Forecasting emergency department presentations

Robert Champion, Leigh D Kinsman, Geraldine A Lee, Kevin A Masman, Elizabeth A May, Terence M Mills, Michael D Taylor, Paulett R Thomas and Ruth J Williams

Abstract

Objective: To forecast the number of patients who will present each month at the emergency department of a hospital in regional Victoria.

Methods: The data on which the forecasts are based are the number of presentations in the emergency department for each month from 2000 to 2005. The statistical forecasting methods used are exponential smoothing and Box–Jenkins methods as implemented in the software package SPSS version 14.0 (SPSS Inc, Chicago, Ill, USA).

Results: For the particular time series, of the available models, a simple seasonal exponential smoothing model provides optimal forecasting performance. Forecasts for the first five months in 2006 compare well with the observed attendance data.

Conclusions: Time series analysis is shown to provide a useful, readily available tool for predicting emergency department demand. The approach and lessons from this experience may assist other hospitals and emergency departments to conduct their own analysis to aid planning.

Aust Health Rev 2007: 31(1): 83–90

What is known about the topic?
There is an extensive body of literature on the application of time series methods in forecasting, particularly in a business context. The methods are well established and often found in standard statistical computer packages. Surprisingly, there are few published examples of the applications of time series methods to forecasting in hospitals.

What does this study add?
This study develops a model for forecasting the number of presentations each month at the emergency department of a hospital in regional Victoria. The forecasts provided by the model are shown to compare very well with the actual outcomes. The study demonstrates how time series analysis can be used for forecasting, at least in the short term, the demand for emergency services in a hospital.

What are the implications for practice?
Every hospital is unique. However, the forecasting methods and approach in this paper could be applied in other hospitals to improve resource allocation and strategic planning.

THE ABILITY TO PREDICT the demand for attendance at the emergency department (ED) of a hospital is valuable — at a micro level for planning rosters for staff in ED, and at a macro level for financial and strategic planning for the hospital. Also, ED attendances are indicators for many aspects of health care in the community, and hence, such predictions may be useful for regional health care planning.

There are many factors that will influence the attendance rate at an ED. Indeed, in England, Milner et al found "startling" variation in ED presentation patterns in 190 health districts.[1] Farmer and Emami[2] also noted regional variation in demand for health services in England and the

Robert Champion, PhD, Senior Lecturer
Terence M Mills, PhD, Professor
Department of Mathematics and Statistics, La Trobe University, Bendigo, VIC.

Leigh D Kinsman, RN, MHSc, Lecturer
Geraldine A Lee, RGN, BSc, PGDE, Lecturer
School of Nursing and Midwifery, La Trobe University, Bendigo, VIC.

Kevin A Masman, BSc(Hons)(Appl Sci), Research Officer, Collaborative Education and Research Centre
Elizabeth A May, RN, MBA, MPPM, Manager, Performance Reporting Unit
Michael D Taylor, MB BS, MHSMN, FRACGP, FACRRM, AFCHSE, Deputy Director, Emergency Department
Paulett R Thomas, RN, PGDipCC, Registered Nurse, Emergency Department
Bendigo Health, Bendigo, VIC.

Ruth J Williams, PhD, Professor
Department of Mathematics, University of California, San Diego, California, United States.

Correspondence: Robert Champion, Department of Mathematics and Statistics, La Trobe University, PO Box 199, Bendigo, VIC 3552. **r.champion@latrobe.edu.au**

effect that this variation will have on forecasting the requirements of such services. Relevant factors that might contribute to this variation may be the demographic structure of the region in which the hospital is located, the socio-economic profile of the region, the number and location of local health care providers such as general practitioners, other EDs, primary care clinics, and nursing homes.

The purpose of this paper is to describe how statistical forecasting methods were used to make short-term predictions for monthly ED attendances at one hospital in regional Victoria. Although the forecasting models that are presented in this paper give the best results for a particular hospital, they may not be optimal for other hospitals where different circumstances prevail. However, the methods of modelling and data analysis are likely to be applicable in other hospitals.

