), and minimal Akaike Information Criteria (AIC). Day of the week and year-round seasonality were controlled for by including Fourier terms, a series of sine-cosine functions capable of approximating periodicity.[20,32] The number of Fourier terms was chosen to minimise the AIC for each seasonal period (up to 7 for day of the week seasonality and up to 365 for year-round seasonality). Each seasonal component can be written in the model equation as

$$\sum_{j=1}^{n}\left[\alpha_j \sin\left(\frac{2\pi jt}{m}\right) + \beta_j \cos\left(\frac{2\pi jt}{m}\right)\right]$$

where *n* is the number of Fourier terms chosen to minimise the AIC (up to 7 for day of the week seasonality and up to 365 for year-round seasonality) and *m* is the seasonal period (7 for day of the week and 365 for year-round seasonality).

Therefore, meteorological and environmental variables, as well as information on flu epidemics and festivities, were retained in the final model only if statistically significant. As festivities, we considered Italian public holidays with school and office closures: New Year's Day, Epiphany, Easter Sunday and Monday, Italian Liberation Day, Labour Day, Foundation of the Italian Republic, assumption day, All Saints' Day, Saint Ambrose's Day (local patron saint), Feast of the Immaculate Conception, Christmas Day, Saint Stephen's day and New Year's Eve. In addition, we created dummies for specific festivities that were responsible for a significant variation in the number of ED visits: New Year's Eve and Assumption Day (August 15th). Diagnostics of the finally selected models were the Jarque-Bera test of normality, and correlation among the residuals according to the Ljung-Box test. Variables and tests were considered statistically significant if p-value was < 0.05.

The ARIMA model was compared with a simple regression model (M1) including only meteorological, environmental, and festivity covariates and with a generalized linear model (M2) also including the Fourier terms to control for seasonality. P-values were calculated by comparing the full model (ARIMA) to M1 and M2 using the likelihood ratio test.

Forecasting Accuracy

Predicted values on validation sets were estimated using one-step forecast.[32] We estimated parameters only on training sets. However, we calculated forecasts on validation sets using all of the data preceding each observation. The accuracy of predictions was evaluated with the Mean Absolute Percentage Error (MAPE), which expresses, as percentages, a unit-free measure of performance:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} * 100$$

with y_t and $\hat{y}_t$ respectively the observed and the predicted number of visits at day t, and n the number of days in the validation set (n=365 in this study).

High demand definition

We proposed a definition of high ED demand as days where the number of visits exceeded the median of the preceding 31 days. The days were defined as green (level 1) if the number of visits exceeded the median by less than 5%, yellow (level 2) if between 5% and 10%, red (level 3) if higher than or equal to 10%. High demand was calculated on the observed and predicted ED visits in validation sets, we thus calculated the proportion of observed high ED demand that is correctly classified by predicted high ED demand (called sensitivity or recall metrics for multiclass classification problems).[33] In addition, we calculated the accuracy of predictions as the number of correct classifications over the total number of observations. All statistical analyses were performed with R (version 3.6.3).[34]

To evaluate the proposed definition, we further calculated high demand as: the number of visits exceeding the median of the preceding 7, 14, and 21 days and the number of visits exceeding the mean of the preceding 7, 14, 21, and 31 days, defining green, yellow, and red levels of high demand as above. We chose 7, 14, and 21 lag days in order to adjust for weekly variation in the number of ED visits by design. We further calculated high demand as defined by the Lombardy Region[35]: when the number of visits exceeded the 91st percentile of the previous year time-series. Low demand days were defined as those with a number of visits smaller than 25th percentile, medium demand days as those with a number of visits between 25th percentile and 75th percentile, high demand days if between 75th percentile and 90th percentile, and finally very high demand days if over 91th percentile.

ED warning system

In the month of January 2020, we established an ED warning system (WS), which was used by the selected hospitals in Milan as a planning instrument for EDs and consists in a transmission of daily reports. This WS continued until February when the COVID-19 outbreak started in Italy. According to the model choices highlighted by the above methodology (validation and calibration of the model were performed with data from 2014 to 2019), parameters were updated weekly and used to establish the WS which operated in January 2020. A hypothetical daily report received from a hospital on the 5th of January 2020 can be found in figure 2. The report included forecasts of the number of visits for the following two days, with 95% margin errors and a high demand indicator (green, yellow or red). The forecasts were made incorporating in the model past meteorological and environmental information via an Application Programming Interface (API) where 2-day future forecasts of meteorological and environmental information were provided by ARPA Lombardia. Weekly information on ILI was downloaded every week from InfluNet, and included in the predictive models. Daily reports were constructed and dispatched automatically using R and R Markdown. During the WS campaign, we established a monitoring service capable of estimating daily sensitivity, accuracy of predictions and MAPE separately for prediction one and two days ahead.

