FLINDERS UNIVERSITY OF SOUTH AUSTRALIA

Hospital Patient Journey Modelling to Assess Quality of Care: An Evidence-Based, Agile Process-Oriented Framework for Health Intelligence

Lua Perimal-Lewis

School of Computer Science, Engineering and Mathematics,
Faculty of Science and Engineering

3 March 2014

A thesis presented to the
Flinders University of South Australia
in total fulfilment of the requirements for the degree of
Doctor of Philosophy

Table of Contents

List of Figures

List of Tables

Abstract

The thesis proposes a novel framework to gain Health Intelligence (HI) using an evidence-based, agile process-oriented approach to gain insight into the complex journey of patients admitted to hospital. This is the first systematic evidence-based research undertaking patient journey modelling spanning the entire hospital system using a process mining framework to complement statistical techniques. This is an innovative research of its kind looking at a large and complex cohort of General Medicine (GM) patients. This research investigated the impact of several system-based differences in models of care upon the Quality of Care (QoC) that can be delivered to inpatients in any hospital in Australia. For example team-based and ward-based models of care were compared using real patient data at Flinders Medical Centre (FMC). Hospital outcomes for patients who were admitted to the "wrong" ward (ward outliers) were compared with patients who were admitted as ward inliers.

Because time spent in the Emergency Department (ED) impacts the overall patient journey, the research also compartmentalised the time patients spent in the ED in order to investigate the influence of these separate time compartments upon QoC and further comparison was made depending on whether the patient was admitted inside or outside working hours.

Having demonstrated the complexities of patient journeys using real hospital data on a complex cohort of patients, the research demonstrates and advocates the use of process mining techniques to automate the discovery of process models for simulation projects. This approach avoids those errors that are more likely when applying hand-made process models in a complex hospital setting.

Process mining is an emerging technology that aims to gain insight into a process. This research applied the process mining framework to analyse clinical processes. Although the application of process mining in the healthcare setting is still in its infancy, the concepts surrounding the framework of process mining are sound. The fundamental elements needed for process mining are historical event logs. Process mining generally relies on event logs generated by Process Aware Information System (PAIS). This research establishes a formal framework for deriving an event log in a healthcare setting in the absence of a PAIS. A good event log is a cornerstone of process mining.

This framework will be generalizable to all public hospital settings because it uses the already-collected hospital Key Performance Indicators (KPIs) for data extraction; building on patient journey data to derive the event log which is then used for various analyses thus providing insight into the underlying processes.

The strength of this work derives from the close collaboration with the practising clinicians at the hospital. This close partnership gives clinical relevance to this research and is the main reason the research is breaking new grounds in improving evidence-based clinical practices to provide patient-centred care. Modelling cannot depict everything in a complex environment such as the healthcare system but a systematic and innovative approach to modelling would depict the main behaviour of the system which will consequently lead to knowledge discovery and health intelligence.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text. There is also no conflict of interest with Flinders Medical Centre (FMC) where the empirical research was undertaken testing the applicability of the framework.

Lua Perimal-Lewis

Date: 3rd March 2014

Acknowledgement

This thesis is dedicated to:

My family and my supervisor, Professor Campbell Henry Thompson

Thank you for your selflessness.

~~~~~~~~~~~~~~~~~~~~~~~~~

Mr Colin Lewis, my husband: "*Thank you for your unconditional love, support and prayers*". Miss Namita Lewis and Miss Samika Lewis, our daughters: "*You are the light of my world. I am sorry for the time away from you*". Mrs Gunalechumi Gunasegaran, my mother; Mr Perimal Gengappan, my father and Dr Enoch Kumar Perimal, my brother: "*You are my pillars of strength. Your unconditional love and prayers helped me through. Thank you for the encouragement"*. Dr Hemabarathy Bharatham, my sister-in-law and Mr Suhail Vihen, my nephew: "*Thank you for sharing your beloved with me*".

My dear friends, Mrs Martha Bhaskaran, Mrs Sarih Raizi and Mrs Haleh Lady: *"Thank you for being there for us and for our children whenever we needed you"*.

Dr Denise de Vries, my supervisor: "*Thank you for your support and encouragement*". Professor Campbell Henry Thompson, my supervisor: "*Thank you for your guidance, encouragement and support. I cherish your integrity. Your actions speak louder than words*".

Mr Paul H Hakendorf: "*Thank you. You were always ready to help with a smile*".

"*Thank you to all the co-authors and colleagues*" (*in alphabetical order*): Professor David Ben-Tovim, Associate Professor Paul Calder, Dr Susan Kim, Dr Jordan Y Li, Ms Rui Li, Dr Shaowen Qin, Mr Mark Reilly, Ms Susan Roberts, Dr Shahid Ullah and Associate Professor Richard Woodman.

~~~~~~~~~~~~

Regard man as a mine rich in gems of inestimable value. Education can, alone, cause it to reveal its treasures, and enable mankind to benefit therefrom.

Bahá'í writings

1 Introduction

Information Technology (IT) has become an integral part of modern lifestyle. The advancement in IT especially developments in the Information, Communication & Technology (ICT) field has enabled instant access to information. The emergence of Clinical Information Systems (CIS) and Hospital Information Systems (HIS) expedites access to patients' health records. According to Hilberts and Gray (2014), ehealth integrates the use of electronic information and communication technologies. The Australian Government has introduced a Personally Controlled Electronic Health Record (PCEHR) to facilitate the sharing of patients' health records electronically if agreed by the patient (Government of South Australia, 2014). The ehealth vision for Australia is based on how Australians across geographic and jurisdictional boundaries interact with the health system to manage their own health (Hilberts and Gray, 2014). The Health Department within each state in Australia is continually looking into ways to consolidate patients' electronic health records for efficient information sharing. One such example is the state-wide implementation of the Enterprise Patient Administration System (EPAS) in South Australia (SA). The EPAS will be rolled out across the SA Health network. The goal of the EPAS project is to provide comprehensive patient information electronically to healthcare professionals and administrative clinical staff at the point of care (Government of South Australia, 2014). To facilitate nation-wide health information sharing, the EPAS will be eventually extended to integrate with the PCEHR (Government of South Australia, 2014).

The union of IT and health has already seen major paradigm shifts in the delivery of healthcare. One such example is the shift from paper-based medical records to Electronic Medical Records (EMRs) or Electronic Health Records (EHRs). The previous paragraph discussed some examples. Although the amalgamation of IT and health has seen the fruition of countless benefits for both the healthcare provider and the patient, there are still many challenges to be addressed. The National Academy of Engineering (2012) has identified the '*advancement of health informatics*' as one of the grand challenges for Engineering in the 21st century. Swiftly adapting the innovative use of emerging state-of-the-art IT research in a timely manner in health will facilitate the advancement in health informatics. The application of process mining to healthcare, as addressed in this research, is one such advancement within the field of health informatics.

One fundamental challenge that remains to be overcome in order to advance health informatics is the lack of a framework to embrace the diversity inherent in a complex hospital structure. Diverse services, units, departments, doctors and allied health professionals should be engaged when treating a patient. These diverse groups would naturally have different goals and requirements and would also have varied IT needs. For example the CIS used in a General Medicine (GM) unit would be different to the CIS used in a Cardiology unit. The challenge therefore is to find innovative ways to enable patient information sharing between these diverse groups involved in a patient's treatment. The information sharing should first happen within the hospital before initiating inter-hospital information sharing. The lack of sharing of patient information between hospitals is widely recognised. In a recent review, Miller and Tucker (2014) argued that, although the IT platform for data sharing exists, the propensity for sharing electronic patient records between hospitals is lower than the tendency for sharing this information internally. The reason for this lack of sharing of inter-hospital patient information was put down to the fear of losing money as a consequence of losing patients to another health institution (Miller and Tucker, 2014).