Terms and concepts

The forecasting methods used below are methods for time series analysis. A time series is a sequence of measurements made over time: in this paper the main time series of interest is the number of presentations to the ED each month over a sequence of months.

In a typical hospital, one has ready access to the series of the number of presentations at ED each month over a long period of time. However, one might not have ready access to other regional geographic or demographic data referred to in the previous section. We wish to explore the practical problem of how well ED demand can be forecast, at least in the short term, using only this time series data. Hence, in this paper, we focus on methods from time series analysis.

We used two methods for forecasting time series, namely exponential smoothing and Box–Jenkins methods. The details of these methods can be found in many books[3] and can be implemented using statistical packages such as SPSS (SPSS Inc, Chicago, Ill, USA). Gardiner[4] provides a useful overview of exponential smoothing methods. In this paper we do not assume that the reader has a detailed knowledge of these methods. Our aim is to show the results of their application to ED time series data.

If $X(1)$, $X(2)$, ... denote observations in a time series, then we denote the observation at time t by $X(t)$. For example, $X(t)$ might denote the number of presentations at the ED in month t. If a forecasting method is used to predict the value of the time series at some time t, then we will denote the predicted value by $X^*(t)$. Hence, the error in our prediction would be $X(t) - X^*(t)$, the difference between the actual value and the predicted value. The ultimate test of any forecasting method is the size of these errors.

Literature review

Many researchers have used time series analysis to assess the impact of some factor on the operation of the ED. For example, some have used time series analysis to investigate the impact of pollution factors on ED attendances. Others have examined the impact of a change in operational procedures on patient flow in the ED. However, there are surprisingly few articles that include time series analysis for forecasting overall demand for ED services. Tandberg and Qualls[5] made the comment that time series methods "have rarely been applied in emergency medicine". Below, we outline the findings in relevant papers.

In a UK study, Milner[6] used time series analysis to forecast the annual demand for emergency services in the Trent region. While the notation we use here is not Milner's, it is designed to assist those who may wish to read his papers. Milner's analysis is based on three time series of annual data for the period 1974–1985. The three time series are:

$E(t)$ = number of ED presentations in year t
$N(t)$ = number of new ED presentations in year t
$P(t)$ = resident population in the region in year t

Then, one can calculate three other series:

$A(t) = E(t) - N(t)$
$R_1(t) = 1000\ N(t)/P(t)$
$R_2(t) = A(t)/N(t)$

Thus, $A(t)$ is the number of people who have attended ED in year t and also attended ED previously; $R_1(t)$ is the ratio of the number of new attendees per 1000 residents; $R_2(t)$ is the ratio of the number of people who are re-attending to the number of new attendees. We infer the precise definition of $R_1(t)$ from Milner et al.[1]

Milner[6] uses Box–Jenkins (or ARIMA [autoregressive integrated moving average]) models and the 1974–1985 data to forecast $R_1^*(t)$, $R_2^*(t)$, $P^*(t)$ for the future years t = 1986, 1987, . . ., 1994. This leads to forecasts for the remaining series as follows.

$N^*(t) = R_1^*(t) \times P^*(t)/1000$

$A^*(t) = N^*(t) \times R_2^*(t)$

$E^*(t) = N^*(t) \times (1 + R_2^*(t))$

By performing this analysis for each of the 12 health districts in the Trent region, he calculates forecasts for the entire region. When he compared the actual total attendances over the entire region for 1986 with his forecasts of total attendances in 1986, he found that the mean percentage forecast error was +0.8% and the mean absolute forecast error was 4.3%. Thus Milner obtained accurate, short-term forecasts.

In a subsequent paper, Milner[7] assessed his long-term forecasts. He found that some of these long-term forecasts were inaccurate, with percentage forecast errors as high as 15%. This is not surprising since, first, he was basing his 10-year forecasts on 12 years of annual data, and second, there can be large variations associated with time series of ED attendances. Thus, Milner's short-term forecasts with Box–Jenkins methods were much more accurate than his long-term forecasts.

