Published in final edited form as:

Ann Emerg Med. 2009 October ; 54(4): 514–522.e19. doi:10.1016/j.annemergmed.2009.06.006.

Forecasting Emergency Department Crowding: An External, Multi-Center Evaluation

Nathan R. Hoot, MD, PhD¹, Stephen K. Epstein, MD, MPP², Todd L. Allen, MD³, Spencer S. Jones, PhD⁴, Kevin M. Bauml, MD⁵, Neal Chawla, MD⁵, Anna T. Lee⁵, Jesse M. Pines, MD, MBA⁶, Amandeep K. Klair, MD⁶, Bradley D. Gordon, MD, MS^{7,8}, Thomas J. Flottemesch, PhD^{7,8}, Larry J. LeBlanc, PhD⁹, Ian Jones, MD¹, Scott R. Levin, PhD¹⁰, Chuan Zhou, PhD¹, Cynthia S. Gadd, PhD, MBA¹, and Dominik Aronsky, MD, PhD¹

¹Vanderbilt University Medical Center, Nashville, TN

²Beth Israel Deaconess Medical Center, Boston, MA

³Intermountain Healthcare, Salt Lake City, UT

⁴University of Utah, Salt Lake City, UT

⁵Mount Sinai School of Medicine, New York, NY

⁶Hospital of the University of Pennsylvania, Philadelphia, PA

⁷Regions Hospital, St. Paul, MN

⁸HealthPartners Research Foundation, Bloomington, MN

⁹Owen Graduate School of Management, Nashville, TN

¹⁰Johns Hopkins University School of Medicine, Baltimore, MD

Abstract

Objective—To apply a previously described tool to forecast ED crowding at multiple institutions, and to assess its generalizability for predicting the near-future waiting count, occupancy level, and boarding count.

Methods—The ForecastED tool was validated using historical data from five institutions external to the development site. A sliding-window design separated the data for parameter estimation and forecast validation. Observations were sampled at consecutive 10-minute intervals during 12 months (n = 52,560) at four sites and 10 months (n = 44,064) at the fifth. Three outcome measures – the waiting count, occupancy level, and boarding count – were forecast 2, 4, 6, and 8 hours beyond each observation, and forecasts were compared to observed data at corresponding times. The reliability and calibration were measured following previously described methods. After linear calibration, the forecasting accuracy was measured using the median absolute error (MAE).

Results—The tool was successfully used for five different sites. Its forecasts were more reliable, better calibrated, and more accurate at 2 hours than at 8 hours. The reliability and calibration of the

© 2009 American College of Emergency Physicians. Published by Mosby, Inc. All rights reserved.

Contact: Nathan R. Hoot 400 Eskin Biomedical Library 2209 Garland Avenue Nashville, TN 37232 Phone: (615) 498-1979 Fax: (615) 936-1427 nathan.hoot@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

tool were similar between the original development site and external sites; the boarding count was an exception, which was less reliable at four out of five sites. Some variability in accuracy existed among institutions; when forecasting 4 hours into the future, the MAE of the waiting count ranged between 0.6 and 3.1 patients, the MAE of the occupancy level ranged between 9.0 and 14.5% of beds, and the MAE of the boarding count ranged between 0.9 and 2.7 patients.

Conclusion—The ForecastED tool generated potentially useful forecasts of input and throughput measures of ED crowding at five external sites, without modifying the underlying assumptions. Noting the limitation that this was not a real-time validation, ongoing research will focus on integrating the tool with ED information systems.

Introduction

Background

The emergency department (ED) serves essential needs in society, delivering emergency health care and simultaneously acting as a safety net provider¹⁻². The annual number of ED visits in the United States rose from 86.7 million in 1990, to 114.8 million in 2005³. During the same period, the number of EDs decreased from 5,172 to 4,611³. Moreover, 47% of American hospitals reported that they were operating at or over their ED capacity in 2007³. These divergent trends of capacity versus utilization may threaten the role of the ED, both internationally and in the United States⁴⁻¹⁰. The Institute of Medicine reported that crowding has led the emergency medical system to reach “the breaking point”¹¹.

No universal consensus exists for the definition of “crowding” in the ED setting¹²; however, it may be described as a mismatch between patient demand for services and provider supply of resources. A portion of the recent literature has focused on techniques to monitor¹³⁻¹⁷ and forecast¹⁸⁻²² crowding using varying definitions. A recent white paper described management approaches that could be facilitated by such techniques in an effort to reduce crowding²³. These include the one-bed ahead strategy, whereby inpatient units continuously anticipate the next patient requiring admission. More flexible staffing could also be implemented, whereby an ED schedules more nursing than necessary on average and allows shifts to end early during periods of low anticipated demand. Despite the possible applications, forecasting tools have not yet seen widespread operational adoption.

