

Data driven process modelling for a hospital emergency department

Andrzej CEGLOWSKI, Leonid CHURILOV, Jeff WASSERTHIEL

Monash University Melbourne, Australia

Abstract. This paper describes how key activities in the emergency department of a major hospital were extracted from workflow history. Analysis of these activities help with modification of both administrative and clinical actions for improved efficiency and effectiveness. Extraction of process from data is a relatively new field. This paper's contributes the innovative determination of processes through data mining, rather than the algorithm-driven approach used to date. Data about patients who present to a major hospital emergency department were used to define clusters of patients who follow common pathways through the emergency department. It is discussed how these "process based" clusters can be used for performance management of the emergency department through evaluation of process inputs, outputs and costs.

1 Introduction

Australian federal and state governments provide funding for public hospitals determined primarily on performance or output, rather than negotiation, history or politics. Clinical and resource homogeneous groups of patients are determined from stored information about patient visits and related to the resources required (Duckett 1998). Homogenous grouping of patients have become known as "casemix" to emphasise the grouping based on similar patient "cases".

The casemix approach has been reasonably successful in predicting resource requirements for inpatient acute care settings, and it now forms a significant part of improvement and management activities (Australian Department of Health and Aging 2003). However, classification of patients who present to emergency departments (a hospital department that specialises in providing care for people who are in need of urgent care) has proven to be difficult, with the best groupings only accounting for some 60% of cost (Bond et al. 1996).

Casemix for emergency department (ED) patients is important because the ED is one of the main routes for admission into Australian hospitals and is becoming a primary source of health care. There have been large increases in presentations to EDs in recent years (Acute Health Division 2001), leading to longer waiting times, patients being directed to alternate facilities, and other issues that have the potential to affect the ability of the ED to save lives. Analyses that treat the ED as a component within a complex healthcare system and the simulation of patient flows within EDs

have contributed greatly to understanding of ED dynamics (Lane et al. 2000), but the absence of an acceptable patient classification limits the accuracy of these methods and ability to satisfactorily account for resource use.

Traditionally, casemix has been based on a combination of clinical information (diagnoses and procedures) and demographic information (age and sex), to result in homogeneous groups with respect to a target variable such as pattern of illness or treatment (Jelinek 1995). Generally the similarities between patients relate to diagnosis, working under the assumption that patients with related diagnoses follow a matching course of treatment and utilise comparable resources. Essentially, casemix strives to yield treatment pathways for patients without explicitly defining the processes incorporated in those pathways – patients are grouped by function (diagnosis), yet the groups are expected to yield a process perspective with associated inputs, outputs and resource requirements.

Since the ED forms such a significant part of the healthcare chain, both in terms of number of patients and potential for life-critical incidents, it is the objective of this paper to present a more effective approach to classification of ED patients. This approach takes a process view. Patients who follow similar processes are likely to consume similar resources. A process based classification can be used to improve understanding of patient flows through the ED, and help with facility design, information system design, resource allocation, reengineering of processes, and training of staff.

The approach described in this paper is a fundamental departure from existing casemix for ED patients, presented as follows: Section 2 provides background to the problem, and looks at related research. The data and methodology is explained in Section 3. Section 4 supplies the clustering results and compares them to existing proposals. Section 5 discusses the insights supported by these results and mentions extensions to the work. The paper concludes with a caution about ED process modification.

2 Background and previous work

There has been much simulation and systems research into hospitals and healthcare (Jun et al. 1999; Preater 2002) in an effort to prevent excessive patient waiting times and the redirection of patients and ambulances to other ED facilities. The general conclusion has been that ED problems cannot be treated in isolation as this simply moves the pressure point within the healthcare system. (Lane et al. 2000; Acute Health Division 2001).

