34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Time Based Clustering for Analyzing Acute Hospital Patient Flow

Sankalp Khanna, Justin Boyle, Norm Good, James Lind, Kathryn Zeitz

Abstract— **This paper describes a novel approach employing time based clustering of health data for visualization and analysis of patient flow. Clustering inpatient and emergency department patient episodes into hourly slots based on recorded timestamps, and then grouping them on required parameters, the technique provides a powerful tool for visualizing and analyzing interactions and interdependencies between hospital patient flow parameters. To demonstrate the efficacy of the approach, we employ time based clustering to address some typical patient flow related queries and discuss the findings.**

I. Introduction

Reflecting the movement of patients through sequences of processes as part of their pathway of care, patient flow is considered to be central to understanding and optimizing hospital performance and thus ensuring good patient care [1-2]. In a recently concluded study [3], poor communication and poor visibility to data were highlighted as the top two contributors to poor hospital patient flow. The need for tools to measure, predict and monitor events critical to patient flow was also identified as a top-ten initiative for addressing poor patient flow.

Visualization of health data is a well regarded strategy for improving patient flow [4-6]. Understanding how patient flow parameters interact with each other provides an effective mechanism for identifying bottlenecks. A ‘whole of hospital’ approach to understanding and overcoming these bottlenecks can help optimize the process and provide a better care environment for patients. Visualization and analysis of health data is however considered a difficult problem given the complex and multidimensional nature of the data.

Cluster analysis provides statistical techniques to group data based on common fields to reduce observations and simplify analysis. Time based clustering has successfully been used for segmentation and analysis in several fields including phonetics [7], photography [8], earthquakes [9] and routing in wireless sensor networks [10].

We propose a time-based clustering mechanism that employs a grouping based on patient admission/presentation and discharge times to cluster patient encounters into hourly intervals and utilizes these hourly interval ‘slots’ as a common index on which patient flow parameters from heterogeneous systems can be measured, visualized and compared. It also facilitates further grouping based on other fields, such as hospital occupancy levels calculated over these hourly intervals to perform more complex analysis thereby providing a powerful tool that can be employed to compare patient flow parameters and identify complex interdependencies between them. We have previously employed time based clustering to study the effect of discharge timing on inpatient occupancy and length of stay [11]. This paper details the methodology and the application of this technique to use inpatient and Emergency Department (ED) information to analyze ‘whole of hospital’ patient flow.

To demonstrate the efficacy of our approach, we raise some commonly asked questions about patient flow and describe how we can utilize time based clustering to address them. We present the results of our analysis and evaluate the findings. We also discuss ongoing efforts to use this methodology to support optimal patient flow and drive improved capacity management in hospital services.

II. Methods

A. Data Sources

For the purpose of this study, we obtained patient record data from two sources. Inpatient admissions and ED presentations data was obtained from 23 reporting public hospitals (representing approx 7800 beds) in Queensland, Australia, for an analysis period of 30 months (913 days), from October 2007 to March 2010. As a subset of this data, single-site data was also obtained for Caboolture Hospital, an urban facility with just under 200 beds in Brisbane, Australia. Picking a single hospital service allowed us to work more closely with clinicians in interpreting the findings. Inpatient admissions and ED presentations data was also obtained from the Royal Adelaide Hospital, a large tertiary hospital, with approximately 650 beds, in South Australia, for an analysis period of 24 months (730 days), from July 2009 to June 2011. This allowed us to generalize the usefulness of our approach by addressing varying models of health care services. Table 1 presents the list of hospitals allocated to each analysis group.

B. Ethics

Ethics approval for this research was obtained from the Queensland Health Central Human Research Ethics Committee and the Royal Adelaide Hospital Research Ethics Committee.

C. Measuring Patient Flow

Contemporary measures employed in patient flow research include bed occupancy levels, measures of patient access to a bed and wait times. In this study, occupancy was calculated using the ratio of occupied beds to census beds (i.e. rated capacity). It was measured over hourly and daily

S. Khanna, J. Boyle, and N. Good, are with the Australian E-Health Research Centre, Level 5, UQ Health Sciences Building 901/16, Royal Brisbane and Women’s Hospital, Herston, QLD, Australia (phone: +617 32533690; fax:+617 32533690; e-mail: Sankalp.Khanna@csiro.au, Justin.Boyle@csiro.au, Norm.Good@csiro.au).