Importance

One challenge associated with monitoring and forecasting ED crowding is the generalization of models beyond the institutions where they were developed²⁴⁻²⁶. The decrease in predictive ability commonly seen when transporting models between sites may be due to the varying definitions of crowding, organizational structures, or workflow paradigms that exist among EDs. We recently described a discrete event simulation to forecast near-future ED crowding²². This tool, called ForecastED, was applied to predict crowding according to seven input, throughput, and output measures²⁷ at the development site. We designed the model with generalizability as a central goal; however, its ability to forecast crowding at other sites has not been shown. Several steps are required to transform a prediction rule into a clinical decision rule that alters patient care – these include derivation, narrow validation, broad validation, and subsequently impact analysis²⁸. The need for broad validation of the ForecastED tool motivated the present study.

Goal of This Investigation

The objective of this study was to externally validate the ForecastED tool for predicting near-future measures of crowding at institutions that are distinct from the original development site. More specifically, we intend 1) to demonstrate whether the model parameters can be fitted for

external sites without changing the underlying assumptions, and 2) to determine whether the forecasting performance is comparable between external sites and the original development site.

Methods

Theoretical Model of the Problem

The ForecastED tool implements a computerized “virtual ED” through a discrete event simulation intended to mimic the operations of an actual ED²². The process of developing the underlying model was theoretical and based on clinical experience. An interdisciplinary team proceeded iteratively within design constraints of forecasting power, minimal input requirements, and fast execution to determine a set of mathematical assumptions, together with software implementing them, specifying the operational structure of a generic ED.

The motivation to use discrete event simulation in forecasting, instead of another technique such as time series regression, is that it operates on patient data at a relatively detailed level, rather than at an institutional summary level – this granularity allows the tool to generate crowding forecasts without being limited to pre-determined outcome measures. An autoregressive model would generate forecasts for a single outcome measure – for instance, the waiting count, the occupancy level, or the boarding count – that must be selected during model development. By comparison, the flexibility of discrete event simulation could allow one model to forecast all of these crowding indicators, among others. This property exists because the model input is a detailed list of patients who are in the ED at the present, while the model output is a detailed list of patients who are projected to be in the ED at a specified point in the near future.

Two design decisions may allow for generalizability of the ForecastED tool: First, specific numerical parameters for each statistical distribution are not built into it; these parameters are flexible and may change before running the simulation. Second, institutional constants such as the number of ED beds or the number of acuity levels are likewise not built into it. In summary, the structural assumptions underlying the ForecastED tool were conserved among all validation sites, while the numerical parameters and institutional constants were intentionally allowed to vary between sites. The assumptions, parameters, and constants have been described in detail previously²².

Study Design

We validated the ForecastED tool using historical data from consecutive patient encounters during a 15-month period (11/1/2005 – 1/31/2007) at five locations, all of which were geographically and operationally distinct from the location where ForecastED was developed. The Institutional Review Board at each participating site approved the study.

We maintained unique numerical parameters and institutional constants for each site using a sliding-window study design, which was applied previously during the initial development and single-site validation of the ForecastED tool²². Given an observation time and institution, this technique uses data from the recent past – four weeks in the present study – to fit parameters for all statistical distributions within the simulation. Then it uses data from the near future – 2, 4, 6, and 8 hours in the present study – for forecast validation. These windows are always relative and are adjusted each time the observation point is moved. The primary purpose of this design was to ensure that the sets of data used to estimate the parameters and to validate the forecasts remained independent at all times. The secondary purpose of this design was to keep the parameters accurate with respect to seasonal variation that may occur at a given institution.

We generated observations in time series at 10-minute intervals between January and December of 2006 ($n = 52,560$) at sites A, B, D, and E. At site C, we repeated this process at 10-minute intervals between March and December of 2006 ($n = 44,064$) due to the unavailability of data from earlier dates.

Setting

Our study took place at five academic, urban, tertiary-care medical centers. Three of the sites are located in the Northeastern region of the United States, while the other two are located in the Midwestern and Western regions. General descriptive characteristics for each participating institution and its affiliated ED are presented in table 1. The hospitals range in size from 425 to 1,171 beds, including 50 to 133 critical care beds. Four of the sites are designated as level 1 trauma centers. Three serve adult populations, while two serve both adult and pediatric populations. The EDs range in licensed capacity from 25 to 39 beds. Three of the institutions have observation units managed by the ED that range in size from 5 to 8 beds. Because the EDs at these sites can access the observation unit for overflow when crowded, these beds are included in the total ED capacity for the purpose of the simulation. Thus, the total ED capacity specified for the ForecastED tool was 33 beds at site A, 35 beds at site B, 25 beds at site C, 39 beds at site D, and 47 beds at site E. Four of the sites triage ED patients using the Emergency Severity Index (ESI), a five-level score where lower values indicate greater urgency²⁹; one site triages ED patients according to a four-level ranking. The number of ED attending physicians employed by the participating sites ranges from 18 to 37. Local policy at each of the five sites allows for ambulance diversion during periods of crowding.