Improving the efficiency and effectiveness of public hospital services in Victoria is being addressed by the “Designing Care” program which aims to redesign processes across the whole health system (Victorian Department of Human Services 2002a). The ED component of “Designing Care” emulates and duplicates ED initiatives that have been successful in other countries and at other hospitals in Australia. These include “fast tracking” of certain patients, decreasing ED volume, and providing

increased supervision of junior medical staff (Victorian Department of Human Services 2002b).

Much work has been done in Australia on determination and agreement of casemix for inpatient classification (Hanson 1998; Australian Healthcare Association 1999; Funding & Financial Policy Branch 2002). Work started in late 2002 on a national ED patient classification, with initial efforts concentrating on identification of appropriate ED data to include (McAlister 2003). Patient classifications have been proposed to aid with ED performance evaluation (Cameron et al. 1990). Characteristically, proposals have grouped ED patients according to combinations of age, urgency of complaint, diagnosis, time in ED and outcome of visit.

A Perth study recorded diagnoses and urgency for ED patients attending four hospitals to develop the typical casemix for the hospitals. In a second phase, resource use was measured for patient attendees to ED and related to the typical casemix (Jelinek 1995). In a later Flinders study, costs were measured for some 17800 patients. Key variables were identified by univariate analysis as cost drivers for ED patient attendance. The cost drivers were urgency, outcome, age, diagnosis and treatment time. A classification tree was built from the cost drivers to determine the minimum number of clusters that could account for most costs (Bond et al. 1996). These classifications are inadequate to describe a significant number of activities within EDs (Table 1).

Table 1: Comparison of cost variance reduction in ED casemix (Bond et al. 1996)

ED Casemix system	Flinders Medical Centre Study (1996)	Perth Study (1992)
Urgency and Disposition groups (UDGs)	43.9%	47.4%
Urgency Related Groups (URGs)	55.3%	57.6%

Data mining and neural networks offer alternative approaches to data analysis. Cullen (2001) used data mining for intelligent feature selection in healthcare. Other data mining in healthcare research relates to investigation of symptoms and treatment (Brossette et al. 2000; Riano and Prado 2000; Lin et al. 2001; Richards et al. 2001; Isken and Rajagopalan 2002; Lee et al. 2002; Williams et al. 2002; Chae et al. 2003). Abston (1999) applied neural networks and other methods to model the pharmacological management of acute myocardial infarction in an emergency department and concluded that the data most descriptive of and pertinent to clinical decision-making seems to be left out of data collected each day in the clinical setting. Abston's conclusion highlights the difficulty of grouping ED patients according to clinical decisions and underscores the need for a change in approach from classic casemix models.

Since this paper involves a process-driven approach to clustering, it is necessary to introduce the relatively new area of process mining. Process mining involves the analysis of data about a process to learn about underlying patterns of activity (List et al. 2001). The result of this analysis are patterns of activity that are objective because they are based on the actual things that took place (Department Technology Management 2003). It is possible to identify the most frequent pathways through a process. Each of these key pathways may be viewed as recurring patterns of activity that may be analysed to identify inputs, outputs and cost structures, or to identify clusters of transaction types.

The context of the problem and the preceding works led to questioning whether clusters of activity could be extracted from ED data to yield homogenous clusters of ED patients. The activities involved in treatment would be the same for each cluster so each cluster could be considered to have matching inputs, outputs and resource consumption. The activities associated with each cluster would comprise activities in *the process of treating* patient instances within that cluster, so process and workflow perspectives could be used to improve understanding of patient flows through the ED. This is process mining with a view to achieving casemix outcomes. The data used and methodology is discussed in the next section.

3 Methodology

The data came from a major city hospital who is partner in this project. The data was comprised of de-identified records of all ED presentations between 1999 and 2002. These records uniquely identified each visit by an ED reference number, and retained codes that permitted identification of repeat visits. The records contained demographic information plus details of the visit such as apparent severity of complaint, key time points and outcome. Initial investigations were limited to random samples within the 56906 records in the 2002 cohort to limit effects of inter-year changes to activities within ED.

