Predicting emergency department admissions

Justin Boyle,1 Melanie Jessup,2,3 Julia Crilly,3 David Green,3 James Lind,3 Marianne Wallis,2,3 Peter Miller,4 Gerard Fitzgerald5

ABSTRACT
Objective To develop and validate models to predict emergency department (ED) presentations and hospital admissions for time and day of the year.

Methods Initial model development and validation was based on 5 years of historical data from two dissimilar hospitals, followed by subsequent validation on 27 hospitals representing 95% of the ED presentations across the state. Forecast accuracy was assessed using the mean average percentage error (MAPE) between forecasts and observed data. The study also determined a daily sample size threshold for forecasting subgroups within the data.

Results Presentations to the ED and subsequent admissions to hospital beds are not random and can be predicted. Forecast accuracy worsened as the forecast time intervals became smaller: when forecasting monthly admissions, the best MAPE was approximately 2%, for daily admissions, 11%, for 4-hourly admissions, 38%; and for hourly admissions, 50%. Presentations were more easily forecast than admissions (daily MAPE ~7%). When validating accuracy at additional hospitals, forecasts for urban facilities were generally more accurate than regional forecasts [accuracy is related to sample size]. Subgroups within the data with more than 10 admissions or presentations per day had forecast errors statistically similar to the entire dataset. The study also included a software implementation of the models, resulting in a data dashboard for bed managers.

Conclusions Valid ED prediction tools can be generated from access to de-identified historic data, which may be used to assist elective surgery scheduling and bed management. The paper provides forecasting performance levels to guide similar studies.

INTRODUCTION
Emergency department (ED) overcrowding is acknowledged as an increasing, worldwide issue.1–3 The implications of impaired function of the ED are increased numbers of ambulance bypass occurrences4 and less favourable outcomes for patients,5–7 including increased mortality associated with access blocked patients.7 Access block has also resulted in last minute cancellation of elective surgical patients with resultant inflating elective waiting lists on which patients spend increasing time.8 Now endemic and critical, access block has been nominated as a threat to patient safety.2 5 Recognition of the adverse outcomes of access block necessitates a shift from the current perception to one in which it is considered as the unsafe situation that it is.5 6

A key advantage in catering for the large numbers of people presenting to an ED is the ability to predict their numbers; consequently there have been a number of previous studies relating to forecasting bed demand.9–21 Some of the models express accuracy only in terms of the degree of fit to historic data rather than prospective forecast performance,20 21 some relate to forecasting patient presentations to EDs only and not subsequent hospital admissions.13 15 17–19 and essentially all are validated at relatively few sites (typically ≤4).

The main objective of this study was to develop and validate a clinically useable software package that accurately predicts the number of ED presentations and subsequent admissions (those patients that require a bed and thus represent a demand on bed management), on any given day of the year, taking into account peak periods such as public holidays. An additional objective was to determine a daily sample size threshold for forecasting subgroups within the data (eg, gender and criticality), as this can be important, for example, in assigning patients to gender specific or medical specialty wards.

METHODS
Study design and setting
Initial development and validation of the model was based on 5 years retrospective analysis (1 July 2002–30 June 2007) of consecutive ED presentations and hospital admissions made to two hospitals in Queensland, Australia. The applicability of the model to other hospitals was assessed by determining accuracy of forecasts of ED presentations at 27 Queensland public hospitals using a later data extract spanning 2005–2009. These validation sites use the same ED and inpatient clinical information systems as the pilot sites, and represent approximately 95% of the ED presentations across the state. The study was approved by human research ethics committees associated with the facilities.

Measurements
In this study, accuracy was treated as the main criterion for selecting a forecasting method, and our assessment of forecast accuracy is based on the principles contained in Makridakis et al.22 The metric used in our evaluations was the mean absolute percentage error (MAPE); further detail of this metric is presented as an appendix.

True out-of-sample forecast accuracy was measured in this study, where data was divided into a training set and evaluated against a separate hold-out set. We considered a 1-year hold-out evaluation period important in order to assess forecasts over summer and winter months. Figure 1 depicts the training and evaluation periods used in the study.

For the initial development and validation of the model, the effect of varying the size of the training
dataset (from 1 to 4 years) was analysed. Also computed were the width of 95% prediction intervals and the number of misses outside this prediction interval. This provides the user of the forecasts with worst and best case estimates and a sense of how dependable the forecast is. As an outcome from the study, it was desired to compare forecasting performance against an existing prediction model at one of the pilot hospitals, and also against other published forecast performance.