In a US study, Tandberg and Qualls[5] endeavoured to make short-term predictions for the number of ED presentations at any hour of the week in a university hospital in New Mexico. Their analysis was based on hourly data collected during 1989 in the hospital ED. The statistical methods they used included simple moving averages, seasonal decomposition methods, and ARIMA models. They found that relatively simple models can be used to describe the number of ED presentations for each hour of the week. They assumed that the hourly patterns that are present during a given week in the ED will not vary much from week to week.

Farmer and Emami[2] compared two statistical forecasting methods in trying to predict the requirements in the acute sector of a hospital. They considered regression methods and Box–Jenkins methods from time series analysis. In their analysis, they argued that the regression methods are not effective forecasting tools in an environment where trends can change suddenly, as might happen in the acute sector of a hospital.

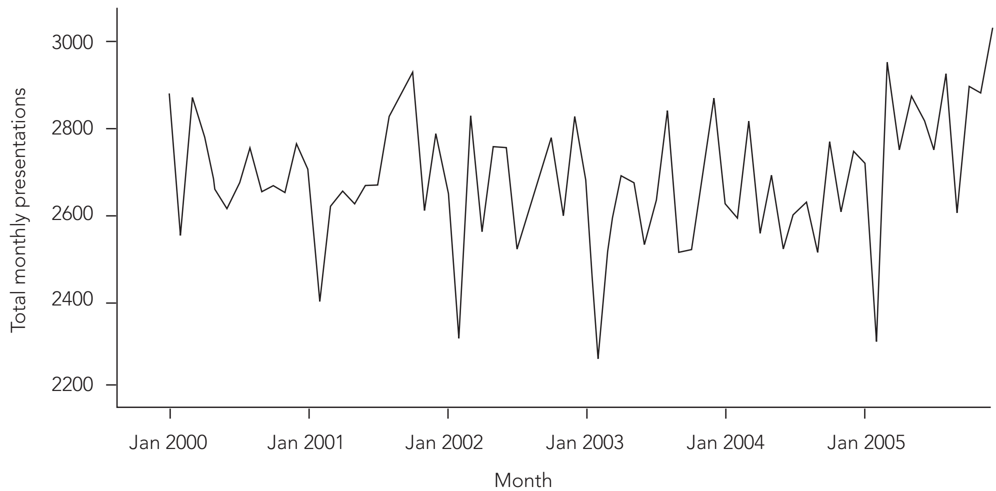
Rotstein et al[8] developed a model for ED presentations at a hospital in Israel based on 3 years of daily time series data. They used a regression model with a linear trend, and 3 types of seasonal factors that represent the effects of (i) the day of the week, (ii) the month of the year, and (iii) the type of day (holiday, half working day, or full working day). They found that their model provided useful forecasts for the next 12 months, although the trend that they detected in the first 3 years of data was not sustained in the year for which they made the forecasts. This lack of flexibility is a common feature of regression models used in forecasting, as Farmer and Emami[2] also indicated.

There are some studies on modelling times series that are related to the series of ED presentations. Sterk and Shryock[9] demonstrated the usefulness of seasonal decomposition methods for forecasting the number of adult inpatient days for each month.

In a UK study, Jones and Joy[10] developed Box–Jenkins models for forecasting the number of daily emergency admissions, rather than presentations, and the number of beds occupied by emergency admissions on a daily basis at Bromley Hospitals NHS Trust. The motivation for this research is that if one can predict emergency admissions, then this ought to assist in planning elective admissions. The authors found that their models gave better fit to the occupancy than to the admissions; they attributed this to the high standard deviation of admissions as a proportion of the mean.