It has been seen in the preceding section that previous attempts at identifying casemix for ED patients grouped patients by cost based on urgency and diagnosis, sometimes combined with demographic information, such as age. Since cost data was not available, it was not possible to duplicate past studies, however effort was made to emulate the groupings using Classification and Regression Trees (CART) and Self Organising Maps (SOM).

CART and SOM are nonparametric grouping methods that seek to minimise diversity within groups and maximise differences between dissimilar groups. The grouping is algorithm-driven, not supervised, so is often referred to as “self-organisation”. Nonparametric grouping relies on data, rather than domain-specific expertise. The methods generally employ large datasets, work well with many input variables and produce arbitrarily complex models unlimited by human comprehension (Kennedy et al. 1998).

- The CART algorithm builds a binary decision tree through brute force. It performs splits based on an exhaustive search of all variables to find an optimal splitting rule for each node. The resultant tree is then pruned to improve overall classification accuracy (Kennedy et al. 1998).
- SOM provide a visual understanding of patterns in data through a two dimensional representation of all variables. Records that have similar characteristics are adjacent in the map, and dissimilar records are situated at a distance determined by degree of dissimilarity. The SOM algorithm repeatedly repositions records in the map until a classification error function is minimised (Kohonen 1995).

In order to facilitate a process-focused approach, a separate data file was obtained that contained the ED reference number linked to one of 57 procedures (investigations such as blood analyses and x-rays, or activities related to treating the patient such as suturing). This data was combined with the records of ED presentations so that each

record now contained demographic and visit information, plus all procedures undertaken during that visit. Working under the presumption that resource use for each patient could be linked to number and type of procedures, it was hoped that discrete groups of procedures could be identified across all records with two or more procedures (Table 2) that would result in “primary pathways” patients take through the ED, in essence providing a set of core processes that account for the majority of work performed in the ED. Patients could be clustered according to the pathways they followed.

Table 2: Overview of ED data used in defining core ED activities. Note that 10 procedures account for the majority of presentations in patients who underwent only 1 procedure.

Description	Count of records
Two or more procedures (including duplicated procedures)	44600
One or no procedures (*)	12211
<i>Top 10 procedures in records with 1 procedure (99% of *)</i>	(11537)
<i>Top 30 procedures in records with 1 procedure (99.9% of *)</i>	(12199)
Missing or corrupted records	95
Total number of records	56906

A second attempt was made to confirm past groupings. These process-based clusters were associated with demographic variables and details about the ED visit, such as whether the patient was injured or not, time spent in ED and outcome.

The ED records were manipulated within SPSS (2001) and SOM investigations were done using Viscovery SOMine software (1999). The results of the above three investigations are presented in the next section.

4 Results

In trying to emulate previous studies, no satisfactory clustering could be achieved, regardless of the variable(s) used in clustering, whether urgency, diagnosis, presenting problem, outcome or other data. Clusters contained a full demographic sweep of patients without any definitive variables. There were isolated pockets of correlation but these were insufficient to satisfy casemix requirements.

When a process-mining approach was tried, 41 clusters of procedures were found. 21 of these clusters accounted for 96.6% of presentations, while just 14 clusters accounted for over 90% of ED presentations. This means that 14 “primary pathways” could be identified that 90% of patients follow. In addition to this remarkable result, 18 procedures could be omitted from future analysis because they did not contribute to the primary clusters. New maps were generated after removing the 18 procedures and 27 clusters identified. Once again just 14 clusters incorporated key pathways for some 90% of ED visits (Table 3).