Forecasting models

With experience, hospital bed managers can identify which days of the week are usually busier than others, and that holiday periods reflect large admission numbers. Thus it is likely that ED forecasting models need to allow for this seasonality, by including variables for the day of week, month of year and holidays, and identifying repeated patterns in the time series data.

We have found that days having matching characteristics are considered to be most ‘similar’ to the day of interest, in terms of patient demand, and accordingly are expected to provide the best basis for prediction. We extracted all days within the historical data that match the day type (Sunday, Monday, public holiday etc) within a 4-week window, centred on the day of interest—that is, 2 weeks prior and 2 weeks following, for use as the basis for prediction (see figure 2). For recent historical data during the current year, information will only be available from the preceding weeks, since the following weeks have yet to occur. We then applied a computational predictive model (eg, smoothing) to the extracted data in order to produce a prediction of patient demand for the specified time period.

Other models considered in the model development included multiple regression, autoregressive integrated moving average...
(ARIMA/Box–Jenkins), and exponential smoothing; further
generic detail of these methods is included as an appendix.

RESULTS
Characteristics of the test sites
Characteristics of the test sites in terms of ED presentations,
admissions and admission rates are shown in figure 3. Analysis
of the data identified the days of the week that represent higher
ED workloads and hospital bed demands. The strong day-of-
week effect is evident in these plots, supporting the inclusion of
this variable in many forecasting models.

The accuracy of the model was subsequently assessed against
27 public hospitals, with mean daily presentation rates ranging
from 60 to 190 patients/day (figure 4A). It can also be seen that
in general, hospitals in regional areas had lower ED presentation
rates than urban facilities.

Forecast accuracy
MAPE for predicting monthly admissions was approximately
2% at both pilot hospitals, while the error for daily admissions
was 16%/11% (facility 104/facility 50): for 4-hourly admissions,
47%/40%; and for hourly admissions, 49%/51%. MAPE
figures were higher in smaller time intervals (4-hourly and
hourly) as the effect of a missed prediction is more pronounced
as a percentage of the smaller number of actual admissions.
When forecasting hourly or 4-hourly admissions, the lowest
MAPE and the lowest number of forecasts outside the 95%
prediction interval occurred during the busiest period (with
largest sample size).

Forecast accuracy of ED presentation data was also modelled
and was found to be better than the forecast accuracy for
admissions (MAPE ~7% vs ~11%, respectively), likely due to the
larger sample sizes. Models needed to include population
growth, otherwise the estimated number of ED presentations
was underestimated. This was not the case for admissions.

The study assessed the effects of varying the length of
training data, and it was found that smoothing models worked
best with as much data as possible (full training dataset), while
regression worked best based on the most recent data (1-year
training data). There were no significant differences in absolute
percentage errors obtained from the best smoothing technique
compared to the best regression (t(0.74, 0.052) = 0.39, p = 0.69).
The ARIMA, regression and exponential smoothing methods did
not offer statistically significant improved forecast performance
over the 4-week rolling window method described above, and so
the multi-site validation and our consequent implemented
model is based on this method (see figure 2), using as much
training data as possible.

Comparison to existing system
The new forecasting models were compared to an existing
prediction system available to bed managers at one of the
hospitals. The existing prediction model was a simple average of
the preceding 2 years, based on calendar position. The MAPE for
daily admissions across an evaluation period (12 July 2006 to 20
May 2007) was 20.5% and 10.6% for the old and new models,
respectively, which represents a reduction in error of 9.9% across
the data tested, or the equivalent of ~5 beds, based on the

Figure 3 Day-of-week and monthly trends at the two pilot development hospitals for patient presentations (top) and admissions (bottom). Left: facility 104 (regional); right: facility 50 (urban).
hospital’s mean admission rate of 50 admissions per day. Multiple comparison testing has been performed on this data and shows that the differences in forecast performance are significant ($F(4,880,0.05(2))=8.4, p<0.001$).

