Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach

Lua Perimal-Lewis  
School of Computer Science, Engineering and Mathematics  
Flinders University of South Australia  
GPO Box 2100, Adelaide 5001, South Australia  
lua.perimal-lewis@flinders.edu.au

Shaowen Qin  
School of Computer Science, Engineering and Mathematics  
Flinders University of South Australia  
GPO Box 2100, Adelaide 5001, South Australia  
shaowen.qin@flinders.edu.au

Campbell Thompson  
Medicine  
The University of Adelaide  
5005 South Australia  
campbell.thompson@adelaide.edu.au

Paul Hakendorf  
Clinical Epidemiology Unit, Flinders Medical Centre  
Flinders Drive, Bedford Park 5042 South Australia  
Paul.Hakendorf@health.sa.gov.au

Abstract

Hospitals are continually struggling to cater for the increasing demand for inpatient services. This is due to increased population, aging, and the rising incidence of chronic diseases associated with modern life. The high demand for hospital services leads to unpredictable bed availability, longer waiting period for acute admission, difficulties in keeping planned admission, stressed hospital staff, undesirable patient and family experience, as well as unclear impact on the quality of care patients receive. This study aims to gain insight into patient journey data to identify problems that could cause access block. Process mining techniques combined with statistical data analysis are adapted to discover inpatient flow process patterns and their correlation with patient types, ward types, waiting time and Length of Stay (LOS). Open source process mining software, ProM, is used in this study. The study is done in collaboration with Flinders Medical Centre (FMC) using data from their Patient Journey Database.

Keywords: patient flow, inpatient journey analysis, process mining, length of stay

1 Introduction

Australian Public Hospitals are generally patient-centred organisations. Hospital managements need to constantly look for ways to improve their patient care processes in order to address challenges caused by the ever-increasing demand for hospital services. One such challenge, which is often in media’s limelight, is the Emergency Department (ED) overcrowding also known as access block.

Fatovich (2002) describes ED overcrowding as a worldwide problem where an ED is unable to provide timely emergency care and as a consequence ambulances are instructed to divert to another facility. The reason for ambulance diversion is simply because of the lack of capacity to safely attend to newly arrived patients.

O’Connell, Bassham et al. (2008) asserts that ED congestions are intensified by regular failure to manage processes involved in progressing patients through the hospital. According to O’Connell, Bassham et al. (2008) better inpatient management and better patient flow will mitigate some of the issues faced by ED.

The study aims to investigate the Flow of Patient or Care Flow within a hospital with the understanding that smooth flow would reduce bottlenecks in hospitals. Smooth patient flow will have positive influence to the overall hospital capacity that would consequently ease ED over-crowding and other systemic issues in a hospital such as bed availability.
2 Background

Public Hospitals are required to conform to certain Key Process Indicators (KPIs). Such conformances are essential because of the competitive nature of government funding and the need to provide justified information to taxpayers. Hospitals do already have mature processes and the ability to report and monitor performances with statistics. However, such data does not show how to improve a process. The term Clinical Process Re-engineering has become a common term when referring to clinical process improvements. Clinical Process Re-engineering could be considered similar to Business Process Re-engineering (BPR). Both initiatives focus on continuous improvements of business processes or clinical processes to gain “competitive advantage”. However, this “competitive advantage” from a hospital’s perspective is service oriented and increasingly becoming patient-centred, endeavouring to service the health care needs of its population. BPR aims at improving core business process and in the hospital setting one of the core business processes can be regarded as the patient journey.

Clinical process redesign is the use of process redesign and change management to health care (Ben-Tovim, Dougherty et al., 2008). The activities in the redesign process focuses on the patient’s perspective. The aim of clinical process redesign is to harmonise the poorly coordinated patient journey as they move across multiple departments, making them simpler whilst looking at the overall design of the clinical processes. Clinical process redesign takes a holistic approach by looking at a wider area in the redesign process, which requires continuous fine-tuning and adjustments to constantly adapt to the ever-transient nature of the hospital system.

