Forecasting Models of Emergency Department Crowding

Lisa M. Schweigler, MD, MPH, MS, Jeffrey S. Desmond, MD, Melissa L. McCarthy, ScD, Kyle J. Bukowski, MBA, BSIE, Edward L. Ionides, PhD, and John G. Younger, MD, MS

Abstract

Objectives: The authors investigated whether models using time series methods can generate accurate short-term forecasts of emergency department (ED) bed occupancy, using traditional historical averages models as comparison.

Methods: From July 2005 through June 2006, retrospective hourly ED bed occupancy values were collected from three tertiary care hospitals. Three models of ED bed occupancy were developed for each site: 1) hourly historical average, 2) seasonal autoregressive integrated moving average (ARIMA), and 3) sinusoidal with an autoregression (AR)-structured error term. Goodness of fits were compared using log likelihood and Akaike’s Information Criterion (AIC). The accuracies of 4- and 12-hour forecasts were evaluated by comparing model forecasts to actual observed bed occupancy with root mean square (RMS) error. Sensitivity of prediction errors to model training time was evaluated, as well.

Results: The seasonal ARIMA outperformed the historical average in complexity adjusted goodness of fit (AIC). Both AR-based models had significantly better forecast accuracy for the 4- and the 12-hour forecasts of ED bed occupancy (analysis of variance [ANOVA] p < 0.01), compared to the historical average. The AR-based models did not differ significantly from each other in their performance. Model prediction errors did not show appreciable sensitivity to model training times greater than 7 days.

Conclusions: Both a sinusoidal model with AR-structured error term and a seasonal ARIMA model were found to robustly forecast ED bed occupancy 4 and 12 hours in advance at three different EDs, without needing data input beyond bed occupancy in the preceding hours.

ACADEMIC EMERGENCY MEDICINE 2009; 16:1–8 © 2009 by the Society for Academic Emergency Medicine

Keywords: crowding, forecasting, emergency service, hospital, operations research

Emergency department (ED) overcrowding has become a significant problem throughout the United States, leading to possible increased health care costs, causing raised stress levels among staff and patients in EDs, and most importantly, adversely affecting patient outcomes.1–9 One aspect of the problem is the difficulty of anticipating the timing and magnitude of overcrowded conditions. The ability to predict crowded conditions, especially hour by hour, could substantially impact ED operations. To this end, we evaluate how time series–based models perform in short-term forecasting of ED occupancy.

Traditionally, ED operations directors have found historical averages to be reliable and accurate for long-term forecasts of ED behavior. For example, a director might use the average ED bed occupancy on Monday evenings at 21:00 over the past 2 years to determine how many staff should be working in the ED at that time. However, short-term forecasting of ED bed occupancy, such as might be useful for calling in additional staff or opening up hospital beds, is likely to need more accurate forecasting techniques.

Several authors have looked to time series techniques, such as autoregression (AR) models, as potentially useful tools in forecasting ED behavior (e.g., patient volume or arrivals, length of stay, or patient acuity) without needing the input of many different

From the Departments of Emergency Medicine (LMS, JSD, JGY) and Statistics (ELI), University of Michigan, Ann Arbor, MI; the Department of Emergency Medicine, The Johns Hopkins University School of Medicine (MLM), Baltimore, MD; and Administrative Consulting, William Beaumont Hospital (KJB), Royal Oak, MI.


This work was supported in part by a pilot grant from the University of Michigan Center for Computational Medicine and Biology (JGY) and by the Robert Wood Johnson Foundation, with whom LMS is a clinical scholar.

Received July 8, 2008; revisions received September 25 and November 11, 2008; accepted November 12, 2008.

Address for correspondence and reprints: Lisa M. Schweigler, MD, MPH, MS; e-mail: lschweig@umich.edu.

© 2009 by the Society for Academic Emergency Medicine
doi: 10.1111/j.1553-2712.2009.00356.x

ISSN 1069-6563
PII ISSN 1069-6563583
predictor variables. The premise of these models is straightforward: an ED’s level of activity in the near future is strongly correlated to its activity now. These studies show that, in general, time series methods provide better statistical fit than more traditional approaches such as multivariate linear regression or historical experience. However, most time series studies of ED behavior have remarked on the ability of time series models to closely fit past events; performance against future behavior has not typically been included. Furthermore, time series approaches have not yet been used to directly investigate ED crowding, instead modeling related behaviors such as patient arrivals per hour. To our knowledge, only one group has studied forecasting related behaviors such as patient arrivals per hour. To our knowledge, only one group has studied the effectiveness of time series methods for predicting future behavior, and in that work they focused on total daily occupancy, rather than hourly forecasts such as might be useful for “higher resolution” real-time operations management.

We contend that three chief requirements of a useful ED crowding forecasting model are 1) that it can be used at different EDs with varying operations environments, 2) that it performs significantly better than the hourly historical average, and 3) that it requires the smallest amount of information possible with which to make predictions. Although large multivariate models no longer constitute a computationally challenging problem, for short-term forecasting purposes they would require a continuous supply of high-fidelity data, often streaming from different administrative units (e.g., hospital bed control, operating rooms). At present, departments with access to such dense operational informatic resources are rare. Our requirement of parsimony of information led us to choose ED occupancy as our crowding metric over other more complex metrics such as the National Emergency Department Overcrowding Scale (NEDOCS) or the Emergency Department Work Index (EDWIN).