Table 3: Primary clusters for patients who have 2 or more procedures

Description	Abbrv.	Clusters for patients with 2 or more procedures													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
Observation	o	X	X	+	x	x	+	+		X	x	+	+	+	+
Venipuncture	vb	x		X	+	x		X	+	x			X		x
Drug (Oral/Sublingual/Optical/Rectal)	drug	+	X	x	x	+	+	x	+	+	x	+	+	+	+
X-ray	xray	+		+		+	+	x	+	+	+	X	+		
Peripheral IV Catheter	iv	+		X		x		x		X			x		+
12 Lead ECG	ecg	+				+		X		+			x		
Infusion of IV fluid (not blood)	inf			X						+					
Full ward test of urine	fwu				X										
CT Scan	ct					X									
Dressing	drs						X							x	
ECG Monitoring	ecgm							X							
Head Injury Observation	hio								X						
IV Drug Infusion	ivi									X					
Nebulised Medication	neb										X				
Plaster of Paris	pop											X			
Random Blood Glucose	rbg												X		
Suture, Steristrip, Glue	sut													X	
Ultrasound	uls														X
Patients with 2 or more procedures (%)		20.3	14.8	10.6	8.1	4.6	3.7	5.8	3.1	4.4	4.7	2.0	2.6	4.2	2.5

Key:

X: Over 80% Patients in this cluster underwent this procedure

x: Between 60% and 80% of patients in this cluster underwent this procedure

+: Between 40% and 60% of patients in this cluster underwent this procedure

Demographic and ED visit details were overlaid on the process-based clusters to check whether past casemix groupings correlated to the process-based groups. There was almost no correlation between number and type of procedures (which act as proxies for resource use) and factors such as age, sex, injury, urgency, time in ED and outcome. The impact of these results on ED processes are discussed in the next section.

5 Discussion

The failure to find discrete groupings of ED patients based on traditional casemix approaches highlights the reason behind the inability of these groupings to account for even 60% of ED patient costs. Although it may seem logical to link patients according to diagnosis, it is likely that the treatments (and resource use) vary considerably, even within diagnosis groups.

While there is not space to discuss the process-based clustering results at length, a few points of interest may be indicated. Common sense would dictate that X-rays and Plaster of Paris would frequently be paired as activities of a single process, and this is seen in Cluster K. Similarly, it would be expected that ECG and ECG monitoring occur as part of the same process, as seen in Cluster G. The principal procedures (indicated by “X” in Table 3) of the 14 main clusters do not overlap with the most common procedures in patients that had only a single procedure (with the exception of “Observation” and “Drug administration” procedures, which are rather generic), so the clusters reflect complete processes, rather than extensions of individual procedures.

Differentiation between clusters may seem trivial if only principal procedures within each cluster are compared, but it must be remembered that the secondary procedures within each cluster (indicated by “x” and “+” in Table 3) provide insight about underlying patterns and similarities between patients in that grouping. It is these patterns that supply the necessary information about the overlap of process and clinical activities. For example, in Cluster K, patients often receive some form of drug (clinical treatment), are transported to the X-ray department (an activity supported by typical process views), are examined and have bones set and Plaster of Paris applied (clinical treatment).

The results have shown that discrete groups of ED patients can be identified that satisfy the casemix requirements of “a reasonable number of clinically meaningful resource homogenous groups based on data that is simple and easy to collect” (Bond et al. 1996):

- The 14 clusters compare well in number to the dozen or so used in previous works, yet account for over 70% of visits to the ED. Over 90% of all visits to the ED can be accounted for by supplementing these 14 clusters with data about the 10 most frequently used single procedures.
- The clusters are certainly clinically meaningful, since they reflect an “as is” analysis of activity in the ED.
- The clusters are resource homogenous in terms of number and types of procedures. Variations within procedures themselves may contribute to some variance, but the clusters allow this variation to be analysed in a meaningful way.
- Since the data is currently being routinely collected, no extra load is placed on staff to collect data, and the variables are defined in a standard and clear manner. The data is formatted to standards that will soon be national, so collation and comparison of datasets should be simple.

There is little potential for manipulating casemix for profit, or gaming. Since the clusters reflect current activities, any sudden change in activities could be detected by referring to earlier data.