Many hospitals around the world have adopted “lean thinking” methodology to improve patient care and have seen significant improvements in their clinical processes. One such success story, according to Richard C (2011), is experienced at Denver Health in America where over the last 5 years, the lean methodology has contributed to revenue increase and decrease in expenses. FMC’s Redesigning Care program has also adopted the concept of “lean thinking”, which enabled the hospital to provide safer and more accessible care (Ben-Tovim, Bassham et al., 2008). Whilst “lean thinking” has contributed to better understanding of the patient flow process and therefore better co-ordination, implementing this concept alone in isolation might not be enough to relieve the hospital-wide crisis where bed availability is still a concern.

Health Care data analysis is traditionally done using various statistical techniques in order to report and hopefully forecast health care performances. New approaches in health care modelling are emerging where more than one technique and approach are used to discover hidden information that might not be easily discovered from one approach. Combinations of techniques are used to complement each other.

Cegłowski, Churilov et al. (2007) proposed combining Data Mining techniques and discrete event simulation for identifying bottlenecks in the patient flow between ED and a hospital ward by providing insight into the complex relationship between patient urgency, treatment and disposal and the occurrence of queues for treatment.

The use of Decision Support System (DSS) in health care is wide spread. DSS in Health Care industry could be divided into 2 broad categories. One category of DSS is used to help physicians with their day-to-day decision makings. An example is a DSS based on clinical practice guideline in the management of diabetic patients (Lobach and Hammond, 1997). The other category of DSS is used by hospital management to make decisions for better hospital resource management. The fundamental information needed for such a system is based on the outcomes of some method of data analysis and modelling. The closer the outcome is in depicting the real scenario the better the DSS output.

Based on this background information and knowledge of common strategies being used by hospitals to better manage their patient journeys, this study aims to build on the existing resources and knowledge gained from some of these practices. The process mining activities proposed under this study will complement the „lean thinking“ strategies currently being practiced at FMC. An in-depth evaluation of the patient flow processes using the same Patient Journey Database will aid in identifying process bottlenecks hidden within the statistics.

3 Data

The Patient Journey Database from FMC records information on the journey or movement of a patient from the time of admission to the time of discharge. Therefore, it only contains information on inpatients or officially admitted patients.

Each admission is given a unique journey number that would remain the same until discharge. Each movement of the patient from one ward to another ward is recorded with a timestamp, so at any point the „start time“ in a ward and the „end time“ in a ward is known together with the ward name being occupied. Each journey is also linked to the doctor treating the patient. The wards are a subset of Units and the Units are a subset of Division. Each journey is also given a status of Inliers, Outliers or Inliers/Outliers. An individual patient could have multiple admissions at various time frames and each of this admission will be allocated with new unique journey number. Timestamp for Admission is the combination of “Date” field and the “Admission Time” field. Timestamp for Discharge is the combination of “Date” field and “Discharge Time” field. The table below shows the relevant fields taken from the Patient Journey Database pertinent to process mining (see Figure 1).

![Figure 1: Snippet of data used for process mining](image-url)
One of the notable criteria of this set of data is the ability to expand or link the patient journey with any other data from another database of interest. For example, the journeys could be linked to a separate database within the hospital that might only collect disease, drug or cost related information. The individual patients are not identifiable at any point.

The repository of data starts from the 1st of April 2003. New or latest data can be easily added to the dataset if deemed necessary for a particular analysis. The patient journey data will be clustered or grouped using appropriate parameters best for the particular analysis or knowledge discovery at hand.

4 Proposed Methodology – A Process Oriented Analysis Approach

4.1 Process Mining

Process mining in healthcare is still in its infancy. Mans, Schonenberg et al. (2008) used process mining techniques to better understand different clinical pathways taken by various groups of patients and used the technique to identify bottlenecks. Reubel and Ferreira (2011) concluded that although process mining techniques have been proven in some instances as being successful in mining health data, there are still room for improvements to identify the right algorithm to handle noise in the data, complexity of data and the ad hoc nature of health data.