In this study, we addressed the following questions: How do AR models perform compared to the hourly historical average in forecasting (up to 12 hours into the future) the occupancy of an ED? Furthermore, do some models perform better at one institution than another? Is there a model that performs sufficiently well regardless of the ED toward which it is applied such that it might constitute a standard? Finally, how far back in time should a model look to generate the best forecasts? Too short of a training time may impair performance by generating imprecise parameter estimates. Too long of a training time may prevent a model from adapting to very recent changes in occupancy behavior. The models were evaluated for the accuracy of their 4- and 12-hour forecasts for a year’s worth of Monday evenings at three large teaching hospitals, which tend to be the most crowded times for many EDs.

METHODS

Study Design and Setting
We conducted a multicenter retrospective analysis of hourly clinical activity at three adult EDs (Site 1 annual census, 98,199; Site 2 annual census, 59,344; and Site 3 annual census, 55,757). No patient- or provider-level identifying information was included, and therefore the study was considered exempt from informed consent requirements by the institutional review boards at all three sites.

Study Protocol
Hourly occupancy was defined as the number of patients within each adult ED and its waiting room divided by the number of permanent beds (excluding makeshift hallway beds, chairs, etc.) in that ED available during the hour in question. Patients who ultimately left before being evaluated were included in the counts while they were still registered as being in the ED. The hourly denominator was corrected for circumstances when greater or lesser numbers of beds were available in each ED (e.g., when ED-adjacent clinic space became available after normal clinic hours). Occupancy values for the adult EDs were collected retrospectively from each institution’s clinical information system for the period beginning midnight, July 1, 2005, and ending 11:00 pm, June 30, 2006, resulting in 8,760 sequential hourly occupancy values for each center.

Evaluation of different statistical models was directed at their goodness of fit and their ability to make forecasts from 15:00 Monday through 02:00 Tuesday for 51 of the 52 Mondays included in the data set. These were times when all three sites frequently experienced occupancy levels that were higher and less predictable than at other times during the week and thus represented a stringent test platform.

Data Analysis
Analysis was conducted in R 2.7.1 (Comprehensive R Archive Network, http://cran.r-project.org) and Matlab R2008a (The Mathworks, Inc., Natick, MA). Prior to building AR-based models, diagnostic time domain analyses were performed to identify dominant frequencies within each site’s occupancy behavior (data not shown). As discussed under Results, 24-hour periodicity was the primary mode at each site, and subsequent time domain models were limited to this frequency.

Following these preliminary model checks, we evaluated in detail three models of ED crowding, including the hourly historical average and two autocorrelation models. The specifications of the models are included in Data Supplement S1 (available as supporting information in the online version of this paper). In brief, they were the historical average, which is the mean occupancy for each site for each hour of the day: a 24-hour seasonal model (seasonal autoregressive integrated moving average [ARIMA] (1,0,1)/(0,1,1)), where occupancy at any time is a function of occupancy both 1 and 24 hours prior, and a sinusoidal model with an AR-structured error term, where occupancy at any time is a function of a 24-hour period sine wave fit to each ED’s diurnal pattern and combined with 1-hour AR. The standard descriptive notation for ARIMA models is ARIMA(p,d,q), where p denotes the number of autoregressive parameters, d is the number of differencing passes, and q is the number of moving average parameters. A seasonal ARIMA is described by
ARIMA\((p,d,q)⁄(sp,sd,sq)\) in which \(sp, sd,\) and \(sq\) provide the additional information on the seasonal autoregressive, differencing, and moving average components of the model, respectively. The two AR-based models were specifically chosen because they account most parsimoniously for both the 24-hour periodicity of ED occupancy behavior and the strong predictiveness of a previous hour’s occupancy on the next hour’s occupancy.

Each model was evaluated in two ways, as summarized in Figure 1. Goodness of fit was evaluated retrospectively using log-likelihood values across the ensemble of 51 Monday evenings in the data set by maximum likelihood regression to the 168 hours (7 days) prior to Monday, \(t = 15:00\). Details of these calculations are also included in Data Supplement S1.

The second means of evaluating model performance was to consider prospective accuracy. As represented graphically in Figure 1, each model was trained on a defined number of hours of prior ED occupancy (annotated as goodness-of-fit domain in the legend to Figure 1) and then allowed to generate a forecast of ED occupancy for a subsequent number of hours (forecast domain). To build the models, we only used observed ED occupancy from the training period; no data from the subsequent prediction period were used in building the predictions. Thus, the forecasting performance was prospectively evaluated in a virtual manner from previously collected data.

Forecast accuracy was examined over 51 consecutive Monday evenings for all three sites over the study year. A forecast was defined as a prediction of ED occupancy either 4 or 12 hours beyond the available data, which in each case was artificially cut off at \(t = 14:00\) for each study day (15:00 was therefore the first hour of forecast). Accuracy was quantified by comparing the predicted occupancy to the actual occupancy during the forecast and calculating the error as the root mean forecast sum of squares,

\[
\varepsilon_{RMS} = \left( \frac{1}{K} \sum_{k=1}^{K} (x_k - \mu_k)^2 \right)^{1/2}
\]

where \(k\) is each hour of a \(K\)-hour forecast, \(x_k\) is the actual occupancy, and \(\mu_k\) is the model-predicted occupancy.