All analyses were performed with R software (V.4.0.2; R Core Team, Vienna, Austria), models and Fourier terms were estimated respectively, using the Arima and the Fourier functions in the R package forecast[36] using the parameter xreg for covariate specification. VIF was calculated using the VIF function in the car package.[37]

RESULTS

ED visit volumes

Between the 1st of January 2014 and the 31st of December 2019 (training set of 1,826 and validation set of 365 days) we observed 2,223,479 visits, 370,633 on average every year. Daily mean number of visits by hospital, temporal, meteorological, and patient characteristics in the training sets are summarized in table 1. Missingness, over the whole period 2014-2019, in meteorological and environmental variables were found in 8 days for temperature, 7 days for precipitation, and 37 days for humidity. Description of training and validation sets, and plots of each hospital's time series are summarized in the Supplementary Material (Supplementary Table 1). The Pearson correlation between predictors varied from weak (absolute correlation<0.3) to moderate (absolute correlation between 0.3 and 0.7), with a maximum of -0.67 between temperature and ILI and 0.61 between NO2 and PM10. VIF was smaller than 5 for all variables, with a maximum of 2.8 for temperature and 1.9 for ILI. We therefore included all the variables in the models, selecting the final model according to the statistical significance of predictors and minimal AIC.

Table 1

Total number of visits and mean number of daily visits by hospital, temporal and meteorological factors, and patient characteristics between the 1st of January 2014 and the 31st of December 2019 in five emergency departments of the city of Milan, Italy.

	N (%)[a]	Mean (Min-Max)[b]		N (%)[c]	Mean (Min-Max)[d]
Hospitals			**Cumulative Precipitation (mm)**		
A	421741 (19)	192 (107-301)	**≤ 0.6**	1678953 (75.5)	1018 (563-1295)
B	457021 (20.6)	209 (65-302)	**0.7+**	544526 (24.5)	1005 (627-1392)
C	272308 (12.2)	124 (61-197)	**NO2 (µg/m³)**		
D	530519 (23.9)	242 (125-337)	**≤ 32**	564957 (25.4)	974 (563-1295)
E	541890 (24.4)	247 (133-346)	**33-44**	570284 (25.6)	1022 (723-1292)
Total	2223479	1015 (563-1392)	**45-57**	536484 (24.1)	1032 (698-1392)
Gender			**58+**	551754 (24.8)	1035 (693-1272)
F	1113405 (50.6)	508 (277-782)	**PM10 (µg/m³)**		
M	1087903 (49.4)	497 (277-661)	**≤ 20**	571339 (25.7)	990 (563-1261)
Age			**21-29**	575530 (25.9)	1017 (688-1295)
≤ 14	360600 (16.4)	165 (55-443)	**30-44**	544147 (24.5)	1023 (710-1392)
15-65	1307139 (59.4)	597 (317-860)	**45+**	532463 (23.9)	1032 (693-1272)
66+	533569 (24.2)	244 (141-385)	**ILI (n. of weekly new cases per 1,000 inhabitants)**		
	N (%)[c]	**Mean (Min-Max)[d]**			
Temperature (°C)			**≤ 1.2**	1310096 (58.9)	1001 (563-1295)
≤ 9.2	564744 (25.4)	1021 (693-1392)	**1.3-2.5**	303072 (13.6)	1031 (799-1256)
9.3-15.6	563764 (25.4)	1033 (813-1261)	**2.6-5.6**	303102 (13.6)	1031 (698-1261)
15.7-22.3	563757 (25.4)	1025 (656-1295)	**5.7+**	307209 (13.8)	1045 (693-1392)
22.4+	531214 (23.9)	980 (563-1292)	**Day before/after festivity**		
Relative Humidity (%)			**No**	2096838 (94.3)	1012 (563-1392)
≤ 50	560870 (25.2)	1018 (563-1295)	**Yes**	126641 (5.7)	1055 (688-1295)
51-62	552865 (24.9)	1009 (637-1292)	**Festivity**		