Selection of Participants

The study included data from all patient visits at each participating site during the study period, with three exclusion criteria applied: 1) Patient visits were excluded if the time of registration or the time of discharge were missing, since we could not accurately determine when the patient was present in the ED. 2) Patient visits were excluded if the patient was admitted directly to the hospital without being placed into an ED bed, because these patients tend not to compete for ED resources. Such patient encounters are referred to as “immediate admissions” in this study. 3) At site A which has a separate psychiatric ED, patient visits were excluded if the chief complaint was purely psychiatric, because these patients were not deemed to contribute much to crowding at that site.

Data Collection and Processing

The following variables are required for each patient visit to estimate parameters for, and to generate forecasts by, the ForecastED tool²²: 1) time of initial registration at the ED, 2) time placed into an ED treatment bed, 3) time of hospital bed request if applicable, 4) time of discharge from the ED facility, 5) triage category assigned to the patient, 6) whether the patient left without being seen, and 7) whether the patient was admitted to the hospital. Each institution collected these data from ED patient-tracking information systems. Two sites used commercial information systems, while three sites used information systems that were developed in-house. The following institutional constants were also supplied to the model as necessary to generate forecasts²²: 1) total ED capacity, including licensed treatment beds and, where applicable, beds within an ED-managed observation unit, and 2) number of acuity levels in the ranking system used to triage ED patients.

Forecasts were obtained at a given time and institution using the following series of steps: 1) All patients who were discharged from the ED during the preceding four weeks were identified. 2) The required parameters for each statistical distribution governing the simulation were estimated using the set of historical patient encounters. Formulas used for parameter estimation are given in appendix 1. 3) All patients who were present in the ED at the observation time of

interest were identified. 4) The simulation was initialized according to the set of current patient encounters, with patients being placed in the virtual waiting room or virtual ED beds as appropriate. 5) The simulation ran 2 hours into the future before terminating. At that time, the state of the virtual ED was noted, and the waiting count, occupancy level, and boarding count were measured. 6) The prior two steps were repeated 1,000 times to obtain an average for each outcome measure. 7) The prior three steps were repeated with the simulation running 4, 6, and 8 hours into the future. The actions of data processing and parameter estimation were automated using a Python language script (version 2.3.5, <http://www.python.org>).

Outcome Measures

The waiting count was defined as the number of patients in the waiting room. The occupancy level was defined as the total number of patients in ED beds divided by the number of treatment beds (this value may exceed 100% when patients are treated in non-licensed areas like hallway beds or chairs). The boarding count was defined as the number of patients with hospital admission orders who await inpatient beds. These three outcome measures, with identical definitions, were used during the development of ForecastED²². Details on calculating these data using raw patient information are provided in appendix 1.

Primary Data Analysis

We calculated the Pearson's r coefficient of correlation to quantify the reliability of the forecasts with respect to the actual operational measure at the corresponding point in the future. For example, when the simulation used the operational status at noon to forecast the occupancy level 8 hours in the future, we compared the resulting forecast against the known actual occupancy level at 8:00 PM that evening. The Pearson's r value reflects the degree of linear association between two measures, without penalizing for any consistent numerical bias. The square of this value describes the proportion of total variation explained by the forecasts. We calculated the Pearson's r with 95% confidence intervals (CI) using 250 iterations of the ordinary non-parametric bootstrap method³⁰.

We recognized that the crowding status of an ED is likely to be autocorrelated. For example, the occupancy level at noon on a given day provides some potentially useful information about the occupancy level at 1:00 PM that afternoon. We considered the present status of an ED to be a naïve predictor of the near-future status of that same ED, so we used this as our control measure for describing the additional utility provided by the forecasts. The autocorrelation gives the Pearson's r coefficient of correlation within a single time series, such that one point in time is compared with a later point in time, following a specified time interval. The usefulness of the simulation forecasts was judged by whether the reliability of the simulation forecasts exceeded the inherent autocorrelation within each actual operational measure. We calculated the autocorrelation coefficients with 95% CI using 250 iterations of the ordinary non-parametric bootstrap method³⁰.

While correlation coefficients are useful to assess the amount of collinearity that exists between two measures, they cannot detect any systematic bias that may exist – that is, any consistent over-estimation or under-estimation that would reduce numerical agreement³¹. To assess whether a systematic bias existed between the simulation forecasts and the actual operational measures, we calculated the mean and standard deviation of the residual error. We would consider a mean near zero, in the context of the associated standard deviation, to demonstrate good calibration.