**Sample size thresholds**

We also compared the forecasting error and width of 95% CIs for different forecasting techniques for when admissions data was partitioned into subgroups of gender and criticality. This analysis determined that forecasting performance is roughly equivalent if sample sizes are greater than 10 admissions or presentations per day. This equivalence was tested formally with multiple comparison testing of the forecast accuracy obtained from these groups ($F(7,2486,0.05(2))=82.3, p<0.001$). Forecasting errors for subgroups with sample sizes less than 10 patients/day were significantly different from the error for the entire dataset. At both pilot hospital sites, 10 cases/day was the cut-off for forecasting without significant difference in error to the entire dataset—that is, to forecast a particular category of interest, there needed to be roughly more than 10 admissions or presentations per day.

**Wider applicability**

Forecast accuracy for ED presentations at 27 public hospitals was quantified to determine wider applicability of the model (see figure 4). The 27 hospitals are geographically dispersed across the state, servicing both urban and regional areas. Figure 4B shows that forecasts for urban facilities are generally more accurate than regional forecasts (as accuracy is related to sample size). Six of the hospitals only commenced routine data collection from 2008, and it can be seen that accuracy for these facilities (indicated by the asterisks in figure 4B) is generally...
poorer than for hospitals with longer data history. Figure 4C indicates that the highest accuracy is observed over weekends and that public holidays have greatest variance. Figure 4D shows that poorer forecast performance is experienced over winter months, particularly the winter of 2009 (and to a lesser extent 2007), which correlates with significantly increased influenza-like ED presentations experienced across this season (data not presented in this study).

Figure 5 shows a screenshot of the software implementation of the presentations and admissions predictive modelling. Forecasts are provided for daily admissions and presentations, patient-flow in 4-hourly blocks, hourly admissions, gender, medical specialty and criticality (Australasian Triage Scale). The web-based application has been installed within the health department, allowing password protected access to forecasts for facilities associated with a user.

DISCUSSION

This study assessed the accuracy of a simple predictive model against de-identified historic data for a range of urban and regional hospitals, and potential exists for the model to be implemented in other facilities. The research also identified that sample sizes of around 10 patients per day are required for accurate forecasting.

Table 1 compares the study to related work. Hoot et al. assessed the ability of logistic regression and neural network models to predict ED overcrowding (ambulance diversion). In addition to the authors’ noted limitations of the work being performed at one institution and the cost of false alarms, our work differed in that we wanted to forecast patient admission numbers for the purposes of proactive bed management, and quantify how close the forecasts were with observed data via an accepted forecast error metric (MAPE). The authors extended their modelling with a good study to forecast ED operating conditions over 400 days, using a sliding window of the preceding 4 weeks of historical ED data. Forecasts were made at 10-minute intervals, which ensures the forecasts reflect current ED status, but consequently requires a relatively high frequency of data integration. Importantly, and this is also a limitation of our study, the authors note there was no intervention based on the tool, and in fact ED personnel did not have access to the forecasts during the study which may have influenced the outcome.
Related work involving the above authors has been performed in predicting hourly ED presentations.13 McCarthy et al13 examined the influence of several temporal, climatic and patient factors on hourly ED arrival rates and found the most important predictor to be hour of the day. The authors report an increase in ED arrivals the day after an official holiday compared to a normal day. Our model allows for this distinction, along with other factors on hourly ED arrival rates and found the most important predictor to be hour of the day. The authors report an increase in ED arrivals the day after an official holiday compared to a normal day. Our model allows for this distinction, along with other factors on hourly ED arrival rates.

Schweigler et al14 applied several forecasting models to historic ED bed occupancy data. Time domain models were limited to 24-hour periodicity, and in comparison we have looked at effects beyond 1 and 24-hour periods earlier (our models included month of year and holiday effects). We also assessed ED workload (presentations) and subgroups within that data. Similar to our study, forecast accuracy in the Schweigler study was assessed against a separate held-out evaluation set not used in the study period. Forecast accuracy in the Schweigler study was assessed against a separate held-out evaluation set not used in the study period.

Champion et al15 used the SPSS Trends package to automatically identify optimal models. Jones et al17 daily presentation data were used to forecast ED overcrowding (ambulance diversion). Wargon et al19 daily presentation data were used to forecast ED overcrowding (ambulance diversion). Reis and Mandl19 daily presentation data were used to forecast ED overcrowding (ambulance diversion).