63-76	554041 (24.9)	1017 (627-1392)	**No**	2144726 (96.5)	1018 (677-1392)
77+	555703 (25)	1016 (786-1278)	**Yes**	78753 (3.5)	938 (563-1253)

ILI=Influenza-like illness

[a]Total number of visits by hospital, gender and age. In parenthesis the percentage of the number of visits out of the total (2,223,479 total number of visits, 2,201,308 with information on age and gender);
[b]Mean, minimum and maximum number of daily visits by hospital, gender and age;
[c]Total number of visits by temporal and meteorological factors (i.e. total number of visits in days with a particular value of temperature, humidity, etc.). In parenthesis the percentage of the number of visits of the total (2,223,479 total number of visits);
[d]Mean, minimum and maximum number of daily visits by temporal and meteorological factors (i.e. mean number of daily visits in the days with a particular value of temperature, Humidity etc.).

Model specification and ARIMA results

All models showed a very strong day of the week and year-round seasonality effect, according to ACF and PACF plots. To normalize residuals, outliers (in the training sets only) were replaced by the mean of the observations of the same day in the other years, consequently all models showed residual normally distributed according to the Jarque-Bera test (number of replaced outliers are presented in Supplementary Table 2). All models showed a lack of fitting on New Year's Eve and/or August 15th, for this reason we chose to define a specific dichotomous variable ("1" for the peculiar festivity, "0" for the other days) capable of detecting this extra variation. Table 2 displays the ARIMA parameters fitted for each model, and the number of Fourier terms that minimized AIC. All models were non-stationary in mean and needed one differencing to make the time series stationary (d=1). ARIMA parameters and Fourier terms were different across hospitals, showing that each time series needed different model specification. Table 2 also displays, for each hospital, the factors that significantly influenced the number of ED visits, and that were included in the models. High temperatures were always associated with a statistically significant increase in ED visit volumes, with a maximum increase of 1.84 daily visits every 1°C increase (hospital E, s.e. 0.18). Relative humidity was significantly associated with a limited decrease of total ED visits (-0.08, s.e. 0.04) for a 1% increment of RH only at hospital D. High levels of cumulative precipitation were associated (except for hospital C) with a statistically significant decrease in ED visits, with a maximum decrease of 0.31 daily visits every 1 mm of precipitation (hospital E, s.e. 0.06). Concerning air pollution, we found an opposite effect of NO_2 and PM10 on ED visits, with a mild significant negative effect for NO_2 in two hospitals (-0.08 and -0.09) and an even milder positive association with PM10 in one (0.03). Except for hospital C, the effect of ILI was always associated with the number of ED visits, showing an increase of daily visits between 0.73 and 1.74 (s.e. 0.29 and 0.41 respectively) at every

unit increase in weekly ILI rates. Festivities were associated with a decrease in ED visits of between 13 and 28 (s.e. 1.45 and 1.98), while special festivities were associated with the greatest decrease of at least 42 ED visits (s.e. 4.94). Autocorrelation function and correlation among residuals according to the Ljung-Box test by hospital and up to 30 and 366 lags can be found in Supplementary figure 1. ACF plots of residuals were overall in significance limits and the Ljung-Box test showed overall no significant correlation between residuals at different lags, except Hospital E which showed residual autocorrelation up to lag 366.

Table 2

Auto-regressive integrated moving average (ARIMA) specifications and covariate effects on the number of ED visits between the 1st of January 2014 and the 31st of December 2018 (training sets).