Patient information sharing within the hospital is an important factor for any modelling of the patient journey hospital-wide. In recognising the diversity of the various groups involved in the treatment of a patient, the framework proposed and demonstrated in this research emphasises that patient information sharing within the hospital is independent of any one IT platform. To undertake modelling of the hospital-wide patient journey, it is essential to determine common variables that uniquely identify a patient and the patient's admission to the hospital. Whilst in the hospital, the patient may move from unit to unit or from ward to ward. Therefore another variable is needed to track this movement. In many hospitals, the tracking of a patient's movement is facilitated by patient management software. Regardless of whether a patient management system exists or not, establishing the minimum data requirement framework that enables tracking of patients through the hospital is essential for hospital-wide patient journey modelling. This minimum data requirement should be easily collated. The ease of data collection is important as is the choice of those variables that can be easily shared. For example, the patient identification number and the date of service delivery are variables that could be easily shared. However specific details concerning treatment of a patient are information that would not be easily shared. The clinical practice relevant to each unit is not of great interest, rather only the minimum data requirement about each patient. Establishing a minimum data requirement framework which could be used for modelling of the hospital-wide patient journey encourages patient information sharing. In this research, hospital-wide patient journey modelling was undertaken to assess Quality of Care (QoC). The minimum data requirement framework was extended to enable suitable collection of QoC data. Therefore a novel framework was proposed to undertake modelling of the hospital-wide patient journey while still accessing data relevant to QoC.

Patient journey modelling is normally undertaken to gain insight into certain patient pathways. The insight gained from a modelling activity could be used as a starting point for process improvement. These process improvement activities are undertaken to improve the delivery of patient-centred care. The aim of this research was to undertake patient journey modelling that would facilitate process improvement activities. The insight gained from undertaking hospital-wide patient journey modelling would enable hospitals to gain Health Intelligence (HI) by looking beyond aggregate statistical modelling techniques. This research therefore proposed and demonstrated the applicability of a novel framework for hospital-wide patient journey modelling. The modelling of the hospital-wide patient journey helped clinicians gain health intelligence into complex patient journeys at Flinders Medical Centre (FMC). The modelling used an evidence-based agile process-oriented approach by developing an innovative way of applying process mining techniques to healthcare.

Process mining was defined by its pioneer van der Aalst (2011) as a new collection of Business Intelligence (BI) techniques under the all-encompassing principle of Business Process Management (BPM). BPM aims to improve business processes by amalgamating IT and management science. As such, applying process mining to healthcare will contribute to an in-depth analytical knowledge that, in turn, will contribute to HI. Unlike many mainstream BI and data mining tools which are data-centric, process mining is process-centric. Process mining aims to gain insight into those processes referred to by the data. The focus is not on fancy-looking dashboards rather it is on a deeper analysis of the data (van der Aalst, 2011).

Processes in hospitals are complex. The goal of process mining is to gain insight into a process by carrying out detailed analysis using historical event data pertaining to that process. Retrospective studies are often criticised because of the possibility of introducing errors due to study bias and confounding variables. However, the invaluable insight and knowledge that could be gained from analysing historical health data should not be dismissed. On one hand, HI is achieved from undertaking modelling of the hospital-wide patient journey using two complementary techniques; process mining techniques and statistical techniques. On the other hand, HI is achieved from the insight registered by the practising clinicians and managers who interpret the results.

As previously mentioned, process mining uses historical event data or event logs for analysis. Generally, these event logs are easily sourced from Process Aware Information Systems (PAISs). This research proposed an innovative framework in which to undertake process mining in healthcare in the absence of PAIS by using data stored in hospital databases. The healthcare sector is rich in data. The vast data repository in healthcare and the complexity of the health data contained therein pose positive challenges to the research community. Simpao et al. (2014) recently reported, the imminent need for the application of sophisticated analytics methods and tools in order to improve patient care, increase efficiency, optimise resources and enhance decision making. The data extraction and the data processing challenges embedded within health data combined with the complexity of health processes demand a close working relationship with clinicians and managers to provide a meaningful interpretation of any knowledge discovered.

Working in close collaboration with clinicians was a key success factor in this collaborative research. In this research, the collaboration with Flinders Medical Centre (FMC) and its practising clinicians shaped the scope of the research. One goal of this research was to apply one recent advancement of IT, namely the application of the process mining framework, to the healthcare domain in an effort to improve the delivery of patient-centred care in a complex healthcare environment.

'*Real*' collaboration is the foundation for long term sustainability underpinned by the '*law of reciprocity*' which lies at the heart of the research. Therefore another goal of this research was to be open to explore matters that would add value to the hospital and its practising clinicians. This in essence demanded an agile attitude from collaborators breaking away from the need to conform to preconceived ideas surrounding the need to apply methodologies with rigidity.

The application of process mining in healthcare is still in its infancy; however the concepts surrounding the framework of process mining are sound. Mans et al. (2008) applied process mining techniques to a stroke care unit in order to better understand the differing clinical pathways taken by a diverse group of patients. Mans et al. (2008) used these techniques to identify bottlenecks. Rebuge and Ferreira (2012) concluded that, although process mining techniques had been proven in some instances to be successful in mining health data, there was still room for improvement when identifying the right algorithm to handle noise in the data, complexity of data and the ad hoc nature of health data. The complex nature of healthcare data and the varied processes within this environment make the use of process mining techniques a viable method to gain insights into these processes (Perimal-Lewis et al., 2012b).

In applying process mining to healthcare, this research undertook a two-fold approach; one using process mining as a framework and the other using process mining as a HI tool. In adapting process mining as a framework for the modelling of the hospital-wide patient journey, three perspectives were explored: the case perspective, the organisational perspective and the control flow perspective. In adapting process mining as a tool, the properties of event logs were explored and an innovative way was used to produce the derived event log. This event log conformed to the properties of the actual event logs of PAIS. Deriving event logs in a complex healthcare environment is far from a trivial exercise. The derived event log enabled process mining analysis to take advantage of ProM toolkit, an open source process mining tool (The Process Mining Group, 2010).

Mans et al. (2013) recently identified 12 scholarly research papers using process mining in healthcare. The published work from Chapter 3 as described by Perimal-Lewis et al. (2012b) was identified as one of the 12 research papers.

Evidence-based research in healthcare is predominantly performed using statistics to improve clinical practice. Evidence-based research is also used as a basis for healthcare process improvement. The output of rigorous research is used as evidence to initiate improvement in clinical practice and clinical process improvement. Sackett et al. (2000) defines evidence-based medicine as *'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients based on best available external clinical evidence from systematic research'.* A clinical epidemiology unit is usually based at most large hospitals. That unit has experienced statisticians producing advanced statistical analysis and modelling for routine hospital reporting. Hospitals are venturing into multi-disciplinary collaborative research, looking for innovative ways to enhance knowledge discovery contributing to the use of techniques from various fields of research. The research presented in this thesis is one such research using the concepts of process mining to aid evidence-based research in healthcare. The derived event log for process mining is sourced around well-established KPIs. This strategy enables process improvements pertaining to these KPIs as a result of gaining deeper understanding of the underlying processes.

The following sections are included to set the context of the application of the framework at FMC:

1.1 Flinders Medical Centre (FMC)

FMC is a public teaching hospital with around-the-clock Emergency Department (ED). It is one of the two major trauma centres in South Australia (SA). Flinders Medical Centre (2013) reported FMC is equipped with approximately 593 hospital beds. It employs more than 3,500 skilled staff. FMC is publicly funded. It has 11 divisions. Each division is managed by its own executive and senior staff. Each management team is responsible for their division including managing their own budget (Flinders Medical Centre, 2013).

1.2 FMC’s Emergency Department (ED)

As reported in 2013, FMC's ED attends to about 74,000 patients per annum. Of these, 43% are admitted to a hospital and 38% admitted as inpatient to FMC. It is one of the busiest EDs in SA. FMC's ED also has a separate paediatric area. Adult patients presenting to FMC's ED are streamed into two streams: one stream is for those patients who are likely to be admitted and the other stream is for patients who are likely to be discharged (Flinders Medical Centre, 2012).

1.3 General Medicine (GM)

The GM service at FMC is the largest service at the hospital. Its patients are those requiring emergency hospital care. These patients have complex multi system disease. The GM units were chosen for this study because they are responsible for the care of the largest inpatient population. The likelihood of identifying outcomes of both statistical and clinical significance is higher when using a large population. On the other hand, the GM population is a heterogeneous population. Therefore careful consideration is required to address this challenge when undertaking these studies.

1.4 Inlier and outliers

It is a well-established perception among clinicians that ward outliers receive inferior QoC compared to ward inliers. Ward inliers are those patients who are admitted to the home ward of the team responsible for the treatment of that patient. On the other hand ward outliers are those patients who are admitted to non-home wards. Clinicians advocate moving the right patient to the right bed believing this to be in the best interests of the patient (Alameda and Suárez 2009, Santamaria et al. 2014).

The literature reviewed thus far has not been able to establish any significant research in this area. Understanding the consequence of being an inlier and an outlier will give insight to the hospital bed managers. When an admitted patient’s exit is delayed from a congested ED, those bed managers are responsible for finding the most appropriate bed for that patient.