J. Lind is with the Gold Coast Hospital, Queensland Health, QLD, Australia (e-mail: James_Lind@health.qld.gov.au)

K. Zeitz is with the Central Adelaide Local Health Network, South Australia Health, SA, Australia (email: kathryn.zeitz@health.sa.gov.au)

periods. Given the use of capacity management protocols such as temporary or over-census beds, hospitals often report occupancy in excess of 100%.

Overcrowding often causes boarding, i.e. ED patients who need admission waiting for a bed to become available. Access Block is defined as having occurred when a patient's boarding time exceeds 8 hours. The frequency of occurrence of Access Block serves as a measure of the level of boarding, and thus poor patient flow.

TABLE I. HOSPITALS ALLOCATED TO EACH ANALYSIS GROUP

Analysis Group	Hospital
23 Queensland (QLD) Hospitals	Royal Brisbane & Women's Hospital Princess Alexandra Hospital Gold Coast Hospital Prince Charles (The) Hospital The Townsville Hospital Cairns Base Hospital Nambour Hospital Ipswich Hospital Logan Hospital Toowoomba Hospital Redcliffe Hospital Rockhampton Base Hospital Caboolture Hospital Bundaberg Hospital Royal Children's Hospital Queen Elizabeth II Jubilee Hospital Mackay Base Hospital Redland Hospital Hervey Bay Hospital Gympie Hospital Maryborough Hospital Caloundra Hospital Gladstone Hospital
Caboolture Hospital	Caboolture Hospital
Royal Adelaide Hospital	Royal Adelaide Hospital

D. *Analysis Design*

To investigate the efficacy of the clustering protocol, we analyzed three flow related problems:

- What does a typical day at the hospital look like in terms of patient flow?
- How rapidly does patient flow change as hospital occupancy increases? How related are inpatient flow and ED patient flow to each other?
- Does increased inpatient occupancy result in increased ED Access Block?

Inpatient and ED datatsets were cleansed by removal of incomplete/inconsistent records. Hourly inpatient admission and discharge rates and length of stay were calculated by clustering patient episodes into hourly slots based upon their admission and discharge times. Hourly ED patient flow and the occurrence of Access Block were also calculated by clustering ED presentation episodes into hourly slots based upon presentation and disposition times. Daily patient flow was then calculated. This process allowed daily peaks and troughs in occupancy to be calculated at the granularity of hourly intervals. The hourly intervals, or 'slots', also served as a common index for overlaying the clustered flow information from inpatient and ED data sources.

To investigate flow parameters across a typical day, the hourly interval data was further grouped by hour of day to calculate various patient flow parameters including inpatient admission and discharge rates, ED presentation and discharge rates, inpatient admission from ED rates, and the frequency of occurrence of Access Block, and mean occupancy levels across the 24 hourly slots through a day. Average occupancy across the day was also calculated as a benchmark position to compare the spread and profile of occupancy variations through the day. This analysis was carried out on '23 QLD Hospitals' and 'Royal Adelaide Hospital' datasets.

Relationships between rates of hourly change in inpatient occupancy and ED occupancy were explored for the '23 QLD Hospitals' dataset by calculating changes in occupancy between each hourly interval and the preceding one and then reclustering this hourly interval data into 1% occupancy interval slots defined by the average inpatient occupancy for the hourly interval under consideration. The rate of change was calculated for both inpatient occupancy and ED occupancy by taking an average across each cluster.

To examine the impact of increased hospital occupancy on boarding and Access Block, we analyzed data from the 'Caboolture Hospital' dataset. Clustered inpatient and ED flow information was grouped into 1% occupancy interval slots corresponding to the average inpatient occupancy for the hourly interval under consideration. Access Block was calculated for each hourly interval to serve as a measure of boarding. To explore the association between Access Block and occupancy, we performed robust regression analysis between the proportion of Access Block cases (i.e. Access Block cases per hour) and occupancy. Robust regression was chosen to allow the model to better handle outliers, i.e. data points that deviate markedly from the rest of the data. Due to the relatively small number of hourly observations at low and high occupancy levels, we also weighted the regression (i.e. assigned a weight to each data point) using iteratively reweighted least squares.

III. RESULTS

Analyzing the interaction between patient flow parameters from '23 QLD Hospitals' across a day revealed various bottlenecks in flow. Figure 1 presents mean patient flow parameters as a function of hour of day for this analysis group. It was observed that inpatient admissions and inpatient discharges displayed variable flow patterns through the day, with both exhibiting peak flow behavior at different times. Further, the discharges peaked much later in the afternoon, contributing to higher levels of occupancy in the day. Peak inpatient occupancy was found to occur between 9am and 10am, coinciding with the peak in ED presentations. Access Block seemed to be fairly well spread out through the day, though it was about 20% higher between 10am and 10pm. Mean hourly occupancy returned a 95% confidence interval band of ±1.4%. Bed occupancy at midnight reflected near-minimum occupancy levels.