In a Spanish study, Martin-Rodriguez and Caceres Hernandez[11] studied the patterns of ED presentations at a hospital in Tenerife. Their analysis is based on a time series of the number of

1 Total monthly ED presentations, Jan 2000–Dec 2005



ED presentations each hour over a 6-year period from 1 Jan 1997 to 31 Dec 2002. In this careful study, the authors use seasonal decomposition methods to obtain detailed estimates of the effects of hour of day, day of week, and week of the year on attendances. Their method is notable by the use of splines to smooth the series of seasonal factors. They note that there is often a mismatch between the ED staffing rosters and the patterns of ED attendances in Spanish hospitals. However, they do not report on the effectiveness of this model in forecasting demand.

This review of the literature on modelling ED attendance is fairly comprehensive. It is surprising that few authors have modelled monthly ED presentation data.

The literature shows that time series analysis can be used for modelling certain series that arise in emergency medicine. Our experience in modelling ED presentation data is similar to that of Farmer and Emami.[2] We found that time series methods tend to respond better than regression methods to fluctuations in the underlying trend. Often, simple time series models will give useful predictions.

Whether the time series is studied on an annual basis, monthly basis, or hourly basis will depend on the purpose of the study. Annual data would be appropriate for strategic planning; hourly data would be useful for roster planning. We have chosen to study monthly attendance because this is the natural reporting period for hospitals and governments, and is routinely examined by the ED manager to monitor trends.

In spite of these general findings, the optimum forecasting model is likely to vary from hospital to hospital.

Methods

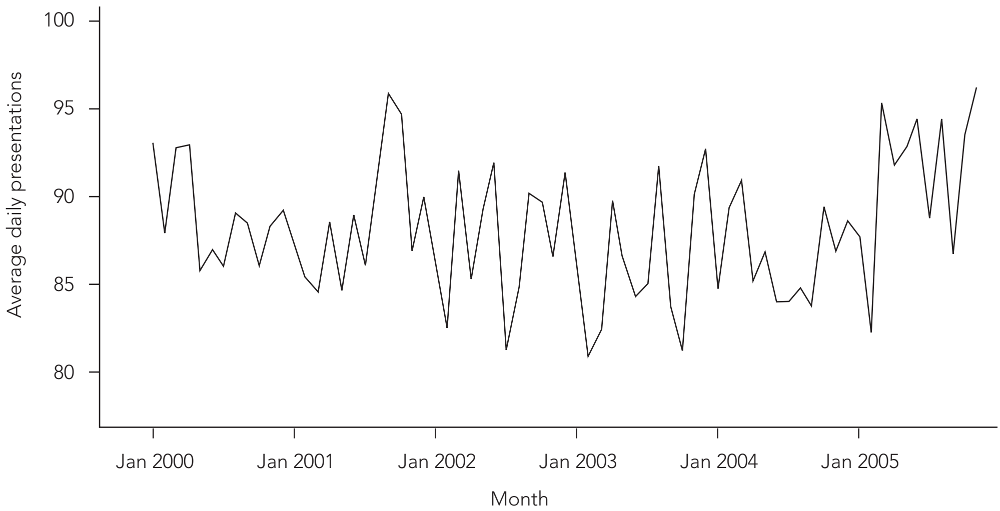
In this section we describe the data on which we based our forecasts and the forecasting models that we used.

Data

The data used in this paper are the time series of the number of patients presenting at the ED in the hospital each month from 2000 to 2005: the series is plotted in Box 1. The data were extracted directly from the ED data system at the hospital.

Presentation numbers were relatively stable between 2000 and 2004, but an increase during 2005 is evident. The mean attendance over the 6 years is 2686 presentations per month. Perhaps the most striking feature of the series in Box 1 is that in each year, except 2004, presentation num-

2 Average daily ED presentations each month, Jan 2000–Dec 2005



bers in February are noticeably low. This is largely due to the number of days in the month. Since February has 2.6 fewer days than the average for other months, we anticipate about 230 fewer presentations in February, all other factors being equal.

It is convenient to remove the length-of-month effect and convert to the variable X(t) = average daily presentations in the month (total presentations ÷ number of days in the month). Henceforth, X(t) will denote this variable. The average daily presentation series, or X(t), is depicted in Box 2. The February "spikes" (and other less obvious length-of-month effects) are removed.