The casemix requirements above overlap with requirements for definition of business processes, and there are a number of process related implications. In general, it may be considered that each patient visit to an ED triggers a sequence of activities aimed at improving their well-being while meeting multiple other objectives such as economic sustainability, disaster contingency and minimal stress for staff. These activities have a business process component that relates to patient administration and workflow, and a clinical component that is complex and variable. Although the primary pathways identified by the process mining approach in this paper are not processes as defined by Davenport (1990), in that there are no

predecessor/successor relationships, they do provide groups of procedures whose individual and cumulative inputs, outputs and costs can be evaluated.

While the immediate benefits to the ED of this work (in terms of real process modifications) have yet to be realised, extensions to the work exist. Clinical business processes for this ED have been modelled in detail using ARIS (Djordan and Churilov 2003), and there is a large library of “clinical pathways” that represent best practice in treatment of numerous diagnoses (Lin et al. 2001). The key pathways identified in this work provide a link between many business and clinical process. It is likely that a “matrixed” view of the ED may be modelled that combines these process and clinical views.

Patient flows in this ED have been modeled using discrete event simulation (Liew et al. 2003). The logical groupings of patients provided in this work will be used to enhance the “granularity” of this simulation model to improve understanding of patient flows and the impact of emergencies on resources.

6 Conclusion

EDs strive for balance between efficiency (more patients may be treated), and effectiveness (quality of care and rapid patient recovery). Previous attempts to identify urgency-related casemix groups that allow for measurement of efficiency and effectiveness in the ED have not been successful. The complexity of clinical treatment and the patient well-being imperative make pure process driven views of ED clinical operations impossible. This paper explained the melding of process and casemix approaches to determine a small number of “primary pathways” – core sets of activities for the ED.

It should be remembered that the intention of ED facilities is to provide timely care, given the urgency of the case, and to retain quality of care, even when the ED is operating at capacity. Any modifications to EDs must be examined in light of these clinical prerogatives. Unlike casemix approaches that artificially group patients based on cost and clinical observations, the data driven approach presented in this paper provides insight into *actual* core procedures, so provides a low-risk avenue for re-engineering of ED processes.

References

- Viscovery SOMine Standard Edition 3.0 (1999) Wien Eudaptics Software GmbH
- SPSS for Windows Release. 11.5.1 (2001) Chicago SPSS Inc.
- Abston K C (1999) PhD *Using the electronic medical record to predict the pharmacological management of acute myocardial infarction* Salt lake City University of Utah
- Acute Health Division (2001) *Emergency Demand Management: A new approach* Melbourne, Victorian Government Department of Human Resources: Feb
- Australian Department of Health and Aging (2003) *Annual Report 2002-'03* <http://www.health.gov.au/index.htm> (Australian Department of Health and Aging) Accessed: 5 Dec 2003