This is the first study of its kind that studied a large cohort of GM patients in order to investigate the impact of ward location of a patient on overall in-hospital Length of Stay (LOS).

1.5 Quality of Care attributes (QoC)

QoC attributes were established based on the historical event logs. These attributes were chosen because they are measurable attributes and the information is already collected by the hospitals for KPI reporting. This process of deriving the QoC attributes that not only surround a process but are associated with KPIs is a generalizable approach which can be adapted by all hospitals in Australia. In this particular instance it was useful to use this strategy to examine the impact of ward allocation on QoC for patients and hence examine the ward allocation process.

The Quality of Care (QoC) attributes used in this research is as per Figure 1-1.

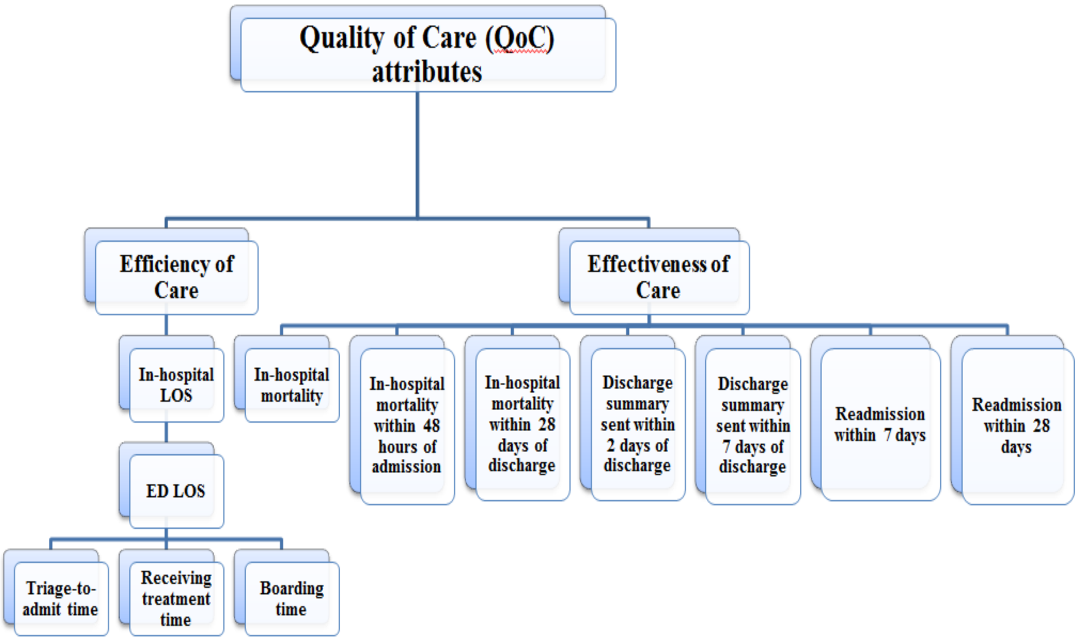


Figure 1-1: Quality of Care (QoC) attributes

1.6 Brief outline of the chapters covered in this thesis

Chapter 3 discusses the process-oriented methodology. It then focusses on the importance of collaboration with domain experts as a key success factor in any process-centric methodology. It proposes a process-oriented analysis approach using process mining techniques in order to gain a better understanding of the inpatient journeys at FMC. Currently, ProM is the only available open access process mining tool (The Process Mining Group, 2010); therefore ProM was used to set the context of the process-oriented approach. ProM was used to demonstrate to the domain experts the potential inputs and the outputs of a process mining methodology. The domain experts had the opportunity to comment and reflect on the applicability of the potential outputs of the process mining. The advantage of using data from a true hospital setting rather than a sample dataset revealed quickly that the success of the process mining will largely depend on drawing clear boundaries and limiting scope to the process of interest. Data analysis was used to define the boundaries or scope of the ward inlier and ward outlier study working in close collaboration with the domain experts. This activity falls under the case perspective of process mining. Patient LOS is one of the KPIs used by hospitals to measure hospital efficiency. The study was interested in testing the hypothesis of ward outliers having a longer LOS than ward inliers. The process-oriented methodology proposed in this chapter was published as one of the 12 scholarly research to adapt an innovative process mining methodology in healthcare (Perimal-Lewis et al., 2012b).

Chapter 4 addresses the inherent challenges in health data. Health data are often classified as structured or semi-structured. It then highlights the properties of health data that are un-structured or semi-structured. The chapter then discusses the properties of event log and highlights the usefulness of these properties for deeper analysis of a process. Such deeper analysis will add insight into a process, therefore gaining intelligence from the knowledge discovery activities. After presenting the benefits of event logs, Chapter 4 then proposes a methodology to derive an event log in the absence of automated event logs. The derived event log forms the data source used for process mining. The derived event log was used for process-oriented analysis. This chapter also discusses the ethical challenges that have to be taken into consideration when working with patient data.

Chapter 5 discusses the approach taken to define ward inliers as opposed to ward outliers. It discusses the concept of ward-based care and the concept of team-based care and how this relates to ward inliers and ward outliers. The hospital process of bed allocation, where hospital patients admitted via the ED are either placed in their home wards or placed in a non-home ward are discussed. Using the historical event log derived in Chapter 4 and with the notion of gaining insight about the process of streaming patients to either a home ward or to non-home ward locations, a novel framework for deriving inlier and outlier populations was established. Although the historical event log came with information about whether the ward a patient occupied was an inlier or an outlier location, there was no established framework to identify this population with conviction. As a result of the process-oriented approach, it was established that there are three distinct populations (status) that could be separated based on their inpatient location of care: 100% inliers, 100% outliers and those who spend their inpatient time partly as an inlier and partly as an outlier. The third population, those who spent their time partly as inliers and partly as outliers had to be further analysed to determine the best definition of inliers and outliers within this group. This chapter further identified and excluded the confounding variables that should be taken into consideration when measuring QoC outcomes. This exclusion process and the reasoning behind the exclusion process are presented together with an evidence-based approach to support the exclusion. To further limit the variation in characteristics of patients in each group, the domain experts identified and grouped patients according to their primary diagnosis. The process-oriented analysis steps taken to examine this ward allocation process, enabled publication of this very first study of its kind on the hospital outcome of ward inliers and outliers (Perimal-Lewis et al., 2012a).

Chapter 6 discusses the heterogeneity inherent in patient journeys. Having established a framework to identify the inlier and outlier population and identifying the attributes to measure QoC based on the underlying processes around KPI reporting, it was then necessary to address the heterogeneity observed in inpatient journeys. As mentioned earlier, GM patients are a diverse and complex group of patients therefore it is vital to ensure that outcome measures (QoC attributes) are assessed on groups of patients with similar characteristics. By adopting the case perspective concept of the process mining methodology to gain insight into inpatient journeys, these patients journeys were clustered into homogenous groups. Clustering techniques are widely used to group the subjects of studies into homogenous groups. Two-step cluster analysis was used to discover from the natural structure of the data, groups of homogenous patients. Based on this grouping, their patient characteristics and their QoC attributes were assessed. Here the QoC attributes were the primary focus in order to establish the homogenous clusters of patients based on patient characteristics. The next important step investigated whether there were any significant differences between the QoC attributes and patient characteristics within the discovered clusters of homogenous inpatient journeys based on their inlier or outlier status defined earlier. This chapter also discusses the finding of cluster-based analysis in comparison with non-clustering approach. It also highlights the importance of addressing the issues of heterogeneity in the data set. This chapter gives an evidence-based foundation from which to advocate the use of complementary techniques (in this case process mining approach with stochastic approach) to take advantage of the strength of each technique in producing robust results. As a result of the work done here, the paper presenting the outcome of '*analysing homogenous patient journeys to assess QoC for patients admitted outside of their home ward*' was published (Perimal-Lewis, 2013).