		Hospitals				
		A	B	C	D	E
Model specification	ARIMA parameters (p,d,q)	(0,1,2)	(1,1,1)	(1,1,2)	(1,1,1)	(1,1,1)
	Fourier terms†	3,13	3,14	3,13	3,16	3,15
Covariate Effects (se)[1]	Temperature (°C)	1.29 (0.15)	1.23 (0.14)	0.68 (0.11)	1.16 (0.18)	1.84 (0.18)
	Humidity (%)				-0.08 (0.04)	
	Precipitation (mm)	-0.2 (0.05)	-0.12 (0.05)		-0.13 (0.07)	-0.31 (0.06)
	NO2 (μg/m3)	-0.08 (0.03)			-0.09 (0.04)	
	PM10 (μg/m3)			0.03 (0.02)		
	ILI (weekly new cases per 1,000 inhabitants)	1.74 (0.41)	1.05 (0.37)	0.73 (0.29)		0.97 (0.46)
	Festivity		-28.23 (1.98)	-12.96 (1.45)	-25.42 (2.23)	-14.56 (2.39)
	Special Festivity*	-43.16 (6.31)	-57.64 (6.36); -62.61 (6.29)	-42.06 (4.92)	-59.86 (7.58)	-63.24 (7.92)
	Day before/after festivity	7.14 (1.5)	9.06 (1.58)	3.75 (1.22)		13.89 (1.96)

[1]Parameter estimates and standard errors in parentheses. Predictors were retained in the final model only if statistically significant (p-value<0.05)
† Number of sine and cosine terms used to approximate day of the week and year-round seasonality
***New Year's Eve for hospitals A, C-D and New Year's Eve and August 15th for hospital B**
ARIMA =Auto-regressive integrated moving average
ILI=Influenza-Like-Illness

Forecasting Accuracy and High demand definition

The accuracy of predictions (MAPE) in the validation sets, sensitivity and accuracy between observed and predicted high ED demand are displayed in table 3. Model performance was good, with small MAPEs in validation sets, ranging from a minimum of 5.5% for hospital D to a maximum of 8.1% for hospital C. The models showed high sensitivity on days with green-level high demand, almost 90% of days with predicted green-level high demand were confirmed from observed values.

On days with yellow-level high demand, sensitivity between predicted and observed demand was scarce, ranging from 0.04 for hospital B to 0.28 for hospital A. Sensitivity of red-level high demand varied between hospitals, with a minimum of 0.25 for hospital A to a maximum of 0.57 for hospital D. Observing Table 3 we can suggest that, for each hospital, at least 54% of the observed red-level high demand days were classified, from predictions, as being at least yellow-level. Accuracy was high, with at least 67% of the days with exactly the same predicted and observed high demand level (green, yellow or red).

All ARIMA models fitted the data significantly better than a simple regression model (M1) and a generalized linear model (M2), with MAPE for M1 and M2 above 13.5% and 9.8%, respectively (Supplementary Table 3). Observed and predicted ED visits in the validation sets (from the 1th of January 2019 to the 31th of December 2019) by date and hospital ~~The scatter plot of observed vs predicted values in the validation set~~ can be found in Supplementary Figure 2.

In Supplementary Table 2 we compared ARIMA results for different temperature specifications: mean, minimum, maximum and apparent temperature. The greatest effect on ED visits was attributed to mean temperature while indicators of performance and AIC where generally superior for mean temperature compared with minimum, maximum and apparent temperature. In Supplementary Table 2 we also calculated, only for outlier days, the relative error mean of observed vs predicted values in order to evaluate if extreme temperatures were better outlier predictors than mean temperature. Number of outliers replaced ranged from 2 for hospital A to 7 for hospital D, results suggested an overall better fit of outliers using minimum temperature (3 out 5 hospitals with smaller relative errors).

In supplementary table 4 we compared the high demand definition used in the ED warning system with similar definitions. There was slight improvement in percentage accuracy between the definition used and the other algorithms and there was no favourite algorithm for all hospitals: hospital B had a maximum improvement of 4% using the mean of the preceding 31 days or the median of the proceeding 21 days, hospitals A and C had an improvement of 2% using the mean of the preceding 31 days, hospital D had an improvement of 2% using the mean of the preceding 21 days, and finally hospital E had an improvement of 1% using the mean of the preceding 21 or 31 days. Using the high demand definition used by the Lombardy Region we did not find any improvement in accuracy, with an overall percentage of matched classification between 50% and 64%. High demand was always predicted less well compared to the definition used in our ED warning system. However, results showed good prediction of very high demand days with a sensitivity between 38% and 67%.

Table 3

Indicators of performance of the developed models: accuracy of predictions (MAPE) in the validation sets, and accuracy and sensitivity of high demand classification.