Process is embedded in every aspect and at every level of one’s daily routines. The output of a process is a result that could be either favourable or not so favourable. Exploring the various activities within a process contributes to deeper knowledge, understanding and discoveries of the intricacies of what actually happens within a process that finally produces the result.

Health care industries are data rich as a result of embracing the notion of paperless system in a big scale. This has introduced some positive challenge to researchers in discovering knowledge from the use of such data. The notion of efficient patient care providing patient-centred approach has seen the emergence of various Health Information Systems as stated by Vezyridis, Timmons et al. (2011). Electronic Patient Management or Tracking Systems have all become not only common but essential systems for any hospital. These information systems store invaluable information that can be used for knowledge discoveries. Process mining enables the discovery of knowledge regarding a process. Process mining uses event or process logs to extract information regarding a process as it has taken place (van der Aalst and Weijters, 2004). These process / event logs do not have to necessarily originate from a Workflow Management System. A process log could be derived from a dataset that contains an order of events which could be used to assemble a process model that portrays the activity of the subject matter. This concept forms the basis of the methodology in this study.

van der Aalst, Reijers et al. (2007) stated that process mining aims to construct a process model from observed behaviour from a process perspective, organisational perspective or a case perspective; and the most significant output of process mining is the discovery of the main process flow. Rozinat, Wynn et al. (2009) stated that process mining normally creates a static model that could be used by the users of the systems to reflect on the process.

In the context of this study, the aim of process mining is to discover the various paths taken by inpatients moving from ED to other ward/wards. This study is also looking to discover the most frequent path taken and the associated parameters concerned with this path. Another aim is to gather the “network” information of doctors and ward where the information presented will indicate a doctor’s ward pattern. Braithberg (2007) argues that improving operational efficiency based on average bed occupancy is too weak to predict a complex hospital system and the dynamic nature of patient flow. Hence by using process mining the aim is to measure operation efficiency by calculating the throughput time for each ward, doctor and a cluster or group of patients. Another aim is to analyse inpatient journey or flow and to indentify the streams that often cause delays for patients to flow outside of ED.

4.2 Data Analysis

The aim of data analysis in this study is to analyse inpatient journeys to reveal macro characteristics of different cluster of patients. This study focuses on LOS of Outliers and Inliers. Outlier is an inpatient who is admitted to Wards other than the Home Ward. An Inlier is an inpatient who is admitted to a Home Ward. Home Ward is a ward that is equipped with appropriate medical team and specialised equipment to treat the patient’s primary disease at the time of admission.

Each patient journey is classified using Inliers, Outliers or Inliers/Outliers statuses. Inliers are patient journeys with 100% of the time at hospital spent in an Inlier ward. Outliers are patient journeys with 100% of the time at hospital spent in an Outlier ward. Inliers/Outliers consist of patient journeys where part of the hospital time is spent in an Inlier ward and the other part in an Outlier ward.

The outcome of this analysis will reveal if LOS is influenced by statuses. Process mining techniques will be used on this cluster of patient journeys for flow bottleneck identifications.

4.3 Collaboration with Clinicians

Regular contact with the clinicians enabled the portrayal of the actual undertakings at the hospital, which enhanced the quality of understanding as well as the relevance of the derived process mining and data analysis results. Clinicians’ insight was also invaluable in explaining the processes within the hospital’s context and the ways to interpret the result of the analysis and knowledge discovery.

The clinicians play a major role in identifying patient journey clusters or pool for a particular analysis and process mining that would give the best representation of activities taking place within a particular cluster of patients.
5 Preliminary Work

5.1 ProM (Process Mining) Toolkit

The Patient Journey Database is a well-maintained database containing historical records of the various movements of a patient during an admission episode. Information contained in this database is useful not only to construct a process model but also to discover hidden knowledge regarding a Patient’s Journey. The following section demonstrates some of the information revealed from undertaking process mining on this database using ProM.

Records from the Patient Journey Database were pre-processed and formatted into MXML format so the data could be read by the ProM toolkit. MXML is an extension of Extensible Markup Language (XML). The paragraphs below show some of the output of ProM upon processing the input data from the Patient Journey Database. For the purpose of ensuring that the pre-processing exercise is done accurately and the output of ProM is in accordance to what is aimed to be achieved, a small subset of data has been used for the ease of manual confirmation of the results. Five patient journeys have been included in this analysis.