Chapter 7 discusses the organisational perspective of the process mining methodology. Having established the importance of addressing the heterogeneity in patient journeys and the impact this heterogeneity has on the outcomes of the measured QoC attributes, it was imperative to undertake further analysis to understand the complex processes in the ED. Modelling a complex ED process without looking into the finer aspects of the relevant processes risks the missing of invaluable insights. Most research has established the impact of long boarding time (overall ED time) on the overall inpatient LOS. A contemporaneous study compartmentalised ED time into two categories. To gain better insight into the interactions of the complex ED processes and their effects on the various outcome measures, this study compartmentalised ED time into a triage-to-admit time and a boarding time. It assessed the outcome of these times as separate measures. To align with the normal processes of how hospitals and their EDs are staffed, the study further investigated the effect of presentation at triage inside and outside working hours. It also looked at the effect of the timing of the admission decision and whether it was taken inside or outside of working hours. These attributes were studied together with the number of patients in the ED at the time of triage and when boarding commenced as this would naturally impact on how quickly the patient would be attended by the medical staff. This study further derived a prediction model for both the ED times as a linear increment of number of patients in the ED during the triage-to-admit period and boarding time. The outcome of the work undertaken in this chapter was been recently published (Perimal-Lewis et al., 2014a).

Chapter 8 discusses the control flow perspective of the process mining methodology. Having established a framework to define the inlier and outlier population for a complex cohort of GM patients, having identified the QoC attributes, having established the need to address heterogeneity in the data and having looked further into the ED processes; process mining were undertaken for the discovery of control flow of the inpatient journey from start-to-end. Hospitals traditionally undertake simulation to better understand the complex nature of the interactions between various hospital variables. These models for simulation are generally derived manually. Having presented the extensive analysis undertaken so far to gain insight into the processes of inpatient journey, it is then imperative to undertake a simulation project which can model the various interactions that are taking place within the inpatient journey from start-to-end. As a starting point, processes within the ED were compartmentalised into three compartments according to recognised and established ED KPIs. This chapter discusses the complexity of the discovered model and the iterative process needed to derive the final control flow model that depicts the patient flow from start-to-end. Having demonstrated the complexity of control flow discovery, process mining was used to automate the discovery of process models for the use of simulation. This strategy is strongly advocated to avoid errors prone to occur in hand-made models. As a result of the work done here, the paper presenting the outcome of '*discovering the process model using process* *mining by constructing Start-to-End patient journeys*' was published (Perimal-Lewis et al., 2014b).

Chapter 9 is a summative chapter. It outlines the novel achievements of this thesis with a section on future work.

2 Literature Review

This literature review was documented in the spirit of investigating the various problems faced by the Australian public hospitals. The review also investigated the process improvement initiatives undertaken by the hospitals to improve the delivery of patient-centred care. This review focussed mainly on investigating those public hospitals with Emergency Departments (EDs).

2.1 Introduction

Australian public hospitals are government funded institutions servicing the healthcare needs of a community within a locality. These hospitals are continually brought into public attention for issues like poor management, lack of timely care, over-crowding or access block, long waiting times at ED, lack of beds and shortage of staff to name a few. Despite all these issues, many initiatives to improve patient-centred care are undertaken by hospitals and government health departments to ensure the welfare of the community.

Each hospital’s core operation is centred on their patients. Increasingly, policies are introduced that promote patient-centred care. Gaining insight into how patients flow through the hospital system is vital in any effort to improve the way hospitals function. The insight gained from such activities directly contributes to formulating strategies to address the challenges facing hospitals.

The issues discussed in this thesis can occur in hospitals world-wide but relate most closely to processes of care within hospitals in Australia, New Zealand, the United Kingdom, Canada and the United States.

2.2 Public hospitals in Australia

Public hospitals provide a range of services from immediate trauma management, specialist care, rehabilitation and other supporting services such as antenatal classes, education on nutrition and services for aboriginal health. Public hospitals rely on government funding that is affected by the volatile political condition of the country. Schafermeyer and Asplin (2003) stated that the United States of America has relied predominantly on market forces to fund their healthcare services. On a similar note, the Australian Government's national health reform plan released on 3rd March 2010 brings significant changes to the governance and public health financing in Australia (Bennett, 2010).

The graph in the Figure 2-1 shows the number of admissions to public hospitals pre 1999 and yearly representation of admissions from 1999. As depicted in the graph, the number of admissions has a steady increase year by year.

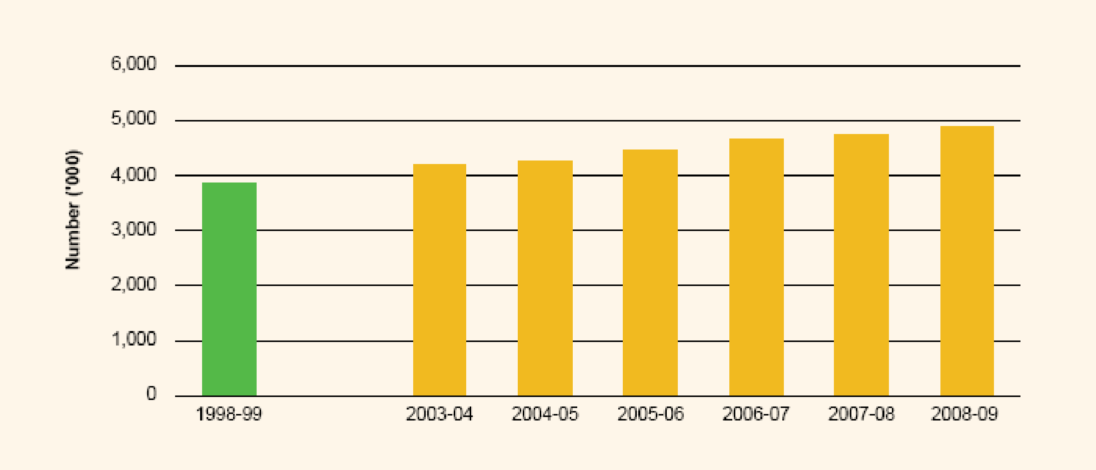


Figure 2-1: Number of admissions in public hospitals, 1998-99, and 2003-04 to 2008-09, (Australian Government Department of Health and Ageing 2010, pg. 16)

An increasing number of admissions to hospitals can be expected as the population of Australia increases.

The demand on Australian public hospitals will also increase as the life expectancy of Australians increases. This is especially apparent in those who are 65 years old or older. The trend in South Australia (SA) as reported by Banham et al. (2011) from a study undertaken from 1999 to 2008 showed that both total life and healthy life expectancy increased (2.0 years among males; 1.5 years among females) and (1.4 years among males; 1.2 years among females) respectively. According to the Australian Bureau of Statistics (2013) in 2056 there will be an increase in the number of people aged 65 years or more with this proportion of the population increasing to between 23% and 25%.

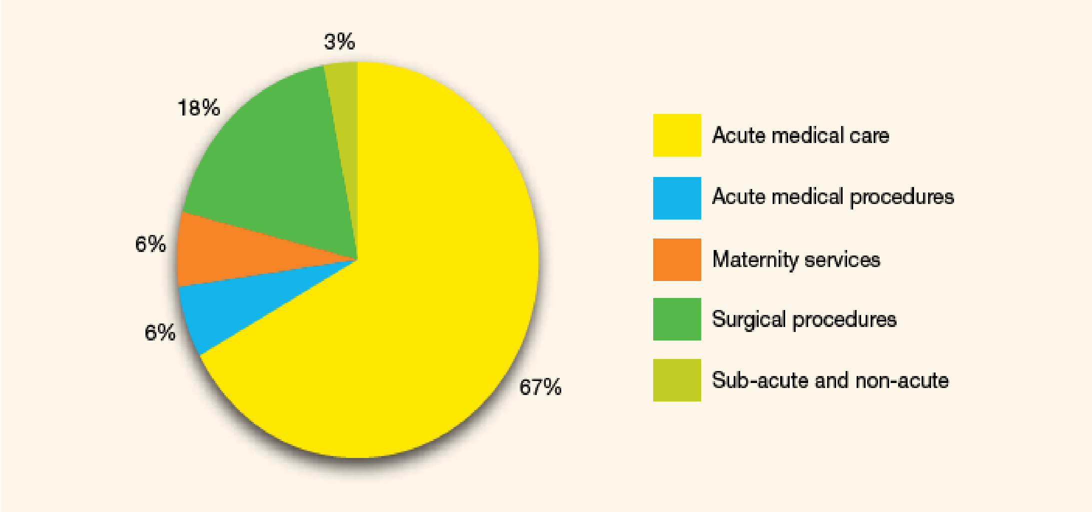


Figure 2-2: Percentage distribution of admissions by service type, public hospitals, 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 53)

The graph in Figure 2-2 shows the various types of admissions to Australian public hospitals. More than 50% of patients admitted required Acute Medical Care. Acute Medical Care is provided for patients with a severe medical condition which can be usually managed with drugs whereas Acute Medical Procedure patients receive their care usually with the involvement of specialised equipment. Patients admitted for Acute Medical Procedure generally complete their treatment in less than a day (Australian Government Department of Health and Ageing, 2010).