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	5.9	72	Green	93	6	1
			Yellow	64	28	8
			Red	46	29	25
Hospital B	5.7	72	Green	92	8	0
			Yellow	85	4	11
			Red	35	15	50
Hospital C	8.1	67	Green	88	8	4
			Yellow	78	10	12
			Red	45	20	35
Hospital D	5.5	76	Green	91	6	3
			Yellow	65	27	8
			Red	35	9	56
Hospital E	6.1	74	Green	90	8	2
			Yellow	59	24	17
			Red	34	28	38

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

ED warning system

In Table 4a and 4b we provided the accuracy of predictions (MAPE), sensitivity and accuracy between observed and predicted high ED demand in January (the operating period of the WS) for one- and two-days horizons. Errors of prediction (MAPE) were slightly higher than in the validation set, with MAPE for one-day horizon always smaller than MAPE for two-days horizons. Accuracy between observed and predicted high ED demand was never smaller than 0.45 and generally smaller than in the validation set.

Table 4

Accuracy of predictions (MAPE), sensitivity, and accuracy between observed and predicted high ED demand in January 2020 (the operating period of the WS) with a one- (4a) and two-day (4b) horizon.

4a

	MAPE	Accuracy (%)	Observed high ED demand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	7.8	52	Green	94	6	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	7.8	81	Green	87	13	0
			Yellow	0	100	0
			Red	17	17	67
Hospital C	8.6	52	Green	100	0	0
			Yellow	67	33	0
			Red	73	27	0
Hospital D	6.6	45	Green	55	36	9
			Yellow	0	33	67
			Red	50	33	17
Hospital E	11	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

4b

	MAPE	Accuracy (%)	Observed high EDdemand	Predicted high ED demand (%, Sensitivity) Green	Yellow	Red
Hospital A	8.1	55	Green	100	0	0
			Yellow	100	0	0
			Red	71	29	0
Hospital B	8.6	71	Green	73	27	0
			Yellow	0	100	0
			Red	25	17	58
Hospital C	9	45	Green	93	7	0
			Yellow	83	17	0
			Red	82	18	0
Hospital D	7.6	48	Green	50	18	32
			Yellow	0	0	100
			Red	33	0	67
Hospital E	11.2	45	Green	100	0	0
			Yellow	100	0	0
			Red	92	8	0

ED=Emergency Department
MAPE= Mean Absolute Percentage Error

DISCUSSION

In this work we proposed and implemented in daily practice, a system to predict the number of ED visits in five hospitals of the city of Milan. The system is based on regression models with ARIMA errors, where ARIMA parameters were allowed to vary between hospitals, according to their specific characteristics, and it provides daily reports on the number of visits predicted for the two subsequent days at the five hospitals participating in the study. The models showed a good overall performance with the MAPEs always smaller than 5.5% and 8.1%. Our results are slightly better than other studies: Marcilio and colleagues[20] forecast daily ED visits with Generalized Linear Models, finding MAPEs between 5.4% and 11.5%, according to different forecasting horizons and controlling for temperature effect.

Jones and colleagues,[16] using similar models, found MAPEs that varied between 8.5% and 15.5%. However, Duwalage et al.[15] using a Generalized Additive Model found MAPEs consistently lower than 5% for 14-day forecasts, which significantly improved including temperature in the model. Although the number of predicted ED visits was close to the observed values, and there was good sensitivity in predicting mild (green) high demand, there was moderate sensitivity in predicting the spike of ED visit volumes (red-level high demand) for some hospitals and acceptable sensitivity for hospital D. This is particularly important for the scope of this study, which aimed to forecast emergency department visits in order to develop a 2-day warning system. For this reason, a better predictive performance of the red-level forecast would be desired. In fact, one of the major reasons for ED overcrowding is the shortage of acute care bed capacity compared with the huge number of visiting patients. Comparing our definition with similar definitions, we found a slight improvement in percentage accuracy, around 1% and 4%, but there was no a favourite algorithm for all hospitals. Furthermore, using the definition of very high demand for ED visits defined by the Lombardy Region, we found sensitivity was better compared to our models, and we plan to implement this in further evolutions of our warning system. However, we found good sensitivity in classifying observed red-level demand as at least yellow from predictions, and accuracy among observed and predicted high demand levels was always close to 70%. The definition of high ED demand is not straightforward as it relies on the specific hospital's characteristics. It is one of the main causes of ED overcrowding, which is the most problematic issue in EDs, thus deserving the effort in trying to predict it. In this study, we proposed a definition based on percentage increases compared to the median of the preceding month, to warn EDs of requests rising over the levels they managed in the preceding month.