The above statistical analysis was identical to the protocol used during the initial, single-center validation of ForecastED²². We performed one additional step of measuring the accuracy, with the goal of making the results more easily interpretable. First, we calibrated the forecasts using

the best-fitting line between predicted and actual operational data. This line was calculated with seven days of time series data ($n = 1,008$) preceding each observation, and the resulting linear transformation was used to obtain a single bias-corrected forecast. The calibration process was repeated over each time series of forecasts, in a manner analogous to the sliding-window study design described above. This step was justified on the grounds that a systematic bias was found to exist during previous research on the ForecastED tool²², and it mirrors the intended real-world application of the tool. Next, we calculated the median absolute error (MAE) between the calibrated forecasts and the actual operational data, with values closer to zero denoting greater accuracy.

We conducted all statistical analyses using R (version 2.8.1, <http://www.r-project.org>).

Results

Summary statistics of the conditions within each participating ED during the study period are presented in table 1. The ED at site C was the least crowded in terms of the total volume (40,193 visits), percentage of patients leaving without being seen (0.3%), median number of patients in the waiting room (0 patients), median percentage of beds occupied (52%), and median number of patients boarding simultaneously (1 patient). The ED at site B had the highest annual volume (62,219 visits), while the ED at site A had the highest percentage of patients leaving without being seen (5.2%), median number of patients in the waiting room (6 patients), and median percentage of beds occupied (91%). The ED at site E boarded the highest median number of patients simultaneously (12 patients). The percentage of total time spent on ambulance diversion ranged among the participating sites from 0.3% at site C to 7.0% at site A. For comparison, the United States average of total time spent on diversion in 2002 was 2.9%, or 7.6% among centers with high annual volumes³².

The number of patient visits excluded from the analysis totaled 1,418 patient visits from site A (0.2% immediate admissions, 1.9% purely psychiatric), 1,035 patient visits from site B (1.3% missing data, <0.1% immediate admissions), 28 patient visits from site C (<0.1% immediate admissions), 0 patient visits from site D, and 14 patient visits from site E (<0.1% immediate admissions). After calculating the time series of simulation forecasts and operational measures, implausible values were found to exist that were associated with episodes of computer system downtime at sites B and C. We removed affected observations from the time series before conducting statistical analysis; these included two segments of 36 hours each at site B and four segments of 24 hours each at site C.

The numerical agreement between the observed occupancy level and the simulation forecasts at site A may be visualized using the scatterplots in the figure. The full set of data representing all outcome measures and participating institutions may be visualized in appendix 2.

The site-by-site reliability, calibration, and accuracy of the forecasts 4 hours in the future are presented in tables 2, 3, and 4, respectively. The full results on the forecasting performance at 2, 4, 6, and 8 hours in the future are presented in appendix 3. Three trends were noted in the results for all operational measures considered: First, the forecasts became less reliable as the forecasts extended further into the future. For example, the forecasts of the waiting room count at site A had Pearson's r statistics of 0.79, 0.71, 0.65, and 0.61 with the observed waiting room count 2, 4, 6, and 8 hours in the future. Second, the reliability of the simulation tended to drop more gradually as the forecasting length increased, compared with the inherent autocorrelation which dropped more rapidly. For example, the forecasts of the occupancy level at site D had Pearson's r statistics of 0.87, 0.79, 0.76, and 0.73 with the observed occupancy level 2, 4, 6, and 8 hours in the future; the inherent autocorrelation was 0.86, 0.61, 0.31, and 0.03 when calculated for delays of 2, 4, 6, and 8 hours. Third, there was sufficient systematic bias, as

indicated by the residual mean differing from zero, to justify an additional calibration step in real-world deployment. For example, when forecasting the boarding count at site B, the model tended to overestimate by 1.5 ± 1.7 , 2.0 ± 1.9 , 2.4 ± 2.0 , and 2.7 ± 2.1 , respectively, at 2, 4, 6, and 8 hours in the future.

The forecasts of the waiting count were more reliable than the inherent autocorrelation at four out of five sites. At sites B, D, and E, the MAE was consistently accurate within 1.5-2.1 patients at 2 hours in the future, and within 1.7-2.5 patients at 8 hours in the future. At site A, the MAE was slightly less accurate, within 2.7 patients at 2 hours in the future and 3.5 patients at 8 hours in the future. This may be explained by the observation that site A had a busier waiting room, with a median of 6 patients and 75th percentile of 12 patients in the waiting room. At site C, the forecasts of the waiting count correlated weakly with the observed waiting count across all forecasting lengths ($r = 0.15$ at 2, 4, 6, and 8 hours). However, the accuracy at site C remained constant at 0.6 patients up to 8 hours in the future.

The occupancy level forecasts exceeded the inherent autocorrelation at all participating sites. At sites A, B, C, and E, the MAE of the occupancy level forecasts ranged from 7.1-9.6% of beds at 2 hours in the future, to 10.0-12.0% of beds at 8 hours in the future. At site D, the forecasts were less accurate, increasing from 11.6% of beds to 16.8% of beds, respectively, at 2 hours and 8 hours in the future. The scatterplot of the actual and predicted occupancy levels at site D, shown in figure E11 of appendix 2, shown an unusual trend: the forecasts of the occupancy level tend to be capped near 100%, yet site D tends to reach occupancy levels exceeding 150% more often than the other sites.