To determine the impact of duration of training time (i.e., the number of hours of occupancy behavior provided to a model to allow predictions), a series of 4- or 12-hour forecasts of occupancy from 15:00 Monday to 02:00 Tuesday were made with an increasingly greater number of training hours, from 168 hours (7 days) to 336 hours (14 days). Forecast root mean square (RMS) error was calculated as described previously, and the mean RMS errors for each site were determined for each training period.

RESULTS

Table 1 summarizes key operational characteristics of the three study sites during the study period July 2005–June 2006, both at the ED and at the hospital level. Appreciable differences are seen in total ED volume, number of ED patients per ED bed per year, number of available inpatient beds, average weekday adult inpatient bed occupancy, percentage of days in study period with inpatient bed occupancy greater than 95%, attending and resident staffing hours, left-before-evaluation rates, size of observation unit, and percentage of patients seen in a minor care area.

A summary of the occupancy data at the three sites is shown in Figure 2. The clinical activity at each site is depicted as a heat map scaled over each day or over each week of the study frame. These images show that 1) the occupancy patterns differ between the three institutions and 2) all three institutions show diurnal variation in bed occupancy, with the busiest times occurring later in the day.

![Figure 1. General analytical strategy. For each study center, data were segmented into 1-week segments. For each segment, three statistical models were examined for goodness of fit and for 4- and 12-hour forecast accuracy. From this ensemble of model fits and model forecasts, summary statistics including log likelihood, Akaike’s Information Criteria (AIC), and root mean square (RMS) accuracy forecasts were generated.](image-url)
Table 2 shows the parameter estimates for the AR-based models. Occupancy means calculated for the historical averages model are not reported.

Goodness-of-fit measures are shown in Table 3. The historical average model consistently produced the best goodness of fit. However, the seasonal ARIMA \((1,0,1)/(0,1,1)\) performed best according to Akaike’s Information Criterion (AIC), which penalizes for increased model complexity. In this case, the historical average model included 24 model parameters (1 for each hour of the day), while the sinusoidal model with autocorrelated error required only 4 to make its
Forecasts (an AR term, a sine coefficient, a cosine coefficient, and an intercept).

Forecast performance is summarized in Figure 3. The box plots depict accuracy, for each site and each model, over either 4 or 12 hours of prediction as the RMS of the summed residuals between observed and predicted occupancy values. These plots show the increase in accuracy achieved by moving from a simple historical average system to more sophisticated models. For each site, the accuracies of the three methods were compared with one-way analysis of variance (ANOVA), followed by post hoc comparisons with Tukey-Kramer statistics. For each site, the two AR-based models outperformed the historical average. In post hoc testing, no differences were noted between the sinusoidal-AR model and the seasonal ARIMA model at each site.

Examination of the effect of training time on forecast accuracy revealed no significant benefit beyond 168 hour (1 week) training periods (data not shown).

**DISCUSSION**

In our study, we show that AR models with seasonal/sinusoidal adjustment consistently outperform the historical average in short-term forecasting of ED occupancy up to 12 hours in advance and do so at several different institutions. Although there is variability in model performance for any given Monday, the reductions in error are potentially important operationally. For example, in moving from a historical average to either the seasonal ARIMA model or the sinusoidal model with AR-structured error, Site 2 would see a roughly 33% improvement in its 12-hour forecasts of ED crowding. We therefore posit that AR-based models should constitute the standard for predictive models, using time-series approaches and ED occupancy as the crowding metric. We found that a training time of 1 week (168 hours) was sufficient to build a model with excellent reliability.

It is not surprising that the two autoregressive models outperformed the historical average: while the historical average is an easy-to-understand approach to predicting long-term future ED volume and occupancy and has a well-founded theoretical basis in queuing theory, it cannot be expected to perform well in situations where there is frequent irregularity in short-term behavior, e.g., unusually high ED occupancy on a Monday night, or increased demand on ED services during a virulent cold and flu season.

The ED literature describes many possible ED crowding metrics, such as staff opinion, leaving-before-evaluation rates, amount of time on ambulance diversion, ED length of stay, or calculated scores such as NEDOCS or EDWIN. However, many of these measures of ED crowding are not easily obtained from EDs that do not systematically make an effort to collect such information. Most EDs do keep records on when patients present to the department and when they depart. From this operational information it is straightforward to calculate ED bed occupancy, defined as number of patients in the ED over number of permanent treatment bays available to that ED. It has been shown that the measure of ED bed occupancy performs no worse than more complex scores such as EDWIN in identifying ED crowding.

Another important consideration in the development of this study was how to interpret the ED bed occupancy metric: should it be treated as a continuous metric or should a threshold approach be used, in which either an ED is crowded or it is not? While some important ED performance metrics may become abnormal at an easily discerned threshold occupancy level,
a crowding metric that can supply a universally applicable threshold of “this ED is now crowded” remains to be developed. The value of picking a dichotomous outcome (e.g., crowded or not, on ambulance diversion or not) is attractive in that it permits evaluating forecast strategies with receiver operating characteristic approaches, but risks limiting generalizability of the forecasting scheme. The goal of our study was not to predict when a participating ED would reach a specific threshold of crowding, but rather to formulate predictions of different occupancy levels. It remains up to the individual institutions using any of the ED crowding metrics to interpret the meaning of the values obtained. Once such a level is established, our results indicate that an AR-based forecasting rule would perform well.