During the operating period of the warning system, January 2020, we found a worse adaptation of the models than in the validation year 2019. This could be due to the ongoing outbreak of COVID-19, as ED visits for non-critical problems were discouraged.[38]

Concerning potential predictors, we found a strong day of the week and year-round seasonality effect, adequately captured by the terms used to approximate periodicity (Fourier terms). Even if the aim of this work was to develop a forecasting model and not an explanatory model, here we found statistically significant effects of meteorological factors on ED visits. Temperature was always positively associated with outcome, with an increase in the number of visits for each 1-degree increase in temperature across hospitals, in accordance with previous results.[17,19,23] As reported in another study,[39] high temperatures are associated with ED visits, especially for the most susceptible population, as persons with diabetes or cancer, so it is important for public health officials to implement adaptation measures to manage the impact of high temperatures on population health. Here we found a slightly better fit for outliers using minimum temperature instead of mean temperature. Nonetheless, we decided to include mean temperature in the ED warning system because it showed the greatest effect on ED visits. Further work has to be done in order to investigate the role of extreme temperature on ED visit fluctuations. The role of precipitations has not yet been well established. To our knowledge only one study measured an indirect effect in reducing ED visit volumes.[19] In accordance with these results, rainy days were found to be mildly associated with reduced numbers of ED visits. NO2 and PM10 had a mild significant effect only in two hospitals and in one hospital respectively, and were discordant, with a negative effect of NO2 and a positive effect of PM10 on the number of ED visits. This may be explained considering that the effect of pollution on ED visits is generally exerted and measured on respiratory conditions, especially asthma, and/or cardiac rather than with total visits and it may be diluted when analysing all ED visits. Only a few studies found a positive association of Total Suspended Particles with all visits but trauma, going in the same direction as the small significant increase in the number of visits related to PM10 we found.[40] In addition, pollution estimated from the monitoring station (classified as from urban traffic) used in the analysis might be of a greater magnitude than that really observed in each hospital. However, even though the hospitals were mostly located on the outskirts of the city of Milan, they are all located in urban areas characterized by a similar air pollution pattern. ILI were found to significantly increase the number of ED visits, as found by other researchers.[41]

This study indicated a moderate to good sensitivity in predicting high demand, showing some difficulties in anticipating the exact red-level days. In the future we aim to investigate models capable to directly predicting ED peaks instead of predicting the number of ED visits such as copulas used for detecting spikes in signal processing in brain circuits[42] or machine learning models. Finally, when interpreting these results, it is necessary to be aware of the possible multicollinearity problem between variables, which may alter the magnitude and statistical significance of coefficients. However, according to Vatcheva 2016,[31] only high correlations between variables would result in a change of sign of the coefficients and furthermore VIFs were always smaller than 5. Correlated factors were the pollution variables (NO2 and PM10), which were never considered in the same model together. Given that the highest correlation was found among temperature and ILI, the effect of these variables on the number of ED visits may potentially be biased due to multicollinearity. However, we included both terms in the models given the fact that they have an effect on ED visits independently from one another.

Another limitation is the choice of the hospitals considered for this work, that is, major hospitals located in the city of Milan. This methodology might not be the feasible for use by small hospitals as they might have low counts or even no visits at all on particular days. A solution can be provided by implementing different statistical models, for example, negative binomial or zero-inflated Poisson models, and would be one of our aims in the next years.

High-demand ED forecasting has a dual nature that should be addressed: first, knowing in advance the number of expected visits would allow a more reasoned choice of the hospital to which request assistance and second, forecasts should be follow immediately by an evaluation of the available beds and of the staff needed to accommodate these expected visits. These two problems were not addressed in this work given that this study was intended to estimate ED demand only and does not include information on hospital capacity, but are fundamental ingredients that should be considered in the future.

In conclusion, we proposed a hospital specific ED warning system based on predictive models developed on previous attendances that can be used as a planning instrument in hospitals to increase resources, and to prevent high patient demand when a higher number of attendances is expected. This is important in any health system that usually deals with scarcity of resources, and it is crucial in a system where use of ED services for non-urgent visits are still high.