The 'Royal Adelaide Hospital' revealed subtle differences in trend when subjected to this same analysis (see Figure 2). While inpatient admissions and discharges still displayed variable flow patterns, admissions were much better spread out through the day while discharge rates peaked earlier in

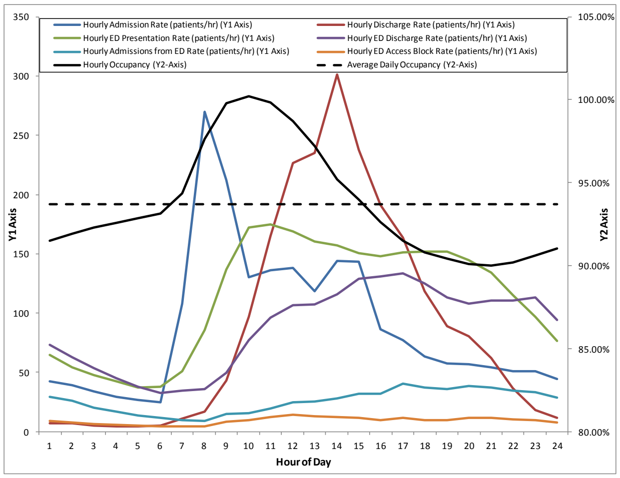

Figure 1 : Mean Hourly Patient Flow - 23 Queensland Hospitals

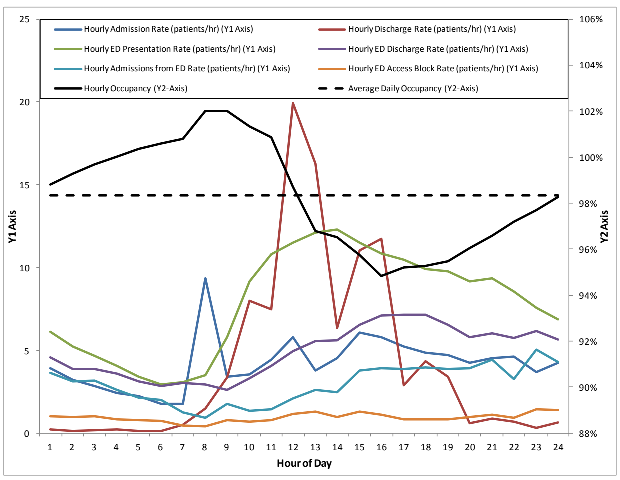

Figure 2 : Mean Hourly Patient Flow - Royal Adelaide Hospital

the day compared to the QLD analysis. Given this, peak inpatient occupancy levels were experienced between 7am and 8am, and had returned to below average levels before ED presentations peaked later in the day. This resulted in an mean hourly occupancy profile that exhibited a lower range of values and a narrower 95% confidence interval of ±0.9%. Unlike in the case of QLD analysis, the midnight occupancy reflected near-average occupancy levels.

Analyzing the relationships between the rate of hourly change in inpatient and ED occupancy as inpatient occupancy increased across '23 QLD Hospitals' revealed a strong correlation between these parameters (ρ=0.924). Analyzing the scatter plot between these parameters (see Figure 3) confirmed these findings with a strong association between these parameters. As occupancy rose, both parameters also returned rapidly increasing average rates of hourly change in occupancy (see Figure 4), with ED occupancy leading the trend over inpatient occupancy.

Examining Access Block as a function of rising occupancy at 'Caboolture Hospital' did not reveal any relationship between the parameters. Figure 5 presents the average and maximum Access Block cases per hour as a function of occupancy level for this dataset. 95% confidence intervals are also plotted for the case of average Access Block cases per