We considered possible applications of short-term forecasts of ED occupancy. For example, some institutions have successfully implemented a “dashboard” approach, in which ED and hospital administrators can make immediate patient flow and resource allocation decisions based on real-time ED and hospital operations data displayed on a computer interface (the dashboard), such as ED volume and hospital bed occupancy. The addition of accurate and frequent short-term forecasts of ED crowding as a dashboard tool could be invaluable in helping administrators mitigate crowded conditions. Short-term forecasts of ED crowding could also prove valuable in regionalizing ambulance traffic. In its examination of the current state and the future of emergency care in the United States, the 2006 Institute of Medicine Report “Future of Emergency Care” called for refinement of methods to enable regional coordination of patient flow between different EDs to help alleviate crowding. Complementary cornerstones of effective regionalization would be up-to-the-minute knowledge of crowding across EDs and a reliable means of predicting their status in the near future. The latter point is critical; delivering

Figure 3. Twelve-hour forecast accuracy for the three models, by study site. The differences between predicted and actual ED occupancy from 15:00 Monday to 02:00 Tuesday for 51 weeks were quantified using the root-mean sum of squares. Models include HA (historical average), SAR (seasonal autoregressive integrated moving average [ARIMA] (1,0,1)/(0,1,1)), and sinusoidal with AR-structured error term (AR-S). One-way analysis of variance (ANOVA) was used to compare the different methods at each site and returned a p-value of < 0.01 for each instance. Post hoc comparisons were performed with the Tukey-Kramer method. The p-values for these results are shown in the figures.
patients to an ED that is currently less than fully occupied but is likely to become so in the near future may not be the most effective triage choice.

Investigators have studied approaches other than historical averages and time-series analysis to forecasting ED behavior. The recent literature discusses numerous methods ranging from multivariable regression analysis to nonlinear techniques, discrete event simulation, and neural networks.\(^\text{11,19,27,30–35}\) As reported, all of these approaches function reasonably well in providing short-term forecasts of various lengths for a variety of ED operational characteristics. However, many of these models may use proprietary software and often require input of many operational variables, some from outside the ED, to generate their forecasts. The AR-based models shown to perform well in our analysis do not have these problems—we demonstrated that they require only one input variable, and they use widely available open-source software (R 2.7.1, Comprehensive R Archive Network, http://cran.r-project.org).

**LIMITATIONS**

It is important to note that the evaluation of the models in this study may be limited by the operational similarities between the three sites studied. All were relatively large, tertiary academic referral institutions. However, the hospitals under consideration are located in varied socioeconomic surroundings and therefore are likely to have different demands placed on them at different times in terms of patient presentations and bed availability. Table 1 shows that there are large differences in some operational characteristics between the three study sites. As seen in Figure 2, the three sites were found to have clear differences in the magnitude of both occupancy and variability of occupancy. Despite these differences in operational characteristics of the three EDs, the models performed similarly relative to each other at all three sites, adding weight to the argument that they might also work in a similar way at other comparable institutions. We would emphasize, however, that extrapolating these models to very different departments, and in particular low-volume sites where departmental occupancy may be zero over several sequential hours, should be done with care.

The premise of this study was to develop AR-based models of ED occupancy that are parsimonious and applicable to any ED. As with any modeling problem, the modeler faces a trade-off between model complexity, practicality, and generalizability. In this study, it is possible that higher-order ARIMA models may have provided even further reduction in the forecasting error than achieved with our approach. Indeed, a department wishing to undertake a systematic statistical consideration of this problem might well explore a large ensemble of related models, possibly incorporating locally available real-time operational covariates to determine which is optimal in their setting. The models explored here are not the final answer, but a reasonable platform from which to move forward.

We evaluated the performance of our models in two ways: model goodness of fit via log likelihood and the AIC and actual forecast performance via ANOVA of the RMS error of the different approaches. The goodness-of-fit measurements did not conclusively favor the AR-based models, but in this instance their interpretation is subtle. Specifically, while the historical average appears to be a very simple model (in that one could easily calculate it by hand), in actuality it is one including 24 separate parameters that must be fit to observed data. As a result, it is significantly advantaged in the calculation of log likelihood and similarly disadvantaged in calculation of AIC, which rewards model simplicity. However, this discussion is in part academic: the AR-based models performed clearly better than the historical average in their forecast accuracy, which is arguably a much more meaningful metric to individuals responsible for clinical operations than goodness of fit.

The models developed in this study were designed to be tools for operations managers that might help them decide when to institute interventions to mitigate ED crowding in their individual institution or regional ambulance network. The model predicts ED occupancy but does not provide insight into causes or consequences, nor would it be expected to shed light on site-specific solutions. However, as the model provides a parameterization of occupancy trends over time, it could readily be implemented as a statistical instrument for evaluating operational changes whose effects might be more complex than simply reducing absolute occupancy.

**CONCLUSIONS**

Using only preceding ED bed occupancy as input, AR-based models with seasonal or sinusoidal adjustment generated robust short-term forecasts of subsequent ED bed occupancy. This forecasting method was found to work equally well at three different institutions with differing operational characteristics, without having to adjust any of the model input variables. The simplicity of this approach makes it attractive for implementation in various applications such as regional out-of-hospital, ED, and hospital operations.

**References**


5. Fee C, Weber EJ, Maak CA, Bacchetti P. Effect of emergency department crowding on time to


Supporting Information

The following supporting information is available in the online version of this paper:

Data Supplement S1. Supplemental material.

The document is in PDF format.