List of abbreviations

ACF: autocorrelation function

AHP: Milan agency for health protection

AIC: minimal Akaike information criteria

API: application programming interface

AR: auto-regression

ARIMA: Auto-regressive integrated moving average

ED: emergency department

ILI: Influenza-like-illness

MA: moving average process

MAPE: mean absolute percentage error

NO2: nitrogen dioxide

PM10: particulate matter with a diameter ≤ 10 µm

PACF: partial autocorrelation function

RH: relative humidity

WS: warning system

Ethics approval and consent to participate This study does not involve human participants or animal subjects. Ethics approval and consent to participate were not required, as this is an observational study based on data routinely collected by the Agency for Health Protection (ATS) of Milan, a public body of the Regional Health Service-Lombardy Region. Among the institutional functions of ATS, established by Lombardy regional legislation (R.L. 23/2015), is management of the care pathway at the individual level in the regional healthcare system and evaluation of the services provided to, and the outcomes of, patients residing in the covered area. This study is also ethically compliant with Italian National Law (D.Lgs. 101/2018) and the “General Authorisation to Process Personal Data for Scientific Research Purposes” (nos. 8 and 9/2016, referred to in the Data Protection Authority action of 13/12/2018). Data were anonymized with a unique identifier in the different datasets before being used for the analyses.

Contributors RM and AGR conceptualised the study and defined the methodology, accessed and verified the data. RM analysed the data set, ST and AA contributed to the literature search, data interpretation, and writing of the manuscript. ST and AA, have made substantial contributions to the revision of the paper. AGR supervised and managed the project.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Data availability statement Data are not publicly available due to property restrictions.

Word count 4862

Acknowledgement Authors thank ARPA Lombardia, in particular Orietta Cazzuli and Gian Paolo Minardi, for their valuable contribution.

REFERENCES

1 Pines JM, Hilton JA, Weber EJ, *et al.* International perspectives on emergency department crowding. *Acad Emerg Med* 2011; **18**: 1358–70.

2 Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. *Emergency medicine journal : EMJ* 2009. DOI:10.1136/emj.2008.062380.

3 Programma Nazionale Esiti - PNE 2018 - Home. https://pne.agenas.it/index.php (accessed Feb 18, 2020).

4 Hwang U, McCarthy ML, Aronsky D, *et al.* Measures of crowding in the emergency department: a systematic review. *Acad Emerg Med* 2011; **18**: 527–38.

5 Asplin BR. Measuring Crowding: Time for a Paradigm Shift. *Academic Emergency Medicine* 2006; **13**: 459–61.

6 Wang Z, Xiong X, Wang S, Yan J, Springer M, Dellinger RP. Causes of Emergency Department Overcrowding and Blockage of Access to Critical Services in Beijing: A 2-Year Study. *The Journal of Emergency Medicine* 2018; **54**: 665–73.

7 Jeyaraman MM, Copstein L, Al-Yousif N, *et al.* Interventions and strategies involving primary healthcare professionals to manage emergency department overcrowding: a scoping review. *BMJ Open* 2021; **11**: e048613.

8 Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. *Emergency Medicine Journal* 2003; **20**: 402–5.

9 McKenna P, Heslin SM, Viccellio P, Mallon WK, Hernandez C, Morley EJ. Emergency department and hospital crowding: causes, consequences, and cures. *Clin Exp Emerg Med* 2019; **6**: 189–95.

10 Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. *Ann Emerg Med* 2008; **52**: 126–36.

11 Howard M, Davis B, Anderson C, Cherry D, Koller P, Shelton D. Patients' perspective on choosing the emergency department for nonurgent medical care: a qualitative study exploring one reason for overcrowding. *Journal of emergency nursing: JEN : official publication of the Emergency Department Nurses Association* 2005. DOI:10.1016/J.JEN.2005.06.023.

12 Angelillo I, Ricciardi G, Nante N, *et al.* Appropriateness of hospital utilisation in Italy. *Public Health* 2000; **114**: 9–14.

13 Hakkoymaz H. How Appropriate are The Emergency Department Admissions? A Study From a Tertiary Hospital Emergency Department. *EJMI* 2020. DOI:10.14744/ejmi.2020.47376.

14 Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med* 2010; **17**: 970–8.