### Table 1: Comparison of forecast performance with related work

<table>
<thead>
<tr>
<th>Author group</th>
<th>Target outcome for prediction</th>
<th>Techniques</th>
<th>Training period</th>
<th>Evaluation period</th>
<th>Findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Study</td>
<td>ED presentations and subsequent admissions</td>
<td>ARIMA, regression, exponential smoothing</td>
<td>1, 2, 3, 4 years</td>
<td>12 months (summer and winter)</td>
<td>MAPE: Daily admissions: 2% Hourly admissions: 38% Daily presentations: 7% Good discriminatory power in predicting ambulance diversion</td>
<td>Simple methods assessed, no outcome evaluation/intervention based on the tool</td>
</tr>
<tr>
<td>Hoot et al10</td>
<td>ED overcrowding (ambulance diversion)</td>
<td>Logistic regression and neural networks; 5 metrics of ED crowding</td>
<td>4 months</td>
<td>2 months</td>
<td>Best performance from seasonal ARIMA and sinusoidal model</td>
<td>Work performed at 1 institution, cost of false alarms, forecast error not quantified via MAPE</td>
</tr>
<tr>
<td>Hoot et al11</td>
<td>ED operating conditions in next 8 h</td>
<td>Sliding window of the preceding 4 weeks of historical ED data</td>
<td>1 month</td>
<td>400 days</td>
<td>Forecast error not measured (only percentage of time observed data was within prediction intervals)</td>
<td></td>
</tr>
<tr>
<td>McCarthy et al13</td>
<td>Hourly ED presentations</td>
<td>Poisson regression: temporal, climatic and patient factors</td>
<td>A random half of a one-year study period</td>
<td>Remaining half of the one-year study period</td>
<td>Accuracy (measured by the Pearson’s r coefficient) decreased as forecast horizon increased from 2 to 8 hours into the future. Hour of the day is the most important predictor in ED arrival rates.</td>
<td>No assessment of triage categories, no intervention based on the tool. Magnitude of forecast error not measured</td>
</tr>
<tr>
<td>Schweigler et al14</td>
<td>Hourly ED bed occupancy</td>
<td>Seasonal ARIMA, sinusoidal, hourly historical averages</td>
<td>1–2 weeks prior</td>
<td>Monday evenings only across 1 year</td>
<td>MAPE across 30 days ranged from 9% to 14%</td>
<td>Time domain models limited to 24-hour periodicity (not month of year and holiday effects)</td>
</tr>
<tr>
<td>Champion et al15</td>
<td>Monthly ED presentations</td>
<td>SPSS Trends package to automatically identify optimal models</td>
<td>5 years</td>
<td>5 months</td>
<td>MAPE: Optimal forecasting performance from seasonal exponential smoothing model</td>
<td>No intervention based on the tool. Magnitude of forecast error not measured (only percentage of time observed data was within prediction intervals)</td>
</tr>
<tr>
<td>Jones et al17</td>
<td>Daily presentation data</td>
<td>Seasonal ARIMA, regression, exponential smoothing, and artificial neural network models</td>
<td>2 years</td>
<td>1 month evaluation across 3 month period</td>
<td>MAPE: Optimal forecasting performance from seasonal exponential smoothing model</td>
<td>No uncertainty statements in the form of prediction intervals</td>
</tr>
<tr>
<td>Wargon et al19</td>
<td>Daily presentation data</td>
<td>Regression model created with SPSS</td>
<td>3 years</td>
<td>1 year</td>
<td>MAPE: Ranged from 8.1% to 17% for 4 individual hospitals or 5.3% when combined</td>
<td>Only presentation data, no prediction intervals</td>
</tr>
<tr>
<td>Reis and Mandl19</td>
<td>Daily presentation data</td>
<td>ARIMA models using SAS package</td>
<td>10 years</td>
<td>2 years</td>
<td>MAPE: 9.4%</td>
<td>Longer training data and higher forecast error than current study</td>
</tr>
</tbody>
</table>

analysis includes daily, 4-hourly and hourly time intervals. Predictions at a daily resolution are expected to be more useful for bed managers than monthly estimates. Our work also included an assessment of the impact of a forecasting model on ED staff and ED work practices.24

In a study based on daily presentation data, Jones et al17 compared seasonal ARIMA, regression, exponential smoothing and artificial neural network models to forecast daily presentations at three facilities. The best MAPE figures across 30 days for each facility ranged between 9.0% and 14.0%. We have improved on this forecast performance: the evaluation period in our study was 12 months and included a range of dates (summer and winter) rather than a 1-month evaluation at the same time of year. Forecast performance for presentations using models developed for this project was 8.0% (facility 1) and 7.1% (facility 2). Additionally, it is usually desirable to provide not only forecast values but also accompanying uncertainty statements in the form of prediction intervals, which have been included in the project models, unlike Jones et al.