The daily presentation rate declined from a yearly average of 88.8 in 2000 to 86.5 in 2004. In 2005, the average rose to 92 presentations per day. The series is quite "noisy", with no obvious pattern. Month-to-month fluctuations about the mean are of the order of five presentations per day.

Box 3 shows measures of monthly seasonal variation of the number of presentations/day. Thus in February there are on average 3.6 presentations per day fewer than the yearly average. Presentations in December were 2.8 presentations per day above average. However, these seasonal effects are barely significant since random fluctuations are similar in magnitude.

Models

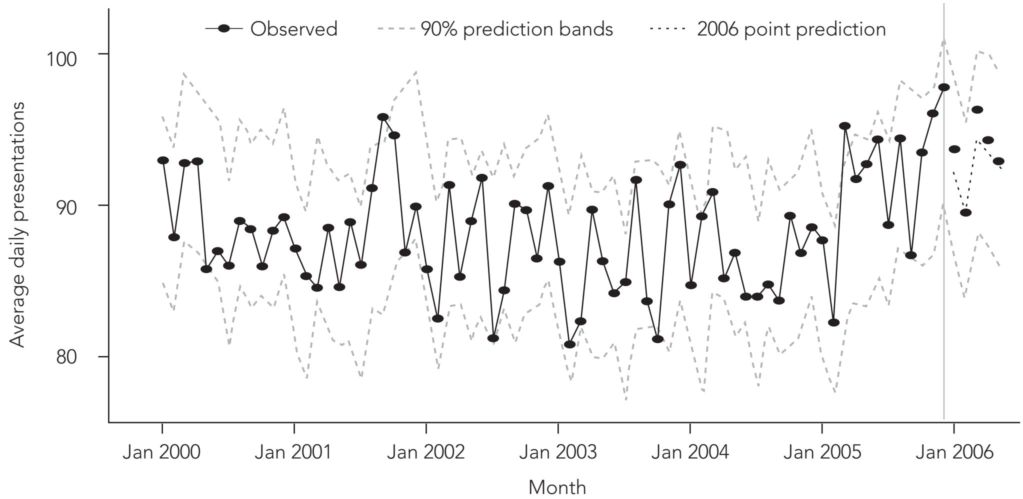
Two classes of models that are widely used for forecasting time series are exponential smoothing models and classical Box–Jenkins, or ARIMA, models. Exponential smoothing methods are based on weighted moving average formulae. ARIMA models are more sophisticated stochastic models that combine elements of moving average methods and regression methods.

Characteristics of the data, such as the existence and nature of any trend and seasonal varia-

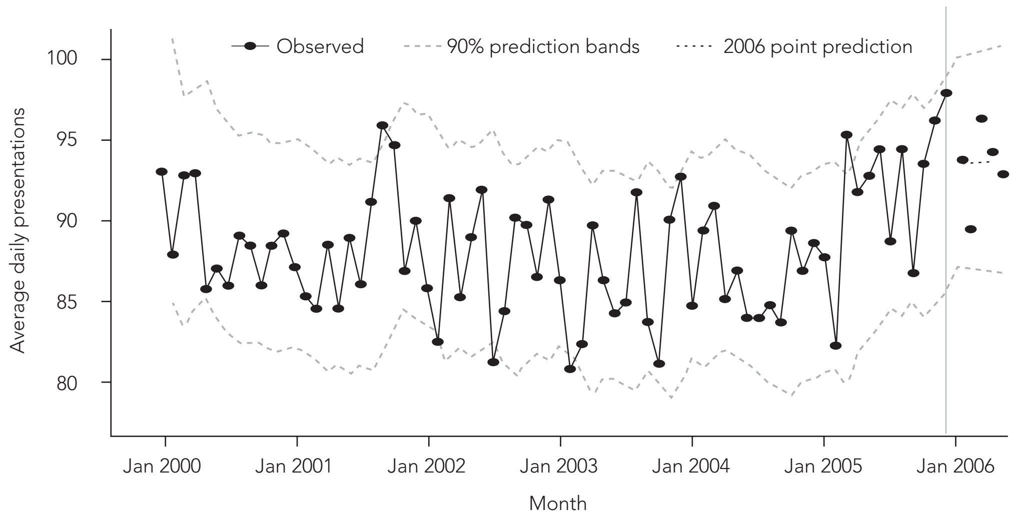
3 Monthly additive seasonal index values