15 Duwalage KI, Burkett E, White G, Wong A, Thompson MH. Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas* 2020; published online Feb 17. DOI:10.1111/1742-6723.13481.

16 Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008; **15**: 159–70.

17 Rising WR, O'Daniel JA, Roberts CS. Correlating Weather and Trauma Admissions at a Level I Trauma Center. *Journal of Trauma and Acute Care Surgery* 2006; **60**: 1096–100.

18 Holleman DR, Bowling RL, Gathy C. Predicting daily visits to a waik-in clinic and emergency department using calendar and weather data. *J Gen Intern Med* 1996; **11**: 237–9.

19 Diehl AK, Morris MD, Mannis SA. Use of calendar and weather data to predict walk-in attendance. *South Med J* 1981; **74**: 709–12.

20 Marcilio I, Hajat S, Gouveia N. Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine* 2013; **20**: 769–77.

21 McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med* 2008; **15**: 337–46.

22 Díaz J, Alberdi JC, Pajares MS, *et al.* A model for forecasting emergency hospital admissions: effect of environmental variables. *J Environ Health* 2001; **64**: 9–15.

23 Macgregor DM. Effect of weather on attendance with injury at a paediatric emergency department. *Emergency Medicine Journal* 2003; **20**: 204–5.

24 Wargon M, Brun-Ney D, Beaujouan L, Casalino E. No more winter crisis? Forecasting daily bed requirements for emergency department admissions to hospital. *Eur J Emerg Med* 2018; **25**: 250–6.

25 Murtas R, Decarli A, Russo AG. Trend of pneumonia diagnosis in emergency departments as a COVID-19 surveillance system: a time series study. *BMJ Open* 2021; **11**: e044388.

26 Morbey RA, Charlett A, Lake I, *et al.* Can syndromic surveillance help forecast winter hospital bed pressures in England? *PLoS ONE* 2020; **15**: e0228804.

27 Italia. Provvedimento del Garante n. 2 del 16 giugno 2004. Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici. .

28 Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. ARPA Lombardia. https://www.arpalombardia.it:443/Pages/ARPA_Home_Page.aspx (accessed March 10, 2020).

29 Steadman RG. Norms of apparent temperature in Australia. *Aust Met Mag* 1994; **43**: 1–16.

30 Ministero della Salute. Influenza. http://www.salute.gov.it/portale/influenza/homeInfluenza.jsp (accessed March 10, 2020).

31 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* 2016; **6**. DOI:10.4172/2161-1165.1000227.

32 Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018.

33 M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *IJDKP* 2015; **5**: 01–11.

34 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020 https://www.R-project.org/.

35 Salute M della. Pronto soccorso, via libera della Conferenza Stato-Regioni alle nuove linee di indirizzo. http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=3849 (accessed April 24, 2020).

36 Hyndman [aut R, cre, cph, *et al.* forecast: Forecasting Functions for Time Series and Linear Models. 2020 https://CRAN.R-project.org/package=forecast (accessed Nov 14, 2020).

37 Fox J, Weisberg S. An R Companion to Applied Regression. SAGE Publications, 2018.

38 Jeffery MM, D'Onofrio G, Paek H, *et al.* Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. *JAMA Intern Med* 2020; published online Aug 3. DOI:10.1001/jamainternmed.2020.3288.

39 Lavigne E, Gasparrini A, Wang X, *et al.* Extreme ambient temperatures and cardiorespiratory emergency room visits: assessing risk by comorbid health conditions in a time series study. *Environ Health* 2014; **13**: 5.

40 Samet JM, Speizer FE, Bishop Y, Spengler JD, Ferris BG. The Relationship between Air Pollution and Emergency Room Visits in an Industrial Community. *Journal of the Air Pollution Control Association* 1981; **31**: 236–40.

41 Schanzer DL, Schwartz B. Impact of seasonal and pandemic influenza on emergency department visits, 2003-2010, Ontario, Canada. *Acad Emerg Med* 2013; **20**: 388–97.

42 Hu M, Li M, Li W, Liang H. Joint analysis of spikes and local field potentials using copula. *NeuroImage* 2016; **133**: 457–67.

Figure 1
Location of the five partecipating hospitals and of meteorological and air pollution monitoring stations in the city of Milan.

Figure 2
Hypothetical daily report received from a hospital on the 5th of January 2020.