Basic Log Statistics reveal statistical information from the data set. Each unique ward’s execution or processing information is represented. The statistical information that can be obtained is on minimum time a patient spent in the ward, maximum time a patient spent in a ward, arithmetic mean, standard deviation, geometric mean, sum and number of times the ward has been used.

Pattern Analysis output for the 5 patient journeys is shown in Figure 2. The pattern reveals 2 distinct flow or movement from ward to ward. It could be concluded from the output below, that ward “FMC” or ED, is frequently used. The aim is to derive common patterns used by a cluster of patients so this information could be used for better capacity planning.

Figure 2: Pattern Analysis – Patient Journey Flow

Individual Journeys could also be further analysed to depict the LOS (see Figure 4). This information combined with diagnostic information could be used to reveal commonalities in overall LOS and type of disease, which would contribute to better capacity planning.

Figure 3: Frequency of Ward Usage

Performance Sequence Analysis facilitates the assessment of performance of the flow of journey categorised as patterns and performance of each ward involved in the pattern. Mean throughput time for each block under a pattern could be discovered. This analysis will aid in discovery of process patterns that could potentially cause issues to the system. For example identifying wards with high throughput would enable investigation into the cause of such behaviour. Similar analysis could be carried out on doctors, which will show the transfer of work between 2 doctors, throughput time as well as the frequency of certain behaviour or pattern relating to a doctor. Figure 5 is a sequence diagram showing the movements of patient between wards for the 5 journeys included in this analysis. The labelled boxes on the top of the diagram are the names of “Wards” involved in these journeys. The numbered scale on the left is the time taken for the journeys to move from one “Ward” to another “Ward”. Measurement unit is in minutes. The coloured lines indicate 5 different journeys. It is also apparent from the diagram, that 2 journeys follow the same path where the movement is from ward “FMC” to ward “6D” to ward “ANG”. The time taken for each movement could be obtained by hovering mouse on top of the horizontal lines between “Wards”.

Figure 4: Journey Length of Stay
The sequence diagram is further processed into a Pattern Diagram. Journeys with same path or pattern are grouped together. For example, in this set of data, 2 journeys have moved from ward “FMC” to ward “6D” and then to ward “ANG”. These 2 journeys are grouped together forming 1 pattern and the rest of the journeys following 3 varied paths forming the other patterns. Figure 6 shows part of the Pattern Diagram for the five journeys. The pattern and analysis results shown are for the 2 journeys with the same path.

Further information for the Pattern shown in Figure 6 is provided in Figure 7. The mean throughput time for each pattern is derived. Measurement unit is in minutes. The mean throughput time for the patterns could also be obtained with mouse hover on the horizontal lines between “Wards”. The integers next to the arrows represent the number of journeys that have used that path. The decimal numbers next to the arrows represents the dependency relationship between the 2 wards involved. A decimal number close to 1 indicates a strong dependency relationship between the 2 wards. Strong dependency relationship shows the flow in that path is likely to happen.

Figure 8 depicts the patient flow derived from the 5 journeys. Each of the square boxes represents a Ward. The number below each Ward indicates the frequency of the Ward usage. The integers next to the arrows represent the frequency of occurrence for each pattern is also listed. Patterns with higher frequency indicate that such a movement pattern in patient journey is common and the behaviours of such a movement or patient journey could be further analysed. For example, this information could be used to investigate the correlation of patterns and bottlenecks.

Figure 8 shows the complete journey derived for the 5 patient’s journey, which reinforces the fact that theflow
discovery is complex and we need to classify and cluster the information properly before attempting to discover the flow pattern through process mining. The output of a flow pattern, when there are many more journeys involved, would become extremely complex and often meaningless without sound knowledge of the best way to derive a pool of data to be analysed. Therefore it is vital to invest in deriving an appropriate cluster of patient journeys and pool size which will ultimately reveal useful knowledge and information that could be used for decision making.