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
−1.3	−3.6	1.2	0.3	0.0	0.6	−3.0	0.8	1.0	0.8	0.3	2.8

4 Simple seasonal exponential smoothing: fitted series (Jan 2000–Dec 2005) and forecasts (Jan 2006–May 2006)



5 ARIMA(0,1,1): fitted series (Jan 2000–Dec 2005) and forecasts (Jan 2006–May 2006)



tion, determine which models are appropriate for a particular series, and for ARIMA models autocorrelation properties play a key role in model identification. Software implementations automate the model selection process. We used SPSS version 14.0, in which the new Trends package provides for automated identification of optimal models from each class based on various performance measures.

Results

We compared the performance of the optimal exponential smoothing model and the optimal

ARIMA model for fitting the daily average ED presentation series for 2000–2005 and, in particular, their success in forecasting presentations in the first 5 months of 2006.

The best exponential smoothing model is a simple seasonal model. The fitted series and forecasts with 90% prediction bands are illustrated in Box 4. The observed values of X in 2006 that we are attempting to forecast are included in the chart for comparison with predictions. The root mean square error (RMSE) is 3.3 presentations per day. The optimal ARIMA model corresponds to a non-seasonal moving average model of order one applied to the first differences of the series, that is, an ARIMA(0,1,1) model. The RMSE for this model is 3.9. The smaller RMSE for the exponential smoothing model is a rudimentary indication that the exponential smoothing model is a little better. The stationary R-squared values, 0.71 and 0.31 for the simple seasonal and ARIMA models respectively, and normalised Bayesian information criterion (BIC) values, 2.5 and 2.8 respectively, confirm this. The fitted series and forecasts are found in Box 5. The point predictions for each model are compared with the observed 2006 values in Box 6.

In general terms, the forecasts from the two models are quite similar. The prediction bands have similar widths and the point predictions are reasonably close. Comparison of point predictions for 2006 with observed values in the charts and in Box 6 reveals that the models are quite successful in predicting presentation numbers. An obvious difference, looking at the 90% prediction bands over the 2000–2005 period, is the regular seasonal fluctuation, particularly the dips each February and July and the peaks each December, predicted by the exponential smoothing model. As a result, the exponential smoothing model predicts the drop in presentations in February 2006, whereas the ARIMA model does not. Inclusion of seasonal terms in the ARIMA model does not improve its performance.

The prediction bands give us a clear visual picture of how to interpret forecasts. Assuming that the factors that drive ED presentations remain much the same in 2006 as in the past we anticipate that the 2006 observations will lie within the prediction bands most of the time (90% of the time in the long run). Of course, if there were changes in circumstances driving presentation numbers after 2005, not reflected in the past data, then significant departure from forecasts may occur. If we consider the broadness of the prediction bands we would concede that the forecasting performance for the Jan to May 2006 period is better than we could hope to expect on average. The error standard deviations in the prediction period for the exponential smoothing and ARIMA models are 1.3 and 2.5 respectively, which are quite a bit less than residual standard errors over the 2000–2005 period. We would normally be content with prediction error standard deviations being a little greater than residual standard errors.

6 Comparison of observed average daily presentations in 2006 with point predictions obtained by fitting 2000–2005 data

		Exponential smoothing		ARIMA(0,1,1)	
Month	***X***	**Forecast**	**Error**	**Forecast**	**Error**
Jan 06	93.7	92.2	1.5	93.5	0.2
Feb 06	89.5	89.5	0.0	93.6	–4.1
Mar 06	96.3	94.3	2.0	93.7	2.6
Apr 06	94.3	93.7	0.6	93.7	0.6
May 06	92.9	92.3	0.6	93.8	–0.9

ARIMA = auto-regressive integrated moving average.