Please note: Wiley Periodicals Inc. is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.
Supplemental Material

I. Statistical models employed

The hourly historical average $\mu_t$ was defined as

$$\mu_t = \frac{1}{N} \sum_{i=1}^{N} x_{t-24i}$$  \hspace{1cm} (eq. 1)

where $N$ is the number of previous days included in the analysis and $x_{t-24i}$ is the observed occupancy value at the $i^{th}$ hour some $i$ days previously.

The seasonal autoregressive integrated moving average (ARIMA) $(1,0,1)/(0,1,1)$ model including a 24-hour seasonal component was defined as

$$x_t - \phi x_{t-1} - x_{t-24} + \phi x_{t-25} = w_t + \Theta w_{t-24} + \Theta w_{t-25}$$  \hspace{1cm} (eq. 2)

where the 1-hour lag coefficient $\phi \neq 0$ and the 24-hr seasonal coefficient $\Theta \neq 0$.

Lastly, given a priori knowledge of strong 24-hour periodicity of ED occupancy behavior, we included a simple sinusoidal model with autoregressive (AR)-structured error,

$$x_t - \beta_1 \sin(t) - \beta_2 \cos(t) - \phi [x_{t-1} - \beta_1 \sin(t-1) - \beta_2 \cos(t-1)] = w_t$$  \hspace{1cm} (eq. 3)

where $\phi$ is the 1-hour lag coefficient. This form takes advantage of the trigonometric identity between a fully specified sine function, $\alpha \sin(t + \varphi)$, which is not amenable to linear regression methods, and $\beta_1 \sin(t) + \beta_2 \cos(t)$, which is. The $\beta$ coefficients in the latter term can be converted to the wave amplitude $a$ in the former equation by $\alpha = \sqrt{\beta_1^2 + \beta_2^2}$. The phase angle $\varphi$ can likewise be recovered by $\varphi = \tan^{-1}\left(\frac{\beta_1}{\beta_2}\right)$.

II. Determination of Goodness of Fit Metrics
For the historical average model, the likelihood \( L \) and log-likelihood were derived from the standard Gaussian form as follows:

\[
L = \prod_{j=1}^{24} \prod_{k=1}^{K} \left(2\pi\sigma_j^2\right)^{-1/2} \exp \left(\frac{(x_{j,k} - \mu_j)^2}{2\sigma_j^2}\right)
\]

\[\log(L) = \sum_{j=1}^{24} \left( K \log\left(\frac{1}{\sigma_j\sqrt{2\pi}}\right) + \sum_{k=1}^{n} \frac{(x_{j,k} - \mu_j)^2}{2\sigma_j^2}\right) \quad \text{(eq. 4)}\]

where \( 1 \leq j \leq 24 \) is the hour of the day, \( K \) is the number of days under study, \( x_{j,k} \) is an individual hourly occupancy value, and \( \mu \) and \( \sigma \) are the mean and standard deviation for occupancy values at time \( j \). The Akaike Information Criterion (AIC) was calculated from the above in the usual fashion

\[
AIC = 2p - 2\log(L) \quad \text{(eq. 5)}
\]

where \( p \) is the number of model parameters, in this case 24.¹

The parameter estimates, log-likelihood values, and AIC values for the two AR-based models were provided by the `arima()` routines, which employ the Kalman filter method, in R 2.7.1 (Comprehensive R Archive Network, http://cran.r-project.org). The likelihood functions for ARIMA models are too complex to be shown. Three parameters (autoregression term, moving average term, and 24-hour seasonal term) were included for the seasonal ARIMA \((1,0,1)/(0,1,1)\) model, and four for the sinusoidal model with AR-structured error term (autoregression term, intercept, cosine component, and sine component). Table 2 in the main manuscript shows the parameter estimates.
 References

Real-Time Identification of Serious Infection in Geriatric Patients Using Clinical Information System Surveillance

William J. Meurer, MD,* Barbara L. Smith, BA,* Eve D. Losman, MD,* Diana Sherman, RN,* Joseph D. Yaksich, RN, ASN,* Jeremy D. Jared, BS,‡ Preeti N. Malani, MD, MSJ,§ and John G. Younger, MD, MS*∥

OBJECTIVES: To develop and characterize an automated syndromic surveillance mechanism for early identification of older emergency department (ED) patients with possible life-threatening infection.

DESIGN: Prospective, consecutive-enrollment, single-site observational study.

SETTING: A large university medical center with an annual ED census of 75,273.

PARTICIPANTS: Patients aged 70 and older admitted to the ED and having two or more systemic inflammatory response syndrome (SIRS) criteria during their ED stay.

MEASUREMENTS: A search algorithm was developed to screen the census of the ED through its clinical information system. A study coordinator confirmed all patients electronically identified as having a probable infectious explanation for their visit.

RESULTS: Infection accounted for 28% of ED and 34% of final hospital diagnoses. Identification using the software tool alone carried a 1.63 relative risk of infection (95% confidence interval CI = 1.09–2.44) compared with other ED patients sufficiently ill to require admission. Follow-up confirmation by a study coordinator increased the risk to 3.06 (95% CI = 2.11–4.44). The sensitivity of the strategy overall was modest (14%), but patients identified were likely to have an infectious diagnosis (specificity = 98%). The most common SIRS criterion triggering the electronic notification was the combination of tachycardia and tachypnea.

CONCLUSION: A simple clinical informatics algorithm can detect infection in elderly patients in real time with high specificity. The utility of this tool for research and clinical care may be substantial. J Am Geriatr Soc 57:40–45, 2009.