2.3 Emergency Departments (EDs)

In Australia, most public hospitals have an Emergency Department (ED). An ED provides emergency / urgent care for patients without a prior appointment. Often the ED is the entry point for patients to access urgent intervention by a medical professional. The ED Team can be challenged by patients presenting with all types of conditions. In a public hospital such as the FMC, the ED does not shut down, and is staffed to attend to patients 24 hours per day and seven days per week. The main focus of this review was on investigating the crisis in the Australian ED as well as some other commonly reported hospital crises. The goal was to gain insight into how the hospital operational issues impacts on the overall functioning and effectiveness of the hospital.

The graph in the Figure 2-3 shows the number of presentations to EDs in public hospitals in Australia. The graph clearly shows that ED presentations are increasing yearly. The Australian Government Department of Health and Ageing (2010) reported that there were

5.0 million ED presentations in 1998 and ED presentations increased to 7.2 million between 2003-04 and 2008-09.

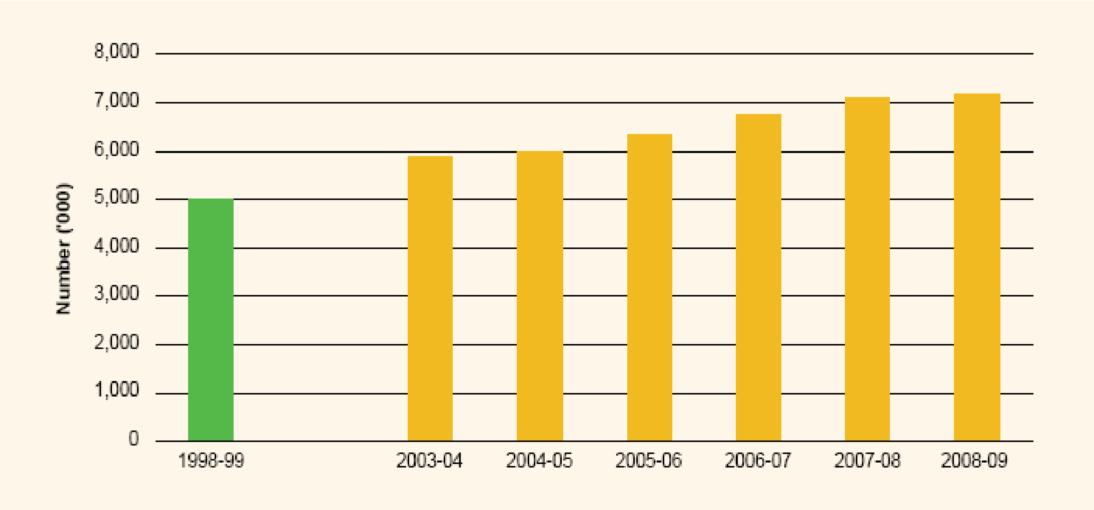


Figure 2-3: Number of emergency department presentations, public hospitals, 1998-99, and 2003-04 to 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 23)

The Australian Institute of Health and Welfare (2011-2012) reported that there was an increase in annual ED presentations of 4.3% on average each year between 2007-08 and 2011-12 amounting to over 6.5 million presentations per annum. Data from both sources confirms that ED presentations are on the rise. The recent data reported indicates that ED presentation had increased substantially more between 2007-08 and 2011-12.

It is evident from the data presented that the demand for services within EDs will continue to increase and hospitals should explore ways to improve processes and enable better delivery of patient-centred care. The issues discussed in the following paragraphs relate to those problems facing EDs that contribute to inefficient functioning of the associated hospitals. A poorly functioning ED impacts upon patient-centred service delivery.

2.3.1 Access block / ED overcrowding

Fatovich (2002) described ED overcrowding as an international problem resulting in the inability to provide timely emergency care. ED overcrowding is a major problem facing many Australian public hospitals. Australia and North America are at the forefront of research relating to this problem as reported by Higginson (2012) and Hwang et al. (2011). When the ED is overcrowded, ambulances are instructed to divert to another facility because of a lack of capacity to safely attend to a newly-arrived patient at the original facility (Delgado et al., 2013). Fatovich (2002) further stated that access block is when a patient in the ED is not able to access on-going inpatient care because of the lack of an inpatient bed and as a result these patients occupy and crowd the ED. This is termed “boarding” in the ED.

Data from a New South Wales tertiary hospital showed that more than 50% of patients receiving treatment in their ED were waiting for beds and 75% of these patients had been in the ED for more than 8 hours (Richardson and Mountain, 2009). In Australia, despite the variation between states and hospital types, on average there are more patients receiving treatment in the ED causing access block than there are patients waiting to be seen. The Australasian College for Emergency Medicine (2011) defined the ED to be “overcrowded” when the number of patients waiting to be seen, the number of patients being treated and assessed and the number of patients waiting to leave the ED exceed either the bed capacity or staffing capacity of that ED. Usually there will be more than enough resources in the ED to treat those patients waiting to be seen if the admitted patients awaiting an inpatient bed are swiftly moved to those beds (Richardson and Mountain, 2009).

Derlet (2002) stated that a large number of seriously ill patients presenting to the ED with multiple medical problems combined with a decreased turnaround time for beds both contribute to ED overcrowding. Patients with comorbidity require a variety of treatments involving more than one speciality and department. Derlet (2002) further stated that a prolonged treatment strategy within the ED designed to avoid hospitalisation coupled with an inability to access on-call specialists have also contributed to ED overcrowding.

Other factors that cause access block include illness associated with the aging community, a higher expectation of QoC by the community, a reduction in residential care options in the community, a significant reduction in local after-hours General Practice (GP) services, a reduction in the after-hours house calls and finally, Casemix payments which encourage hospitals to perform simple elective surgeries (Cameron et al., 2009).

Richardson and Mountain (2009) identified more factors that can cause ED overcrowding such as seasonal influences (*e.g. during winter / flu season*). EDs are more congested on Mondays because of an influx of admissions for elective surgery and a reduction in weekend discharges.

Richardson and Mountain (2009) also argued that the belief that ED overcrowding was caused by patients who do not require urgent care (*e.g. conditions which could have been treated by a GP*) was a myth. Richardson and Mountain (2009) further warned that, if this myth is given credence, attention will be diverted from finding solutions for the real major ED crisis; the large amount of patients being admitted to hospital.

As ED overcrowding intensifies so do the varied strategies proposed to ease ED overcrowding (Paul and Lin, 2012). Despite a broad range of research in this area, there are still a growing consensus for quantitative and feasible measures of ED crowding that can be adapted for use at multiple sites (Hwang et al., 2011). Lucas et al. (2009) stated that future strategies to reduce ED overcrowding should be aimed at streamlining the admission process of patients from the ED, increasing inpatient capacity and increase the capacity of higher acuity nursing units to accept patients.

In conclusion, access block in the ED is not just an ED problem. Therefore looking at the patient flow through the hospital in order to understand bottlenecks would give better insight to decision makers for developing strategies to relieve ED congestion.

2.3.2 Presentation and waiting times in ED

Patients are triaged upon presentation to the ED. There are five Australasian Triage Scale (ATS) categories and each category has a recommended maximum time within which assessment and treatment must commence. Patients are triaged according on how urgent they need treatment (Australian Government Department of Health and Ageing, 2010).

Patients within ATS 1 category require immediate resuscitation, ATS 2 category patients require emergency treatment within 10 minutes, ATS 3 category patients require urgent treatment within 30 minutes, ATS 4 category patients require semi-urgent treatment within 60 minutes and ATS 5 category patients require non-urgent treatment within 120 minutes.

The Australian Government Department of Health and Ageing (2010) reported an increase in the proportion of patients in the emergency and urgent categories and a decrease in the semi-urgent and the non-urgent categories; as per the graph in Figure 2-4. The proportion in each category only changed by about one per cent from 2003-04 to 2008-09.