Considering the forecasts of the boarding count, the reliability at site D exceeded the inherent autocorrelation, but this did not hold true for any other site. At sites B and C, which boarded the fewest patients at one time among the five participating institutions, the MAE of the boarding count forecasts remained close to 1.0 patient up to 8 hours in the future. At sites A, D, and E, the MAE ranged from 1.4-2.0 patients at 2 hours in the future, to 2.3-3.8 patients at 8 hours in the future.

Limitations

There are a number of limitations to our study, including the use of historical patient data. While the ForecastED tool is intended for real-time application, we considered it technically more feasible to validate the tool at five institutions using offline analysis instead of live deployment. The differences between non-concurrent and concurrent data cleaning may have affected the results. The technique of cleaning data in real time is more difficult due to inherent uncertainty of the information, so an accurate determination cannot always be made whether a given patient meets the inclusion criteria. For example, when examining data from a center where psychiatric patients should be excluded, all patients having solely psychiatric complaints can be identified in retrospect. However, when considering patients in real time as they present to triage, one cannot immediately know who will be referred to the psychiatric ED. Because of this, the results may be considered optimistic for each participating site; however, the percentages of excluded data were small, so it seems unlikely that the real-time performance would degrade substantially due to this effect. The results may still provide useful information regarding the generalizability of the ForecastED tool.

A selection bias may exist in the study, with respect to which institutions participated; the five sites were chosen with an intentional focus on academic, metropolitan EDs. Thus, the validation results may not be representative of rural or non-teaching hospitals. We justify the sampling of participating sites on the grounds that the crowding burden disproportionately affects EDs like the ones described here^{3,33}, and a method to forecast crowding offers more potential value to busy, crowded EDs. An additional, practical constraint required all participating institutions

to have computerized patient tracking systems. This research would have been difficult to perform otherwise, given that several data elements needed to be obtained on a per-patient basis over a large span of time.

Many simplifying assumptions were made in the process of applying the ForecastED tool across five institutions. The discrete event simulation represents a generic ED, and it does not capture the rich variety of strategies that physicians and administrators may employ to improve ED operations. It treats all beds identically, so it does not handle trauma bays differently from unmonitored beds. It does not model fast-track beds, which may only be open during specified hours of the day and are not equipped to handle patients with severe conditions. The schemes used to allocate beds in the ED and hospital are simple – this may explain why forecasts of the occupancy level rarely exceed 100%, and why forecasts of the boarding count do not exceed the autocorrelation in reliability. In its present form, the simulation does not capture the different roles that physicians may serve during a shift. The ForecastED tool was designed to represent the common denominator across institutions, and despite this, it performed generally well in forecasting the waiting count and the occupancy level.

Discussion

We successfully estimated parameters for, and obtained forecasts from, the ForecastED tool at five different institutions. All required numerical parameters were fitted using historical data from ED information systems, and no modification to the model's assumptions, as originally described, was necessary to generate forecasts in different settings²².

The results show that, with respect to its ability to forecast the waiting count and the occupancy level, the ForecastED tool generalizes fairly well among the participating institutions. The degrees of reliability and accuracy varied among sites; however, after controlling for the intrinsic difficulty of forecasting operational data as measured by the autocorrelation, the simulation gave additional predictive information. Similar observations were made at the original development site of the ForecastED tool²². The forecasts of the occupancy level appeared to remain useful up to 8 hours in the future, while the forecasts of the waiting count may be most useful 4 hours or less in the future. The pattern noted in the occupancy level forecasts at site D suggests a future refinement to the ForecastED tool: The simulation assumes that an occupancy level of 100% is only exceeded for the most critically ill patients²², even though some institutions – notably site D in this study – routinely place patients in hallways or doubled in rooms, achieving occupancy levels of 150% or even 200%. This assumption may need to be relaxed in future research.

The ForecastED tool generally provided little additional predictive information, beyond the autocorrelation, for the number of boarding patients. This observation held true across four of the five external validation sites. This was not the case for the original development site²², and this difference may provide useful information regarding potential future improvements for the ForecastED tool. It has been suggested that extended boarding of patients in the ED substantially exacerbates the crowding problem^{7,34-36}. Furthermore, some researchers have reported that improving access to hospital beds may lessen the burden of ED crowding³⁷⁻³⁸. Based on these two lines of reasoning, one might improve the forecasts of the boarding count by making the process of simulated hospital bed allocation more robust within the ForecastED tool. This might be achieved through the use of granular data from the affiliated hospital, such as the present inpatient occupancy and projected lengths of stay.