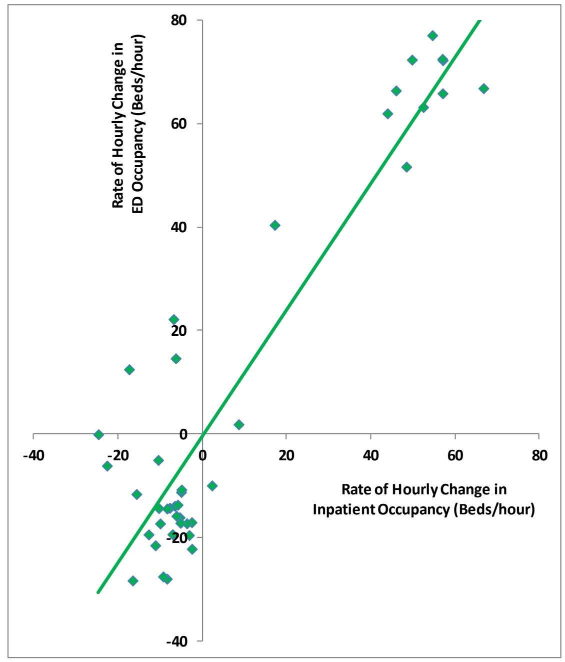

Figure 3 : Relating Hourly Changes in ED and Inpatient Occupancy

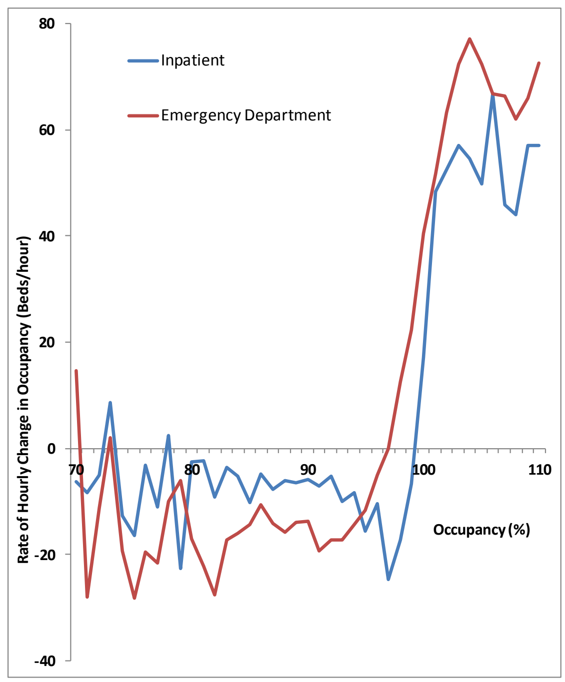

Figure 4 : Rate of Hourly Change as a function of Inpatient Occupancy

hour. We see from this figure that Access Block was reported at all levels of hospital occupancy. Analyzing the correlation between these parameters indicated that little or no correlation existed between hospital occupancy level and Access Block (ρ=0.351). These findings were confirmed by the regression analysis. Figure 6 presents the robust regression model of Access Block and occupancy. Based on the slope of this regression fit, we conclude that Access Block is not overly influenced by hospital occupancy in the data analyzed.

IV. DISCUSSION

The key contribution of this study is a novel temporal grouping approach that clusters patient encounters into hourly intervals using admission and discharge timestamps. Hourly interval flow data is then utilized for measuring, visualizing and comparing patient flow parameters. It is also grouped on analysis parameters like occupancy levels for more complex analysis. To demonstrate the efficacy of our approach, we posed some typical patient flow queries and sought to answer them using our methodology.

Analysing interaction between patient flow parameters across a day allows us to identify times when hospitals exceed average levels of occupancy. Observing the relative position of the admission and discharge peaks allows us to interpret the relationship between admission and discharge flow and peak occupancy, and provides evidence to support initiatives like early patient discharge and staged admissions that can help improve flow. Analyzing the relative position of average occupancy with respect to minimum, maximum and midnight levels also provides a better understanding of how effective the midnight census may be in truly representing patient flow in the hospital. The differences in flow patterns across the analysis groups suggests that interactions vary across various sites and analysis needs to be done at a service level to better inform capacity management initiatives.