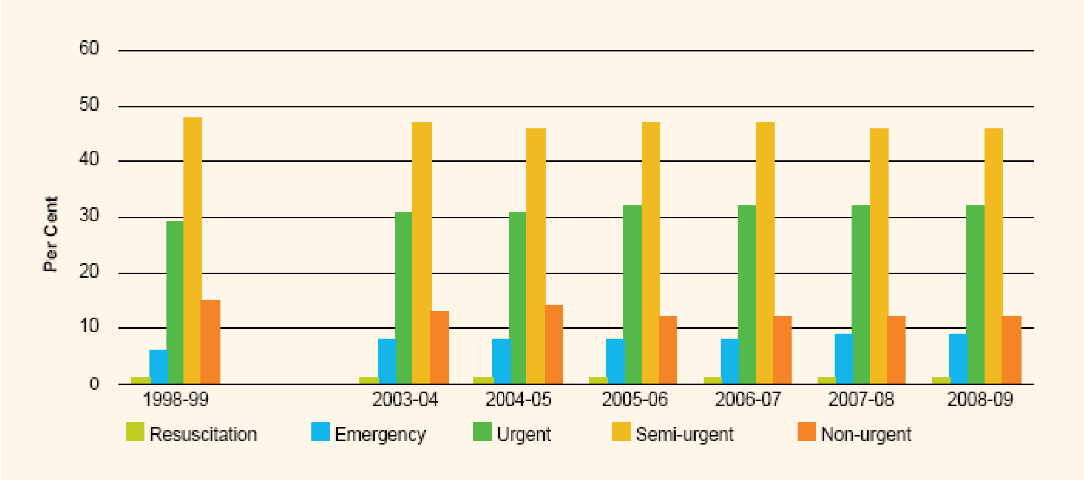


Figure 2-4: Proportion of emergency department presentations, by triage category, public hospitals, 1998-99, and 2003-04 to 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 25)

The graph in Figure 2-5 shows the percentage of patients treated within the recommended triage categories. The proportion of ATS 1 category patients who required immediate resuscitation and were treated within the recommended time has always been very close to 100%. Treatment time for all other triage categories has reduced from 1998-99 to 2003-04 with one percent point improvement in 2008-09 compared with the previous year; 2007-08.

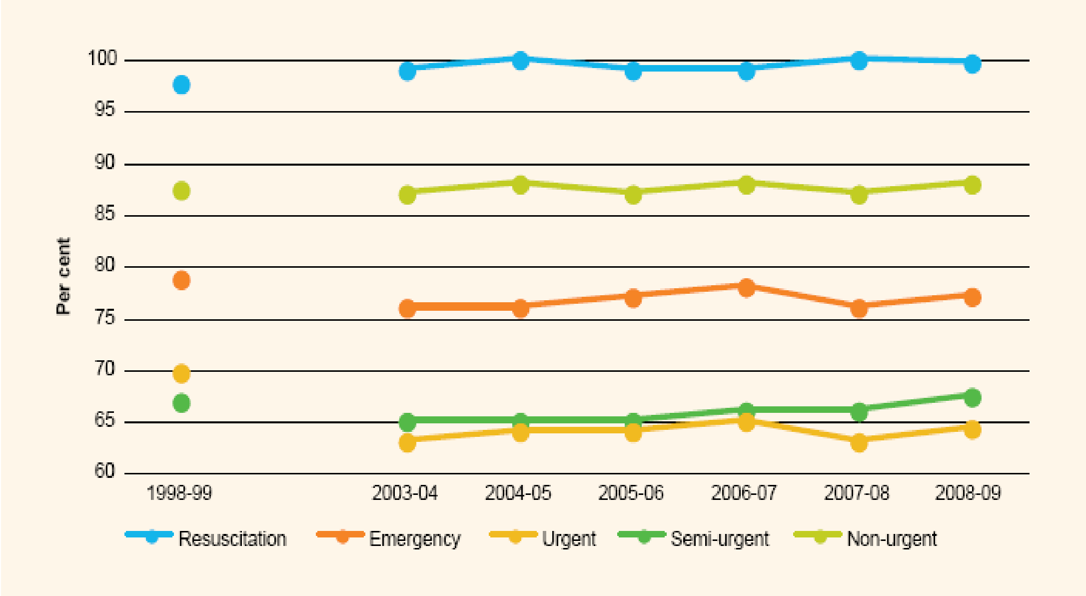


Figure 2-5: Percentage of emergency department presentations seen within recommended time by triage category, public hospitals, 1998-99, and 2003-04 to 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 27)

The median waiting time for emergency treatment in public hospitals decreased from 25 to 23 minutes for all emergency presentations, from 2003-04 to 2008-09. In 2008-09, the change was from 24 minutes to 23 minutes from the previous year (Australian Government Department of Health and Ageing, 2010).

The Australian Institute of Health and Welfare (2011-2012) reported that the overall proportion of patients who were *'seen on time'* was 54% in Northern Territory and 76% in New South Wales and South Australia. 50% of patients received their treatment within 21 minutes and 90% received their treatment within 108 minutes of presentation to the ED. 28% of ED patients were admitted as hospital inpatients.

O'Connell et al. (2008) reported that the processes in the ED are often time consuming and this is often due to the admission of patients who are old and frail. Often these patients have multiple health problems which take longer to assess. It is also harder to stream these patients to a medical team because inpatient teams often specialise in single-organ problems.

The data presented in Figure 2-5 showed that there were no major improvements in the waiting times at the ED for ATS 2 to ATS 5 categories. Apart from ATS 1 category patients; who were seen according to the recommended time, compliance to these recommended times were poor for all other ATS categories.

Since the proportion of ED presentations in each category only changed by about one per cent from 2003-04 to 2008-09, this should facilitate the development of good prediction models. Yet, prolonged ED waiting times are still reported. Most ED time-based studies have focussed on boarding times; further segmentation of ED waiting times could give insight into the implementation of targeted process improvements. Therefore, in this research, ED waiting times were categorised into phases in order to gain a deeper understanding of the complex interactions that take place in the ED.

2.3.3 Waiting list for elective surgery

The Australian Government Department of Health and Ageing (2010) reported that there were 595,009 elective surgery admissions to public hospitals, with a median waiting period of 34 days in 2008-09. Of these, 86.2 per cent were admitted within the recommended waiting period. The graph in Figure 2-6 shows the trend in the percentage of elective surgery patients admitted within the recommended waiting period in 1998-99 and 2003-04 to 2008-09. Access to elective surgery may be measured by the percentage of patients admitted within the recommended waiting period. Most surgeries are performed during the week in public hospitals.

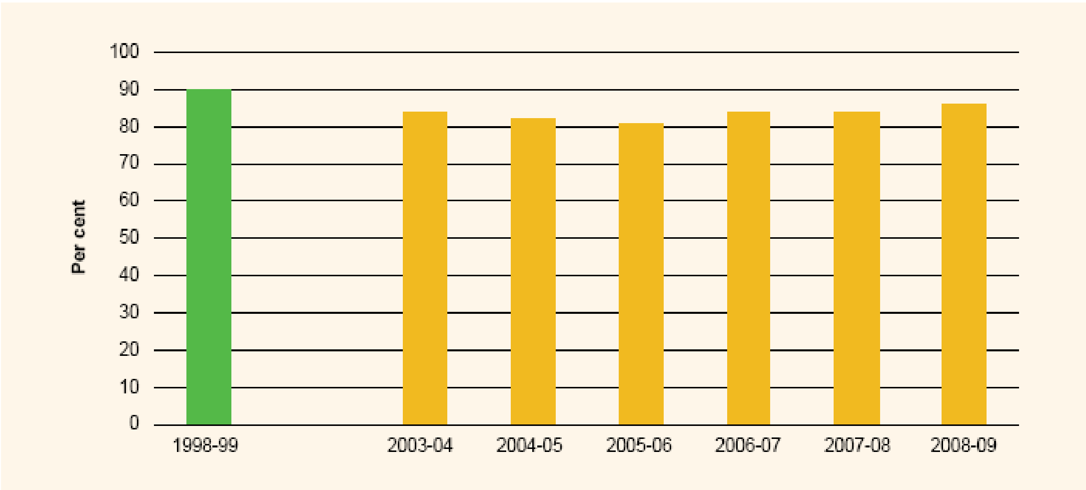


Figure 2-6: Percentage of elective surgery patients admitted within the recommended waiting period, public hospitals, 1998-99, and 2003-04 to 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 21)

Dunn (2003) undertook research during a period of industrial action when a hospital stopped taking patients for elective surgery but their ED continued to function as per normal and the ED was staffed according to normal staffing arrangements. During the study period the following results were observed:

- 5.9% reduction in hospital occupancy
- 23% reduction in the mean occupancy of the ED
- 72% reduction in the frequency in which the ED was required to operate in excess of its bed capacity
- 51% of the reduction in ED occupancy was due to decreased numbers of admitted patients awaiting hospital transfer out of the ED
- 37% reduction in the mean waiting time
- Improvements in waiting times for ATS categories 2–5

Dunn (2003) concluded that modest improvements in bed access resulted in highly significant reductions in ED waiting times.