Numerous challenges are associated with forecasting ED crowding – for instance, the lack of uniformly accepted definitions for crowding and common operational data³⁹⁻⁴⁰. The ForecastED tool was developed with the goal of addressing the above issues, through

generating forecasts in a target-agnostic manner²². The variability noted between sites in our study may be attributable to differences in operating room schedules, differences in policy for managing patient flow, and unique factors that influence input from the community. Substantial complexity exists among the hospital factors that affect crowding, and this may augment the complexity of forecasting tools in later research. The practical value of forecasting ED crowding depends upon real-time data, which is not yet available in many centers; furthermore, this work could gain additional utility if regional information sharing within hospital networks became common practice⁴¹. Most importantly, any forecasts of ED crowding must be combined with a plan to intervene and alleviate crowding in order to achieve practical value; until this occurs, we refrain from speculating on whether the forecasts are sufficiently accurate.

In summary, we validated the ForecastED tool, without modifying any assumptions from the original description, to forecast ED crowding at five institutions separate from the development site. The tool generated potentially useful forecasts of input and throughput measures of ED crowding, and further opportunities may exist to improve the forecasts of output measures²⁷. Future research will address the question of how to operationalize the ForecastED tool, with the goal of determining whether it can improve management of ED workflow to lessen the crowding burden.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The first author was supported by the National Library of Medicine grant LM07450 and National Institute of General Medical Studies grant T32 GM07347. The research was also supported by the National Library of Medicine grant R21 LM009002-01.

References

- [1]. Gordon JA, Billings J, Asplin BR, et al. Safety net research in emergency medicine: proceedings of the Academic Emergency Medicine Consensus Conference on "The Unraveling Safety Net". *Acad Emerg Med* 2001;8(11):1024–9. [PubMed: 11691663]
- [2]. American Academy of Pediatrics Committee on Pediatric Emergency Medicine. Overcrowding crisis in our nation's emergency departments: is our safety net unraveling? *Pediatrics* 2004;114(3):878–88. [PubMed: 15342870]
- [3]. TrendWatch Chartbook 2007. American Hospital Association web site; [Accessed July 18, 2007]. Available at: <http://www.aha.org/aha/trendwatch/2007/cb2007chapter3.ppt>.
- [4]. Li G, Lau JT, McCarthy ML, et al. Emergency department utilization in the United States and Ontario, Canada. *Acad Emerg Med* 2007;14(6):582–4. [PubMed: 17470903]
- [5]. Proudlove NC, Gordon K, Boaden R. Can good bed management solve the overcrowding in accident and emergency departments? *Emerg Med J* 2003;20(2):149–55. [PubMed: 12642528]
- [6]. Espinosa G, Miró O, Sánchez M, et al. Effects of external and internal factors on emergency department overcrowding. *Ann Emerg Med* 2002;39(6):693–5. [PubMed: 12023721]
- [7]. Fatovich DM, Nagree Y, Sprivilis P. Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia. *Emerg Med J* 2005;22(5):351–4. [PubMed: 15843704]
- [8]. Shih FY, Ma MH, Chen SC, et al. ED overcrowding in Taiwan: facts and strategies. *Am J Emerg Med* 1999;17(2):198–202. [PubMed: 10102327]
- [9]. Rehmani R. Emergency section and overcrowding in a university hospital of Karachi, Pakistan. *J Pak Med Assoc* 2004;54(5):233–7. [PubMed: 15270179]
- [10]. Graff L. Overcrowding in the ED: an international symptom of health care system failure. *Am J Emerg Med* 1999;17(2):208–9. [PubMed: 10102330]