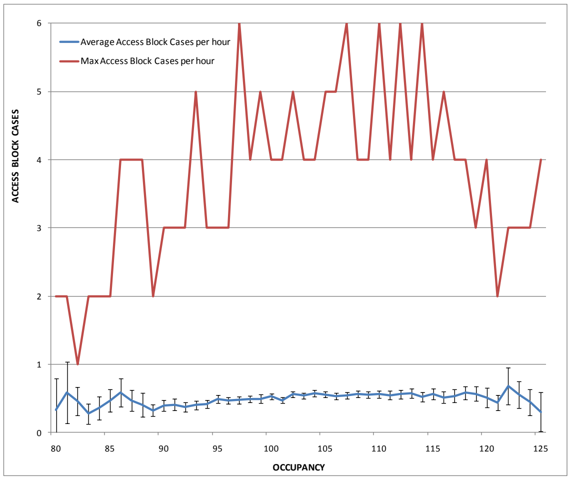

Figure 5 : Access Block (with 95% CI) as a function of Occupancy

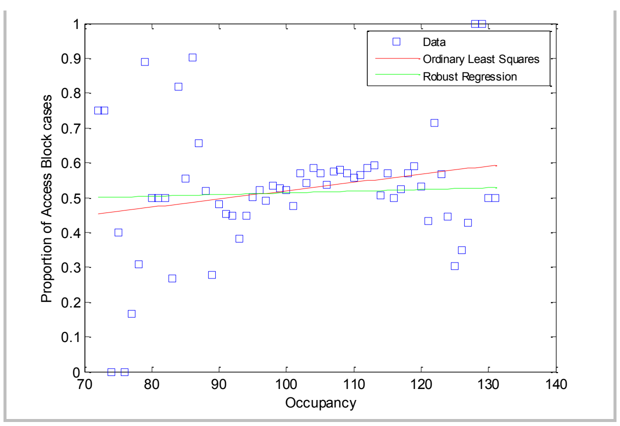

Figure 6 : Robust Regression Model of Access Block and Occupancy

Visualizing the relationship between the rate of hourly change in inpatient and ED occupancy reveals a correlation between the two flow parameters and confirms that ED overcrowding is closely related to hospital crowding. Plotting these against rising occupancy however reveals that ED occupancy starts trending towards crowding faster than hospital occupancy, suggesting the need for further analysis.

Investigations into the association between boarding and hospital occupancy suggests no meaningful relationship between Access Block and occupancy at the chosen hospital. Further analysis into this relationship across numerous hospital sites is recommended.

The performed investigations establish the efficacy of our time based clustering approach for patient flow visualization and analysis. We are currently working towards using this methodology to perform more complex analysis on 'whole of hospital' patient flow, including analyzing the complex interdependencies between patient flow parameters like inpatient admission and discharge rates, quality indicators like length of stay and Access Block, safety indicators like number of adverse incidents, and mortality. We are also employing this methodology to drive simulations to analyze and quantify the potential impact of strategies like early discharge, and to test and design optimal capacity management strategies for individual hospital services.

REFERENCES

[1] K. Konnyu, L. Turner, B. Skidmore, R. Daniel, A. Forster, and D. Mohan, '*What input and output variables have been used in models of patient flow in acute care hospital settings?*', Ottawa Hospital Research Institute, 2011, http://www.ohri.ca/kta/docs/KTA-Patient-Flow-Evidence-Summary.pdf [last accessed March 2012]

[2] Scottish Executive. '*A guide to service improvement: Measurement, analysis, techniques and solutions*', St Andrew's House, Edinburgh, 2005, http://www.scotland.gov.uk/Resource/Doc/76169/0019037.pdf [last accessed March 2012]

[3] AHA Solutions, 'Results and Report of the 2012 Patient Flow Challenges Assessment: Hospitals Consider Patient Flow Essential to Care and Competitiveness', 2012, http://www.aha-solutions.org/what/pfca.shtml [last accessed March 2012]

[4] E. Chazard and R. Beuscart, 'Graphical representation of the comprehensive patient flow through the Hospital', *AMIA Annu Symp Proc*, vol. 2007, pp. 110–114, 2007.

[5] J. J. Thomas and K. A. Cook, 'A visual analytics agenda', *IEEE Computer Graphics and Applications*, vol. 26, no. 1, pp. 10– 13, Feb. 2006.

[6] J. A. Fitzgerald and A. Dadich, 'Using visual analytics to improve hospital scheduling and patient flow', *J. Theor. Appl. Electron. Commer. Res.*, vol. 4, no. 2, pp. 20–30, Aug. 2009.

[7] B. Eberman and W. Goldenthal, 'Time-based clustering for phonetic segmentation', in , *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings*, 1996, vol. 2, pp. 1225–1228 vol.2.

[8] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox, 'Temporal event clustering for digital photo collections', *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 3, pp. 269–288, Aug. 2005.

[9] S. Hainzl, G. Zöller, and J. Kurths, 'Self-Organization of Spatio-Temporal Earthquake Clusters', in *Nonlin Proc Geophys*, 2000, vol. 7, pp. 21–29.

[10] O. Zytoune, Y. Fakhri, and D. Aboutajdine, 'Time based clustering technique for routing in wireless sensor networks', in *2011 International Conference on Multimedia Computing and Systems (ICMCS)*, 2011, pp. 1–4.

[11] S. Khanna, J. Boyle, N. Good, and J. Lind, 'Impact of admission and discharge peak times on hospital overcrowding', *Stud Health Technol Inform*, vol. 168, pp. 82–88, 2011.