- Australian Healthcare Association (1999) "Special Issue on the George Palmer Symposium" Australian Health Review **22** (2)
- Bond M, Baggoley C and Erwich-Nijhout M (1996) *Costings in the Emergency Departments A Report for the Commonwealth Department of Health and Family Services, South Australia*
- Brossette S E, Sprague A P, Jones W T and Moser S A (2000) "A data mining system for infection control surveillance" *Methods Inf. Med.* **39** (4-5): 303-310
- Cameron J, Baraff L and Sekhon R (1990) "Case-mix classification for emergency departments" *Medical care* **28**: 146-158
- Chae Y M, Kim H S, Tark K C, Park H J and Ho S H (2003) "Analysis of healthcare quality indicators using data mining and decision support system" *Expert Syst. Appl.* **24** (2): 167-172
- Cullen P (2001) Ph.D *Feature selection methods for intelligent systems classifiers in healthcare* Chicago, Loyola University of Chicago
- Davenport T H and Short J E (1990) "The new industrial engineering: information technology and business process redesign" *Sloan Management Review* **31** (4): 11-27
- Department Technology Management (2003) *Process Mining* http://tmitwww.tm.tue.nl/research/process_mining.shtm (Technische Universiteit Eindhoven) Accessed: 8 Dec 2003
- Djordan V and Churilov L (2003) "Business interactions in an acute emergency department in Australia: A clinical process modeling perspective" *The 6th Pacific Asia Conference on Information Systems (PACIS 2002)* Tokyo, The Japan Society for Management Information
- Duckett S J (1998) "Casemix funding for acute hospital inpatient services in Australia" *The Medical Journal of Australia* **168**: S17-21
- Funding & Financial Policy Branch (2002) *Casemix Funding in Victoria* <http://casemix.health.vic.gov.au/index.htm> (Victorian Department of Human Services) Accessed: 4 Dec 2003
- Hanson R M Ed. (1998) *Casemix: Moving forward* The Medical Journal of Australia Sydney, The Australian Medical Association
- Isken M W and Rajagopalan B (2002) "Data mining to support simulation modeling of patient flow in hospitals" *J. Med. Syst.* **26** (2): 179-197
- Jelinek G A (1995) *A Casemix information system for Australian Hospital Emergency departments A Report to the Commissioner of Health, Western Australia*
- Jun J B, Jacobson S H and Swisher J R (1999) "Application of discrete-event simulation in health care clinics: A survey" *Journal of the Operational Research Society* **50** (2): 109-123
- Kennedy R, Lee Y, Van Roy B, Reed C and Lippmann R (1998) *Solving data mining problems through pattern recognition*, Prentice Hall
- Kohonen T (1995) *Self-organizing maps* Berlin ; New York, Springer
- Lane D C, Monefeldt C and Rosenhead J (2000) "Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department" *Journal of the Operational Research Society* **51**: 518-531
- Lee I N, Liao S C and Embrechts M J (2002) "Important variable selection techniques with multiple solutions for medical information applications" *Med. Inform. Internet Med.* **27** (4): 253-266
- Liew S K, Churilov L and Brailsford S (2003) "Treating ailing emergency departments with simulation: an integrated perspective" *International Conference on Health Sciences Simulation* Orlando, Florida, The Society for Modeling and Simulation International
- Lin F R, Chou S C, Pan S M and Chen Y M (2001) "Mining time dependency patterns in clinical pathways" *Int. J. Med. Inform.* **62** (1): 11-25
- List B, Schieder J, Min Tjoa and Quirchmayr G (2001). "Multidimensional business process analysis with the business warehouse" *Knowledge discovery for business information systems* Abramowicz W and Zurada J Eds. Boston, Kluwer Academic Publishers: xvii, 431
- McAlister S (2003) *Work on the National Minimum Data Set for Non- Admitted Patient Emergency Department care (NAPED NMDS)* Personal communication to Ceglowski A

- Preater J (2002) "A Bibliography of Queues in Health and Medicine" Health Care Management Science
- Riano D and Prado S (2000). "A data mining alternative to model hospital operations" *Medical Data Analysis*. Berlin, Springer-Verlag. **1933**: 293-299
- Richards G, Rayward-Smith V J, Sonksen P H, Carey S and Weng C (2001) "Data mining for indicators of early mortality in a database of clinical records" *Artif. Intell. Med.* **22** (3): 215-231
- Victorian Department of Human Services (2002a) *Designing Care* <http://designingcare.health.vic.gov.au/> (Victorian Department of Human Services) Accessed: 6 May 2003
- Victorian Department of Human Services (2002b) *Summary of findings from project annual reports: Hospital demand management strategy 2001-2002* Melbourne, Metropolitan Health & Aged Care Services Division: Nov
- Williams G, Baxter R, Kelman C, Rainsford C, He H X, Gu L F, Vickers D and Hawkins S (2002). "Estimating episodes of care using linked medical claims data" *AI 2002: Advances in Artificial Intelligence*. Berlin, Springer-Verlag. **2557**: 660-671