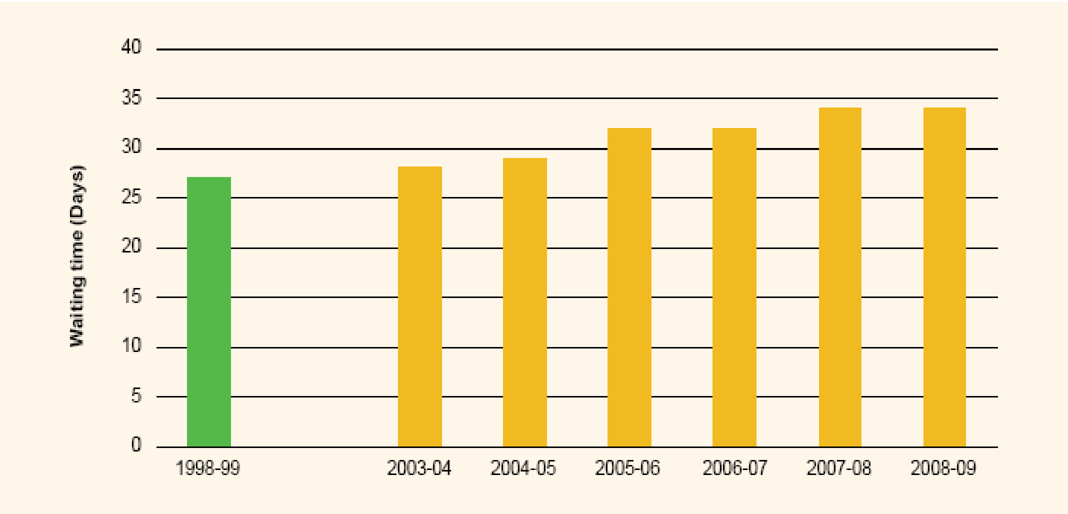


Figure 2-7: Median waiting time for elective surgery patients, public hospitals, 1998-99, and 2003-04 to 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 22)

Figure 2-7 shows the median waiting time for elective surgery patients. The median waiting time was 27 days in 1998-99, 28 days in 2003-04 and had increased to 34 days for 2007-08 and 2008-09.

These reports indicate accelerated access to an inpatient bed is a useful strategy to reduce ED waiting times. On the other hand the waiting list for the elective surgery patients might steadily increase if ED access to available beds is prioritised. Improving the access of emergency patients to inpatient beds cannot be implemented at the expense of reducing the amount of elective surgery.

However, process improvements around enabling efficient discharge practices may aid in creating the required beds. Efficiency within the hospital may be affected by other issues such as staffing levels. Therefore, this research looked at the outcomes for ED patients who present to ED inside or outside of working hours. If staffing levels affect the delivery of care in the ED, they may also impact on other services in the hospital including the discharge processes.

2.3.4 Hospital capacity

The Australian Government Department of Health and Ageing (2010) reported that there were 56,478 available beds in 756 public hospitals in Australia in 2008-09. This equated to 2.5 beds per 1000 population.

The Australian Government Department of Health and Ageing (2010) identified that hospital services can be assessed according to the number of available beds. Figure 2-8 shows the average number of beds per 1,000 population at all hospitals on a state by state basis. An available bed includes same-day beds, neonatal cots, hospital-in-the-home beds and overnight beds that are immediately available for the patient.

	Public hospitals		Private hospitals	
State	Beds	Beds per 1,000 population	Beds	Beds per 1,000 population
NSW	19,805	2.6	6,714	0.9
Vic	12,869	2.3	7,391	1.3
Qld	10,805	2.4	6,414	1.5
WA	5,369	2.4	3,109	1.4
SA	4,874	2.7	2,308	1.3
Tas	1,275	2.3	1,035	1.9
ACT	875	2.6	391	1.2
NT	606	3.5	104	na
Aust	**56,478**	**2.5**	**27,466**	**1.2**

Figure 2-8: Average number of available beds per 1,000 populations, all hospitals, 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 46)

Howell et al. (2008) stated that active bed management has enhanced collaboration between physicians in different departments and between physicians, nurses and case managers.

Braitberg (2007) argued that hospital managers are focussed on improving operational efficiency. Average bed occupancy is used for modelling the required bed numbers for that community. However, analyses based on averages are too weak to predict a complex system within a typical public hospital and the complex dynamics of patient flow within that hospital.

Analysing the number of beds alone is not a sufficient indicator to assess the service delivery. Appropriate resources (*e.g. nurses, beds in the correct ward*) should be available to provide care for each patient occupying a bed. Further insight is needed into how hospital capacity affects the delivery of care. This research looked at the consequences of admitting patients to a ward that is distant to their home ward. This often occurs because of a lack of beds in a particular home ward.

2.3.5 Staffing / Resources

The Australian Government Department of Health and Ageing (2010) reported that there was the equivalent of 246,895 full-time staff in Australian public hospitals in 2008-09. This number includes the allied health professionals such as radiographers, laboratory technicians, physiotherapists, pharmacists and other health professionals who provided additional support during a patient's admission. Figure 2-9 shows the breakdown of these full-time staff.

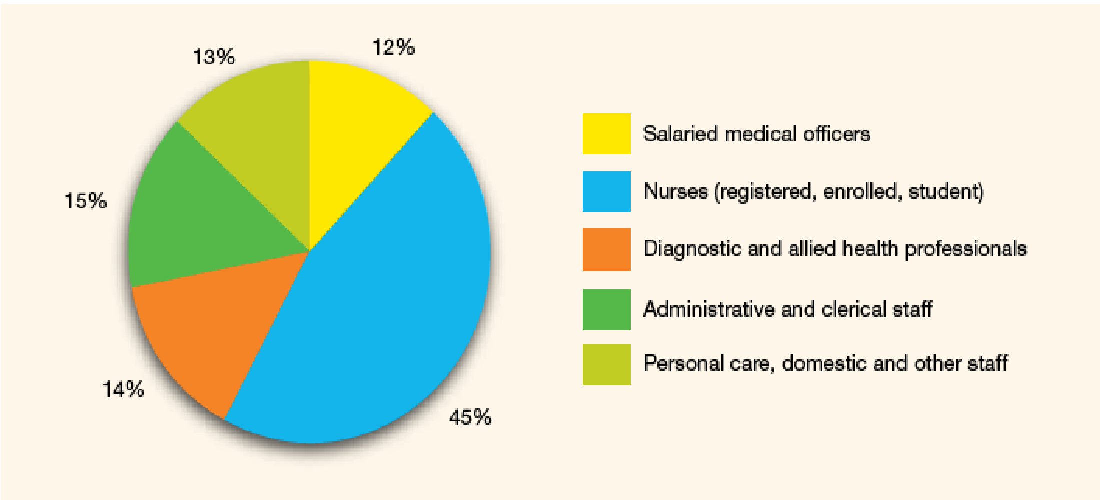


Figure 2-9: Percentage distribution of full-time equivalent staff by category, public hospitals, 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 56)

Patient flow through the hospital is also dependent on a number of support services for the hospital inpatient. The level of support or care received by the patient might influence the timing of discharge; more support might facilitate an earlier discharge. This could be especially true with allied health support such as services provided by a physiotherapist, audiologist, occupational therapist, psychologist, dietician and/or chiropractor.

Detailed study of hospital staffing was outside the scope of this research largely due to the inaccessibility of such data. However, this research was able to look into a surrogate measure of resources. Although the ED is staffed to operate 24 hours; other parts of the hospital are staffed differently inside and outside working hours. For example, the pathology laboratories will have a shorter turnaround time during working hours and there will be more administrative staff to facilitate patient discharges during working hours. Because the ED is dependent on these support services, a misalignment of the availability of hospital resources was hypothesised to impact upon ED efficiency. Therefore, the efficiency of the ED in delivering care was assessed inside and outside of working hours.

2.3.6 Physician autonomy

Hospitals employ a large number of highly qualified personnel. Physician autonomy is seen as one of the issues that contributes to the uncoordinated and often disengaged hospital processes. Abraham and Reddy (2010) specified that the challenges faced by the hospitals in patient transfer include uncoordinated inter-departmental interactions. Abraham and Reddy (2010) concluded that the uncoordinated inter-departmental issues involved superficial status differences between clinical and non-clinical staff and competing professional hierarchies. These issues also lead to another disruption to smooth patient transfers between wards and/or units leading in turn to inefficient work hand-over between care providers.