- [11]. Committee on the Future of Emergency Care in the United States Health System. Hospital-based emergency care: at the breaking point. National Academies Press; Washington, DC: 2006.
- [12]. Hwang U, Concato J. Care in the emergency department: how crowded is overcrowded? *Acad Emerg Med* 2004;11(10):1097–101. [PubMed: 15466155]
- [13]. Bernstein SL, Verghese V, Leung W, et al. Development and validation of a new index to measure emergency department crowding. *Acad Emerg Med* 2003;10(9):938–42. [PubMed: 12957975]
- [14]. Reeder TJ, Burleson DL, Garrison HG. The overcrowded emergency department: a comparison of staff perceptions. *Acad Emerg Med* 2003;10(10):1059–64. [PubMed: 14525738]
- [15]. Weiss SJ, Derlet R, Arndahl J, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). *Acad Emerg Med* 2004;11(1):38–50. [PubMed: 14709427]
- [16]. Asplin BR, Rhodes KV, Flottemesch TJ, et al. Is this emergency department crowded? A multicenter derivation and evaluation of an emergency department crowding scale (EDCS) [abstract]. *Acad Emerg Med* 2004;11:484–485.
- [17]. Epstein SK, Tian L. Development of an emergency department work score to predict ambulance diversion. *Acad Emerg Med* 2006;13(4):421–6. [PubMed: 16581932]
- [18]. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med* 1994;23(2):299–306. [PubMed: 8304612]
- [19]. Asplin BR, Flottemesch TJ, Gordon BD. Developing models for patient flow and daily surge capacity research. *Acad Emerg Med* 2006;13(11):1109–13. [PubMed: 17015412]
- [20]. Jones SS, Thomas A, Evans RS, et al. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med* 2008;15(2):159–70. [PubMed: 18275446]
- [21]. McCarthy ML, Zeger SL, Ding R, et al. The challenge of predicting demand for emergency department services. *Acad Emerg Med* 2008;15(4):337–46. [PubMed: 18370987]
- [22]. Hoot NR, LeBlanc LJ, Jones I, et al. Forecasting emergency department crowding: a discrete event simulation. *Ann Emerg Med* 2008;52(2):116–25. [PubMed: 18387699]
- [23]. Eitel DR, Rudkin SE, Malvey MA, et al. Improving service quality by understanding emergency department flow: a white paper and position statement prepared for the American Academy of Emergency Medicine. *J Emerg Med.* May 29;2008 [Epub ahead of print].
- [24]. Jones SS, Allen TL, Flottemesch TJ, et al. An independent evaluation of four quantitative emergency department crowding scales. *Acad Emerg Med* 2006;13(11):1204–11. [PubMed: 16902050]
- [25]. Hoot NR, Zhou C, Jones I, et al. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med* 2007;49(6):747–55. [PubMed: 17391809]
- [26]. McCarthy ML, Aronsky D, Jones ID, et al. The emergency department occupancy rate: a simple measure of emergency department crowding? *Ann Emerg Med* 2008;51(1):15–24. [PubMed: 17980458]
- [27]. Asplin BR, Magid DJ, Rhodes KV, et al. A conceptual model of emergency department crowding. *Ann Emerg Med* 2003;42(2):173–80. [PubMed: 12883504]
- [28]. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144(3):201–9. [PubMed: 16461965]
- [29]. Wuerz RC, Milne LW, Eitel DR, et al. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000;7(3):236–42. [PubMed: 10730830]
- [30]. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986;1(1):54–77.
- [31]. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307–10. [PubMed: 2868172]
- [32]. Burt CW, McCaig LF, Valverde RH. Analysis of ambulance transports and diversions among US emergency departments. *Ann Emerg Med* 2006;47(4):317–26. [PubMed: 16546615]
- [33]. Andrulis DP, Kellermann A, Hintz EA, et al. Emergency departments and crowding in United States teaching hospitals. *Ann Emerg Med* 1991;20(9):980–6. [PubMed: 1877784]
- [34]. Fromm RE Jr, Gibbs LR, McCallum WG, et al. Critical care in the emergency department: a time-based study. *Crit Care Med* 1993;21(7):970–6. [PubMed: 8319477]

- [35]. Schull MJ, Lazier K, Vermeulen M, et al. Emergency department contributors to ambulance diversion: a quantitative analysis. *Ann Emerg Med* 2003;41(4):467–76. [PubMed: 12658245]
- [36]. Schneider SM, Gallery ME, Schafermeyer R, et al. Emergency department crowding: a point in time. *Ann Emerg Med* 2003;42(2):167–72. [PubMed: 12883503]
- [37]. Dunn R. Reduced access block causes shorter emergency department waiting times: An historical control observational study. *Emerg Med (Fremantle)* 2003;15(3):232–8. [PubMed: 12786644]
- [38]. McConnell KJ, Richards CF, Daya M, et al. Effect of increased ICU capacity on emergency department length of stay and ambulance diversion. *Ann Emerg Med* 2005;45(5):471–8. [PubMed: 15855939]
- [39]. Solberg LI, Asplin BR, Weinick RM, et al. Emergency department crowding: consensus development of potential measures. *Ann Emerg Med* 2003;42(6):824–34. [PubMed: 14634610]
- [40]. Asplin BR. Measuring crowding: time for a paradigm shift. *Acad Emerg Med* 2006;13(4):459–61. [PubMed: 16581934]
- [41]. Sprivulis P, Gerrard B. Internet-accessible emergency department workload information reduces ambulance diversion. *Prehosp Emerg Care* 2005;9(3):285–91. [PubMed: 16147477]

Figure.

Actual (x-axis) versus predicted (y-axis) occupancy levels at site A. The data represent forecast lengths of 2 hours (top left), 4 hours, (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point in time. A small amount of uniformly distributed random noise was added to the x-axis; this was necessary to improve legibility since observed patient counts are constrained to integers. The time series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

Table 1

General characteristics of the participating validation sites.