Key words: sepsis; aged; sentinel surveillance; informatics

Emergency departments (EDs) are an important source of health care for elderly people. Data from the National Hospital Ambulatory Medical Care Survey from 2003 have described high utilization of EDs by older Americans, with infections being among the most frequent and most serious reasons for these visits and the need for subsequent hospitalization.¹ Unscheduled admission to the hospital is an important prognostic factor for elderly patients. Estimates suggest that, of elderly patients with chronic conditions who survive hospitalization, 33% will die within 6 months.² Although infection is common and frequently serious, its presentation in older patients is notoriously difficult to detect. Senile changes in the immune system and the inflammatory response serve to mute classic symptoms and signs of invasive bacterial processes.³ For clinical care and for clinical research, a hospital environment wherein caregivers are personally unfamiliar with patients and where levels of activity are often near, at, or even above 100% facility capacity seriously compounds subtle presenting signs and symptoms in this population.

Improved strategies for early detection of serious infection in older emergency patients would be beneficial for several reasons. Earlier detection during ED evaluation presumably would shorten delays to comprehensive evaluation, antibiotics, and in some instances advanced monitoring and goal-directed therapy. Just as valuable would be a means of rapidly and reliably identifying such patients to recruit them for observational and interventional studies in the ED. EDs are increasingly challenging environments in which to comprehensively identify and recruit subjects for clinical research. Methods of identifying serious infection in
elderly people would be expected to improve clinical research in this population as well as improve care.

The current report describes and characterizes a novel strategy for identifying elderly ED patients likely to have serious infections. The method was built around a simple algorithm that continuously interrogated a clinical information system for abnormal vital signs and laboratory data in older patients. Flagged patients were electronically brought to the attention of a study coordinator (BLS) who completed patient evaluation at the bedside.

**MATERIALS AND METHODS**

This was a prospective, consecutive-enrollment, single-site observational study performed at the ED of a large university medical center with an annual census of 75,273.

**Study Protocol**

**ED Informatics Structure**

The clinical information system used in this study is shown in Figure 1. In brief, the system was built on SQL 2000 Enterprise (Microsoft Corporation, Redmond, WA). The user interface through which clinical information was entered manually was Centricity 7.5.x (General Electric Healthcare, Piscataway, NJ). Vital signs were also entered automatically using a monitor capture server interfaced to the department’s bedside patient telemetry units. Laboratory data were transmitted automatically from the clinical laboratory’s server.

**Automated Surveillance Algorithm**

The surveillance protocol was developed as a tool for an ongoing study of infection in elderly people. Inclusion criteria for that study included patient age of 70 and older, time from ED admission to identification as a subject in the study of no more than 6 hours, and two or more systemic inflammatory response syndrome (SIRS) criteria accumulated at any time during the ED stay (respiratory rate ≥20/min, temperature ≤36 or ≥38°C, heart rate ≥90 betas/min, and total white blood cell count ≤4,000 or ≥12,000/mm³).³

These criteria were used to develop a rules engine that would execute a scheduled query of the ED’s SQL server every 15 minutes from 9:00 a.m. to 4:00 p.m. every weekday. Whenever age and SIRS inclusion criteria were met, an alphanumeric page (electronic notification) containing a patient identifier, the patient’s location within the ED, and which inclusion criteria had been met was sent to a study coordinator. The coordinator was responsible for the second element of patient identification, a confirmation with the physician in charge that a noninfectious explanation for the patient’s condition was not present. Although this criterion appears circuitous, it was designed to avoid physicians’ preconceptions as to what might or might not

![Figure 1. Informatics architecture used in the current study. The patient identification strategy was centered on the department’s SQL server. An admission, discharge, and transfer (ADT) server drives the database of current patients. Once the ADT server placed a patient’s identifiers into the SQL database, that patient was under surveillance for the duration of his or her stay. Systemic inflammatory response syndrome criteria could enter the system manually (through emergency department (ED) staff by way of the user interface), through automated capture of patient telemetry, or through information from the medical center’s clinical laboratory server (for white blood cell count information). The rules engine executed its query every 15 minutes from 9:00 a.m. to 4:00 p.m. on weekdays. When candidates were identified, pertinent data were sent to the study coordinator through the medical center’s alphanumeric paging system.](image-url)
constitute a serious infection or even sepsis in the study population. From this point forward, this criterion will be stated as “suspected infection” for the sake of clarity.

Measurements
The strategy was characterized in two phases. From July 1 through August 15, 2007, the validity of the strategy as a classification tool was evaluated in all patients aged 70 and older presenting between 3:00 a.m. and 4:00 p.m. on weekdays. Because the enrollment criteria allowed a patient to be identified as long as 6 hours after presentation, the 9:00 a.m. start time could capture patients presenting as early as 3:00 a.m. Structured chart review of this population was conducted to confirm or exclude an ED diagnosis of infection and a final hospital discharge diagnosis of infection. Three trained chart abstracters whose interrater agreement was assessed using the Fleiss kappa score (0.81) conducted reviews. The performance of the surveillance algorithm and the combination of the algorithm followed by confirmation by the study coordinator were quantified using the sensitivity, specificity, and relative risk of an affirmative screen for identifying an ED diagnosis of infection, hospital discharge diagnosis of infection, and the need for intensive care unit admission. To determine short-term (90-day) mortality, electronic medical records of the study hospital and federal Social Security mortality data were examined. For the latter, records were not studied for at least 9 months after enrollment to maximize the probability of capturing all deaths within the cohort. Survival analysis of patients identified by the algorithm, compared with all other patients aged 70 and older presenting during the study window, was performed using standard proportional hazards analysis.