Wargon et al18 used the SPSS package to determine a regression model, and evaluated its forecast performance across one year using 3 years of training data from four different hospitals. They report MAPEs ranging from 8.1% to 17% for each hospital, and 5.3% for the four EDs combined. This overall combined performance is better than we achieved in our study, although performance at an individual hospital level was not as high. The authors also report that the calendar variables influencing forecasts were different in each ED, suggesting the need for local calibration, as opposed to our model which requires estimates of local population growth only.

Reis and Mandl19 used the SAS package to fit ARIMA models to nearly a decade of ED presentation data, and report a MAPE of 9.4% when validated against the final two years of the dataset. Our study achieved lower forecast error for daily presentations (MAPE ~7.1%) using only half the length of training data.

The work described in this paper differs from many of these studies in that we desired to predict admissions as distinct from all presentations, as this represents demands made on hospital beds. Moreover, when compared to other regression attempts which only report the degree of fit (R²) of the forecasting model to the data,20 21 the validation of the models in this project was based on a ‘hold-out’ set of data not included in developing the models.

LIMITATIONS

The forecast models assessed in the study, including the implemented model, are fairly simplistic. Other techniques such as dynamic regression, state space models and neural network forecasts were not pursued due to the limited timeframe of this study, and the desire to use models that could be easily replicated into other EDs. Also this study did not include outcome evaluation—that is, taking baseline measurements at the commencement of the project and comparing with post-model implementation measurements. Such evaluation is essential to quantify the potential benefits of the model, such as reduced ambulance bypass occurrences and elective surgery cancellations.

A precursor to measuring change is obtaining user buy-in and achieving regular use. While we have made the forecasts available to potential users via a web-based application, a challenge remains in achieving behavioural change in response to this information. One idea is to identify appropriate trigger points to escalate bed management response, and the authors have recently commenced mapping the procedures and process required to achieve such change management.

CONCLUSION

This study quantified the accuracy of a simple demand predictive model, and established useful performance benchmarks to guide others developing models in other facilities. The research also identified a minimum sample size required for accurate forecasting (10 admissions or ED presentations per day).

We recommend that all EDs implement a method to forecast admissions which can easily be generated from access to historic data. Such insight into the number of people accessing overburdened hospitals via overcrowded EDs will enable strategic planning and forethought that can benefit patients and health carers alike.

Acknowledgements We would like to acknowledge the support of the Queensland Emergency Medicine Research Foundation, Queensland Health data custodians and staff from the Centre for Healthcare Improvement, and thank Mr Remy Le Padellec and Mr Derek Ireland for their contributions to this work.

Competing interests None.

Ethics approval Initial model development and validation: The Gold Coast Health Service District Human Research Ethics Committee & Toowoomba Health Service District Human Research Ethics Committee. Subsequent multi-site validation: Queensland Health, Central Office Committee Human Research Ethics Committee.

Contributors JB: acquisition of the data, data analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical expertise, obtained funding, administrative, technical, or material support. JL: study concept and design, analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, study supervision. DG: study concept and design, analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, study supervision. JC: study concept and design, analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, study supervision. GF: critical revision of the manuscript for important intellectual content, study supervision. MW: analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, administrative, technical, or material support, study supervision. MJ: analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, administrative, technical, or material support, study supervision. PM: analysis and interpretation of the data, critical revision of the manuscript for important intellectual content, administrative, technical, or material support, study supervision. DF: critical revision of the manuscript for important intellectual content, study supervision.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


Predicting emergency department admissions

Justin Boyle, Melanie Jessup, Julia Crilly, et al.

*Emerg Med J* published online June 24, 2011
doi: 10.1136/emj.2010.103531

Updated information and services can be found at:
http://emj.bmj.com/content/early/2011/06/24/emj.2010.103531.full.html

These include:

**References**
This article cites 22 articles, 4 of which can be accessed free at:
http://emj.bmj.com/content/early/2011/06/24/emj.2010.103531.full.html#ref-list-1

**P<P Email alerting service**
Published online June 24, 2011 in advance of the print journal.
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

Advance online articles have been peer reviewed, accepted for publication, edited and typeset, but have not not yet appeared in the paper journal. Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/