5.2 Inliers vs. Outliers LOS Analysis

Inliers and Outliers are one way to cluster patients. Clinicians have the perception that outlier patients end up staying longer in the hospital compared to Inliers, as they might be treated in a ward that is not the speciality ward for the patient’s condition. However, this remains a hypothesis. This analysis attempts to establish the correlation between LOS of patients and their Inlier/Outlier status. The first cluster of patients included in this analysis contains patients admitted to wards belonging to the division of General Medicine as these journeys portray the diversity of movement.

The challenge experienced so far has been to find the correct cluster of data or patient journey within the division of General Medicine that would appropriately represent the characteristics needed to give a holistic representation taking into consideration of the various factors that influence the status of a patient journey at a point in time.

The results of the very first analysis from the cluster selected were contrary to the hypothesis, i.e., the LOS for Outliers is not necessarily longer than Inliers. Upon consultation with the Clinicians and looking at the data closely, it was established that LOS study from grouping the journeys using status field alone was not an accurate representation to derive LOS for each status group. Another filter was applied to ensure that all the journeys considered were exclusively within the General Medicine division. This cluster of data has been confirmed by the Clinicians to be the best representation of the data for the analysis.

The next stage in this analysis is to separate the time spent at ward “FMC” or ED from the overall LOS. It has to be noted here, that these are inpatient records and in theory ward “FMC” or ED time should be zero however, this is not the case for most of the records, which indicates that many inpatients are spending time at ED when they should already be placed in a ward.

This brings another question to be clarified before LOS results could be finalised. “The question is whether time spent in ward “FMC” should be classified as Inliers time or Outliers time?” Following on from this clarification, LOS might be able to be calculated. Before this decision could be made, the cluster or the pool criteria has to be checked with the Clinicians.

Inpatient LOS has become one of the many ways used to measure performance of a hospital. Thomas, Guire et al. (1997) states that patient mean LOS has been used to measure quality of care and hospital efficiency in terms of resource usage. Thomas, Guire et al. (1997) further asserts that lower than normal LOS could indicate that hospitals are discharging patients early possibly sacrificing quality of care. Hospitals react differently to the continuous rising cost of health care. One way is to reduce the average inpatient LOS and unfortunately some hospitals reduce the number of beds in the hospital as a direct response to increasing cost of healthcare.

6 Discussion

The process undertaken has proven to be a viable approach in analysing the inpatient journey. The main challenge has been in defining the boundaries of the parameters needed and defining the parameters in accordance to the actual practice rather than analysing or undertaking process mining based on the field value only.

According to the Clinician’s view, the possibility of an Inlier staying longer is very viable as often when there are lack of beds the patients who are less sick are transferred to Outlier wards and patients who need more specialised care wait in ED before ending in an Inlier ward and consequently contributing to a longer Inlier LOS. The other scenario could be that a sicker patient staying longer at the hospital might eventually end up in an Inlier ward after being at various Outlier wards.

The collaboration with Clinicians has been an invaluable experience in this process. Hospitals are already undertaking various statistical data analysis for various reporting purposes to conform to KPIs and the approach taken in this study is to further break down the information and data to discover hidden knowledge.

7 Conclusion & Future Work

A smooth patient journey is an important aspect of a patient’s experience in the hospital. It could be as important as the actual treatment given to the patient, as a smooth journey will aid the healing process of a patient with physiological tranquillity.

After the finalisation of the pool or cluster of patient journeys as an output of the data analysis, process mining techniques will be used for knowledge discovery of the inpatient journey. LOS of Inliers versus Outliers will be analysed using the process-oriented approach outlined here to investigate if Outliers have or have not followed and optimum flow. Patient journey control flow will be analysed using various process mining algorithms that are available within ProM, which at the same time will lead to the identification of the more effective or appropriate algorithm to use for this kind of analysis.

Process mining results in conjunction with the usage of the proposed methods are perceived to offer added benefit to the already successful implementation of “lean thinking” and possibly enhance the improvement in areas where “lean thinking” approach alone is inadequate to reveal insight to access block.

8 References