The strategy was also characterized according to the time required, from presentation, time of triage, and the time that a patient was placed in an ED bed, for inclusion criteria to be met. Presumably, whenever criteria were present, they would be detected at the time of the next executed query, which was never more than 15 minutes distant. Rather, the delay from presentation until detection is a reflection of the time required for sufficient signs of infection to accumulate in these patients.

The second phase of characterization occurred between July 1 and November 15, 2007, and was limited to patients in whom automated detection occurred; patients from the first phase who met two or more SIRS criteria were included. The goal of this phase was to examine the frequency and utility of the observed combinations of SIRS criteria. In these individuals, the association between the particular SIRS criteria met and the likelihood that a suspected infection was present was evaluated using log odds.

Statistical Analysis
All analysis was performed in R 2.6.0 (The Comprehensive R Archive Network, www.r-project.org). The epibasix package was used to calculate relative risk, sensitivity, and specificity, along with appropriate confidence intervals. The irr package was used to calculate the Fleiss kappa.

The local institutional review board approved all protocols.

RESULTS
Between July 1 and August 15, 2007, 583 patients aged 70 and older were evaluated during the daily study window, 248 of these were admitted and included for further analysis: of these, 69 (28%) had an ED diagnosis of infection of any kind, and 84 (34%) had a final hospital discharge diagnosis of infection. The clinical characteristics of all patients with an ED infection diagnosis (regardless of whether they were detected by the system) are shown in Table 1.

An electronic notification was generated in 61 (25%) of those admitted, and 13 (5%) of those admitted had a suspected infection. The combination of electronic notification and suspected infection was a strong indicator of overall degree of illness; only one patient meeting these criteria was not admitted to the hospital. This patient left against medical advice and subsequently returned and was admitted.

The surveillance system correctly identified all potentially eligible patients. The clinometric performance of the overall identification strategy is reported in Table 2. Patients triggering the SIRS rules engine were approximately 1.6 times as likely to have an infectious diagnosis as those who did not trigger this system. Patients initially identified by the surveillance system who also had suspected infection were three times as likely to have an ED diagnosis of infection as those not triggering the system (P < .05).

To determine the extent to which the detection algorithm identified patients who were truly seriously ill, a follow-up chart review was conducted on these 248 patients. All patients were first tracked through the study site’s electronic medical record. Any patients for whom that search did not confirm survival or death at 90 days post-ED visit were submitted to the Social Security Death Index. Queries to this data set were not conducted for at least 9 months after the ED visit to maximize the likelihood of uncovering undetected deaths in the cohort. Kaplan-Meier curves for these data are shown in Figure 2. Of all patients aged 70 and older admitted to the hospital from the ED, 13.7% experienced mortality at 90 days. In contrast, patients identified first by the computerized algorithm and then by follow-up confirmation of a likely infectious diagnosis suffered 37.5%

Table 1. Characteristics of Patients with an Emergency Department Diagnosis of Infection (N = 69)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR)</td>
<td>81.5 (10)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>33 (47.8)</td>
</tr>
<tr>
<td>Hospital length of stay, days, median (IQR)</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Index hospitalization mortality, n (%)</td>
<td>5 (7.2)</td>
</tr>
<tr>
<td>90-day case fatality, n (%)</td>
<td>10 (14.5)</td>
</tr>
<tr>
<td>Infectious source, n (%)</td>
<td></td>
</tr>
<tr>
<td>Urine</td>
<td>22 (31.9)</td>
</tr>
<tr>
<td>Lung</td>
<td>13 (18.8)</td>
</tr>
<tr>
<td>Skin</td>
<td>7 (10.1)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>6 (8.7)</td>
</tr>
<tr>
<td>Multiple sites</td>
<td>6 (8.7)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (4.3)</td>
</tr>
<tr>
<td>Undetermined</td>
<td>12 (17.4)</td>
</tr>
</tbody>
</table>

IQR = interquartile range.
mortality in the same time frame ($P < .01$ according to proportional hazards modeling).

The median time from ED arrival to notification was 109 minutes (interquartile range (IQR) 59–177), from triage to notification was 93 minutes (IQR = 50–170), and from placement in room to notification was 89 minutes (IQR = 47–164). One reason for delay in notification was the time needed for collection, measurement, and reporting of peripheral white blood cell counts. In addition, patients frequently did not meet two SIRS criteria at triage. Rather, their disease course tended to evolve over the early hours of observation and treatment.

Because the SIRS criterion respiratory rate has variously been noted as 20 or more per minute or more than 20 per minute, a sensitivity analysis of this distinction was conducted, which removed patients who met the SIRS criteria by virtue of a respiratory rate of 20 breaths per minute or less. The specificity of the electronic notification for ED infection diagnosis increased to 86.0% (95% CI = 81.0–91.1%), and the sensitivity decreased to 30.4% (95% CI = 19.6–41.3%).