7 Standard deviations in forecast errors for Jan 2006 to May 2006

Fit data up to	Exponential smoothing	ARIMA
Sep 2005	4.0	4.9
Oct 2005	3.5	4.0
Nov 2005	1.9	2.9
Dec 2005	1.1	2.5

ARIMA = auto-regressive integrated moving average

An important forecasting strategy[6] is to update forecasts as more data become available. We would expect that better forecasts will be obtained if more recent data are available to model. To demonstrate this, we predicted the January to May 2006 values of the series by fitting the series up to different endpoints in 2005. The gradual decline in the standard deviations in forecast errors in Box 7 illustrates a gradual improvement in forecasting performance as more recent data are used in the forecasting models.

Discussion

Although emergencies are unpredictable events, the attendance patterns at the ED can be predicted. We have found that there is little variation between months in the number of ED presentations per day. There is some variation, but it is barely significant. We have demonstrated that average daily ED attendance each month can be predicted using well known techniques and simple models from time series analysis.

The average daily presentations time series is a relatively simple series, and we found that relatively simple models give good results. However, more complex series, such as those with changing trend patterns and/or daily or hourly data that display regular cycles, can be readily analysed by the methods we have used. Both the exponential smoothing and Box–Jenkins classes of models are quite adaptable to a wide range of data characteristics.

Forecasting methods do assist in evidence-based management. Thus, other hospitals may be able to use these methods for their own internal planning and business analysis.

This study raises many questions about related series. Do time series for different triage categories or International classification of diseases (ICD) groups share similar characteristics, or do they differ? Can we predict emergency admissions as distinct from emergency presentations? We could also examine ED presentations on a finer time scale, such as daily or hourly. Accurate prediction of these series would facilitate planning of nursing rosters and allocation of staff within the department, and can potentially assist in bed occupancy prediction.

Acknowledgements

The authors acknowledge financial support from the Faculty of Health Sciences, La Trobe University and Bendigo Health.

Competing interests

Although Bendigo Health is the organisation on which this case study was based, Bendigo Health is not referred to in the article. However we wish to declare that the authors Masman, May, Taylor and Thomas are employed at Bendigo Health and Mills has an honorary position at Bendigo Health.

References

1 Milner PC, Nicholl JP, Williams BT. Variation in demand for accident and emergency departments in England from 1974 to 1985. *J Epidemiol Community Health* 1988; 42: 274–8.
2 Farmer RDT, Emami J. Models for forecasting bed requirements in the acute sector. *J Epidemiol Community Health* 1990; 44: 307–2.
3 Chatfield C. The analysis of time series: an introduction. 6th ed. Boca Raton: Chapman & Hall/CRC; 2004.
4 Gardiner ES. Exponential smoothing: the state of the art. *J Forecast* 1985; 4: 1–28.
5 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.
6 Milner PC. Forecasting the demand on accident and emergency departments in health districts in the Trent region. *Stat Med* 1988; 7: 1061–72.
7 Milner PC. Ten-year follow-up of ARIMA forecasts for attendances at accident and emergency department in the Trent region. *Stat Med* 1997; 16: 2117–25.
8 Rotstein Z, Wilf-Miron R, Lavi B, et al. The dynamics of patient visits to a public hospital ED: a statistical model. *Am J Emerg Med* 1997; 15(6): 596–9.
9 Sterk WE, Shryock EG. Modern methods improve hospital forecasting. *Healthc Financ Manage* 1987; 41: 96–8.
10 Jones SA, Joy MP. Forecasting demand of emergency care. *Health Care Manag Sci* 2002, 5(4): 297–305.
11 Martin Rodriguez G, Caceres Hernandez JJ. [A method for ascertaining the seasonal pattern of hospital emergency department visits] [Spanish]. *Rev Esp Salud Publica* 2005; 79(1): 5–15.