	Site A	Site B	Site C [†]	Site D	Site E
<i>Medical center factors</i>					
Inpatient capacity (# of beds)	702	425	520	1,171	620
Critical care capacity (# of beds)	92	50	70	133	77
Trauma service (accreditation)	level 1	level 1	level 1	N/A	level 1
Population served (adult/pediatric)	adult	both	adult	both	adult
Metropolitan setting (urban/rural)	urban	urban	urban	urban	urban
<i>Emergency department factors</i>					
Licensed capacity (# of beds)	25	35	25	34	39
Observation unit capacity (# of beds)	8	N/A	N/A	5	8
Triage system categories (# of levels)	4	5 (ESI)	5 (ESI)	5 (ESI)	5 (ESI)
Attending physicians (# on staff)	30	30	18	33	37
<i>Operational data for 2006[*]</i>					
Total volume (# of patients)	54,611	62,219	40,193	53,904	49,794
Diversion frequency (% of time)	7.0	1.0	0.3	3.6	2.4
Proportion eloped [‡] (% of patients)	5.2	3.2	0.3	3.2	1.4
Proportion admitted (% of patients)	22.5	19.1	22.8	30.5	34.2
Waiting count (# of patients)	6 (2, 12)	3 (1, 7)	0 (0, 1)	3 (1, 6)	2 (1, 5)
Occupancy level (% of beds)	91 (73, 106)	77 (57, 91)	52 (32, 72)	85 (59, 113)	79 (62, 98)
Boarding count (# of patients)	8 (5, 11)	2 (1, 3)	1 (0, 2)	9 (5, 14)	12 (9, 16)

* Operational data are presented as counts, percentages, or median (interquartile range) as appropriate.

[†] All operational data from site C were calculated based on 10 months of patient visits (3/1/2006 – 12/31/2006), and the total volume was adjusted by a factor of 6/5 to extrapolate for the calendar year.

[‡] The proportion eloped is defined as the percentage of patients who register in the ED and leave prior to being assessed by a physician.

Table 2
Reliability of the simulation versus autocorrelation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
<i>Waiting count</i>					
Simulation (r)	0.71 (0.70, 0.71)	0.52 (0.51, 0.53)	0.15 (0.14, 0.16)	0.56 (0.56, 0.57)	0.65 (0.64, 0.66)
Autocorrelation (r)	0.50 (0.49, 0.51)	0.45 (0.44, 0.46)	0.08 (0.06, 0.09)	0.41 (0.40, 0.42)	0.44 (0.44, 0.45)
<i>Occupancy level</i>					
Simulation (r)	0.78 (0.77, 0.78)	0.80 (0.80, 0.81)	0.77 (0.77, 0.77)	0.79 (0.79, 0.79)	0.83 (0.83, 0.83)
Autocorrelation (r)	0.52 (0.51, 0.52)	0.46 (0.45, 0.47)	0.37 (0.36, 0.37)	0.61 (0.61, 0.62)	0.60 (0.60, 0.61)
<i>Boarding count</i>					
Simulation (r)	0.72 (0.72, 0.73)	0.41 (0.40, 0.41)	0.26 (0.25, 0.27)	0.70 (0.70, 0.71)	0.57 (0.57, 0.58)
Autocorrelation (r)	0.73 (0.73, 0.74)	0.48 (0.47, 0.49)	0.17 (0.16, 0.18)	0.65 (0.64, 0.65)	0.68 (0.68, 0.68)

*The Pearson coefficient of correlation is presented with lower and upper bounds of the 95% confidence interval in parentheses.

Table 3

Calibration of the simulation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
Waiting count (# of patients)	0.6 ± 6.9	-2.5 ± 4.1	-0.3 ± 1.2	3.6 ± 9.2	1.2 ± 5.2
Occupancy level (% of beds)	-0.3 ± 15.4	-0.4 ± 13.4	7.0 ± 16.9	-5.6 ± 23.5	2.6 ± 13.3
Boarding count (# of patients)	0.2 ± 3.0	2.0 ± 1.9	1.0 ± 1.5	-1.2 ± 4.4	-2.0 ± 4.4

*The forecasting residuals are summarized with the mean ± standard deviation.

Table 4

Absolute error of the simulation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
Waiting count (# of patients)	3.1 (1.6, 5.1)	2.4 (1.3, 3.7)	0.6 (0.4, 0.7)	2.0 (1.0, 3.2)	1.6 (0.8, 2.7)
Occupancy level (% of beds)	10.4 (4.9, 17.5)	9.0 (4.3, 15.4)	11.1 (5.2, 19.1)	14.5 (6.9, 25.0)	9.0 (4.3, 15.4)
Boarding count (# of patients)	1.9 (0.9, 3.2)	1.0 (0.5, 1.6)	0.9 (0.4, 1.2)	2.8 (1.3, 4.7)	2.7 (1.3, 4.5)

*The median absolute error is presented with the interquartile range in parentheses.