The performance of the system from July 1 through November 15 was further considered in all patients for whom automated notification occurred. Specifically, in these patients, which patterns of SIRS criteria were mostly likely to be associated with patients going on to meet the second tier of the inclusion criteria related to suspected infection was of interest. For this analysis, 434 patients were included. The associations between each combination of possible criteria and a patient being considered likely to have a serious infection are shown in Figure 3. Although the combination of tachycardia and tachypnea had the weakest association with patients being considered candidates for study, this combination’s frequency (48% of all cases) led it to be the pattern from which most patients of interest were identified.

## DISCUSSION

The current study found that a two-stage patient identification protocol identified all potentially eligible patients for an investigation of serious infection in elderly ED patients. Infectious diagnoses were common in the sample; patients included had both apparently serious (SIRS) and nonserious infections (incidental urinary tract infection). As a result, the sensitivity for identifying all patients with infection was low. A validated, inclusive definition of serious infection in emergency and critical care is elusive; therefore, the low sensitivity observed was not surprising. Nevertheless, for older patients requiring admission, the automated strategy with a single follow-up question performed by a study coordinator produced a population three times as likely to have an ED infectious diagnosis as the general ED population.

High patient volume can affect timely and reliable study candidate identification. Furthermore, geriatric populations are frequently underrepresented in ED research.
Busy frontline medical providers are less likely to take an additional step to contact a study coordinator when an eligible patient arrives. Therefore, an automated system to reliably identify candidates could eliminate this potential barrier to subject recruitment. This can have beneficial effects including but not limited to a reduction of selection bias due to ED volume, less demand on and distraction of clinical staff, simplified departmental monitoring by study coordinators, and shorter overall study duration because of faster enrollment. Each of these would be expected to reduce the cost and administrative burden of conducting ED research.

Prior investigations have described the use of information systems for real-time screening of patients for research and clinical purposes. An automated system, which similarly used the hospital information system to identify syncope patients as part of the development of a clinical decision rule, has been described.\(^5\) The current work expands upon this by using clinical parameters from the information system as opposed to the reported reason for visit. Additionally, the majority of patients in the current study “evolved” into meeting inclusion criteria; for potentially seriously infected patients, a one-time screening would have been insufficient. Others have explored the use of computerized clinical information data as an “early warning system” to identify inpatients who are in distress or are in a pre-arrest state.\(^6\) The surveillance schema developed for the current study has potential utility in this respect as well. Extension of surveillance to identifying clinical parameters that predict a well-defined serious disease process before clinically apparent deterioration would be beneficial. The present work on identifying serious infections in older ED patients demonstrates the feasibility of such a system in the scope of existing hospital information databases. Adaptation of this process to a diagnostically more-concrete disease could potentially greatly improve the sensitivity of such a strategy.

This work has several important limitations. The characteristics of patients without an ED diagnosis of infection were not collected, and thus the cohort studied cannot be differentiated from the full population of older ED patients. An additional limitation is that this investigation was performed at a single academic medical center, with a relatively demographically homogenous catchment area.

Clinical informatics systems differ widely between institutions. It is therefore not possible to comment specifically on the ease with which this strategy might be deployed in other EDs, although in the SQL environment in which the algorithm was developed, the entire query was implemented in just a few lines of code and ran fast enough not to interfere meaningfully with other clinical information system data management tasks.

---

Figure 3. Relationship between all possible combinations of two or more systemic inflammatory response syndrome (SIRS) criteria and the odds of ultimately having a study coordinator confirm a likely infectious diagnosis. All patients who generated an automated notification in this study (July–November) are included. Although the combination of tachycardia and tachypnea had the weakest association with a patient being suspected of having an infectious diagnosis, the high frequency of this combination of criteria resulted in most patients being captured from that pattern. In contrast, patterns including both abnormal temperature (T) and abnormal white blood cell count (WBC) were more strongly associated with a suspected infectious diagnosis. HR = heart rate; RR = respiratory rate.
CONCLUSIONS
The current study prospectively evaluated a simple two-part syndromic surveillance strategy consisting of an automated clinical information system search algorithm followed by a single-question confirmation. This work describes the methodology for automated patient identification. The results indicate that such strategies are feasible as a tool for conducting clinical research and provide valuable proof of concept as a tool that could be useful in a variety of research and clinical applications. Further work is needed in optimizing the sensitivity and specificity of the method and in adapting it for applications in direct clinical care.

ACKNOWLEDGMENTS
This work was funded in part by a grant from the Nathan Shock Center for Aging Research, the Claude D. Pepper Older Americans Independence Center (funded through the National Institute on Aging, NIA P30AG024824), and the John A. Hartford Foundation (JGY, EDL). This work was presented in part at the 2008 American Geriatrics Society and Society for Academic Emergency Medicine annual meetings.

Conflicts of Interest: The authors have no conflicts of interest relevant to this research to report.

Author Contributions: All authors critically reviewed this manuscript for important intellectual content and approved its final form. Study concept and design: W.J.M., E.D.L., P.N.M., J.G.Y. Acquisition of data: B.L.S., D.S., J.D.Y., J.D.J., and J.G.Y. Analysis and interpretation of the data: W.J.M., E.D.L., and J.G.Y. Drafting of the manuscript: W.J.M. and J.G.Y. Provision of statistical expertise: W.J.M. and J.G.Y. Administrative and technical support: B.L.S. and J.D.J. Obtaining of funds and overall study supervision: J.G.Y.

Role of Funding Source: The study sponsor was not involved in the design, methods, subject recruitment, data collections, analysis or preparation of the manuscript. In addition, the sponsor had no role in the decision to publish this manuscript.

REFERENCES