Physician autonomy is a common problem in almost all hospital settings. Doctors, often acting in the best interests of a patient, sometimes prefer for a patient to remain in a ward longer than absolutely necessary to ensure all tests are completed even in situations where such testing could be completed in an outpatient clinic. Another example of disruptive physician autonomy is when one unit refuses to give over a long-standing agreed bed allocation to another unit which might be more in need of beds.

Examining data on the effect of such physicians' decisions on the overall patient flow and hospital capacity would lead to better co-operation between the various personnel involved in a patient's care. Adding transparency to complex hospital processes will be very useful for care optimisation.

Further research into this area was not feasible from the data supplied by the hospital.

2.3.7 Length of Stay (LOS)

The Australian Government Department of Health and Ageing (2010) reported that in 2008-09 the average LOS for patients admitted to public hospital for at least one night (overnight admissions) was 6.3 days. Patients admitted to public hospitals stayed longer than patients admitted to private hospitals as shown in Figure 2-10.

State	Public hospitals	Private hospitals	All hospitals
NSW	6.5	5.4	6.2
Vic	6.3	5.6	6.1
Qld	6.0	5.5	5.8
WA	6.2	4.8	5.7
SA	6.9	4.8	6.3
Tas	7.6	n.p.	n.p.
ACT	5.9	n.p.	n.p.
NT	5.9	n.p.	n.p.
Aust	**6.3**	**5.3**	**6.0**

Note: 'n.p.' not published.

Figure 2-10: Average length of stay (days) for overnight admitted patients by hospital sector, 2008-09 (Australian Government Department of Health and Ageing 2010, pg. 65)

The LOS differences between the public and private hospitals are not comparable without considering the different types of patients presenting to these hospitals.

LOS data are often used by the hospitals as a measure of efficiency of care. The average LOS alone was not deemed a sufficient indicator of hospital performance.

In this research LOS data was used and analysed in different context. Firstly; it was used to assess the LOS of ward inliers and ward outliers and secondly; it was used as one of the overall QoC attributes.

2.3.8 Patient flow

Richardson and Mountain (2009) reported that EDs are designed largely for a continuous flow of patients. They provide an initial diagnosis and treatment and then on-going treatment is provided as an inpatient outside of the ED.

Richardson and Mountain (2009) also reported that patient flow is highly dependent on ED occupancy. A small variation in treatment rate is cumulative and this can be manifested in an increase in the number of patients waiting for treatment with a priority ahead of most new arrivals.

In order to gain further insight into the consequence of ED occupancy and the effectiveness of care delivery within the ED, this research assessed the impact of the number of patients in the ED upon the delivery of efficient care.

O'Connell et al. (2008) asserted that ED congestion is intensified by any failure to manage those processes involved in progressing patients through the hospital. Some of the contributing factors for this failure were:

- the disorganised nature of discharge planning
- the reliance on visits by medical staff for decisions
- the lack of a shared understanding among staff, patients and carers concerning the likely patient path
- an increase in undesirable events and length of stay if patients are not admitted to their home ward

These statements emphasise the need for better inpatient management ensuring a better patient flow which in turn should reduce issues faced by the ED.

In order to gain insight into the complex patient journey, hospitals need to look beyond conventional statistical modelling. To improve patient flow, the hospital management need to understand more about the hospital-wide patient journey. Therefore, the focus of this research was to use recent advances in Information Technology (IT) to undertake hospital-wide patient journey modelling.

2.4 Strategies used in hospital research to improve overall hospital performances

The majority of the strategies reviewed in this section primarily relate to the strategies used at the FMC because the case studies were based on data supplied by this hospital. However, within each sub-section, the use of the relevant strategies in other healthcare organisations are referenced, showing these strategies are also used in other settings.

2.4.1 Lean thinking

The National Primary and Care Trust Development Programme Website (2003) defined Lean Thinking as a philosophy widely used in manufacturing industries which is equally applicable to the healthcare industry. The concepts of lean thinking are to simplify processes, to identify patient care processes that add value, to facilitate effective care flow and to eliminate waste. Lean Thinking is a quality-improvement methodology (Andersen et al., 2014).

Cooper and Mohabeersingh (2008) stated that lean thinking in healthcare has the following benefits:

- it gives the opportunity to resolve inefficiencies
- it achieves costs reduction
- it achieves a higher turnaround with the same staff and processes in a more efficient way
- it provides careful and accurate ways to implement change

Cooper and Mohabeersingh (2008) also highlighted the role of the Practice Manager in implementing Lean Thinking. Leadership skills, good communication skills and a highly influential personality are essential in driving change from inception to implementation.

Process mapping is one of the main activities undertaken as part of Lean Thinking. According to Trebble et al. (2010), process mapping enabled the mapping of a patient's pathway or a process of care according to the patient's experience. The process is broken into a series of consecutive steps and events from admission to discharge and include activities and interactions between the patient and relevant staff or departments. The aim of process mapping was to maximise clinical efficiency and competence. Redundant and unproductive steps are removed when undertaking process mapping thus eliminating waste. As a result of process mapping, patient pathways consist of steps focussing on those events and activities seen as most important or valued by the patient. This improves QoC and efficiency of clinical management.

Many hospitals around the world have adopted Lean Thinking to improve patient care. These hospitals have seen benefits in improved quality, access, efficiency and reduced mortality (Mazzocato et al., 2010).

Successful implementation of Lean Thinking at FMC has seen improvements in ED performance. Lean Thinking has contributed to a better understanding of the patient flow process and therefore better co-ordination. Lean Thinking implementation has reduced the number of patients leaving the ED without being seen by a doctor, stabilised staffing levels, reduced the number of adverse events, shortened the LOS for medical patients without increasing the rates of unplanned readmissions (Ben-Tovim et al., 2007).

Lean Thinking has been successfully implemented at FMC's ED (King et al., 2006). Lean Thinking has also been reported as an effective improvement methodology and a framework for change management in nursing work at FMC (O'Neill et al., 2011). However, according to a recent report by the National Health Performance Authority (2012), the period of time until most admitted patients (90%) departed ED for admission to a ward was 14 hours and 1 minute at FMC compared to 12 hours and 25 minutes at the Royal Adelaide Hospital (RAH). Both FMC and the RAH belong to a peer group of similar hospitals developed by the National Health Performance Authority (2012) ensuring a fair comparison. The hospital environment, possibly due to its complex nature, needs a strategy that would enable continuous process improvement to occur. A lack of evidence concerning the sustainability of the application of Lean Thinking has been observed by others (Andersen et al., 2014). Mazzocato et al. (2010) reported that although Lean Thinking is a hospital-wide initiative, most reported Lean Thinking implementation is targeted at improving unit specific processes. Therefore, most of the process improvement initiatives do not cross the organisational boundaries.

Lean Thinking methodology practiced at FMC could benefit from the process mining techniques undertaken in this research. One area of Lean Thinking that could directly benefit from process mining is the process mapping activity, undertaken at the initiation of Lean Thinking project. This will be further discussed in Chapter 8.

2.4.2 Redesigning the Patient Journey

Redesigning the patient journey requires looking at the patient journey with the view of creating a smooth journey for the patient and removing bottlenecks in the journey. Such bottlenecks can hinder the smooth progress of the patient through the various steps involved in the care process. Such patient-centred care is a positive step and redesigning the patient journey helps with overall flow within the hospital system.

O'Connell et al. (2008) reported that redesigning the patient journey for a large patient stream or group will be very beneficial. A strategy known as “fast track zones” was implemented in New South Wales hospitals. This strategy streamed lower urgency or complexity patients into a separate zone in the ED. They were treated by an experienced group of medical staff who could also be supplemented by nurse practitioners and experienced physiotherapists. This strategy enabled quick decision making.

At the John Hunter Hospital, ED patients were streamed into high acuity/high complexity; low acuity/high complexity; or fast track (low acuity/low complexity). As a result, that hospital had improved access block and triage times and an improvement in their ED LOS (O'Connell et al., 2008). Another success story was at FMC where ED patients were streamed into two categories – those who were likely to be admitted and those who were likely to go home. This streaming contributed to improved access for patients and reduced the number of undesirable events (O'Connell et al., 2008).

According to Ben-Tovim et al. (2007), patient streaming at FMC has seen an instantaneous halving of the number of patients who left without treatment. Streaming of the patient journey also reduced the average time spent in the ED from 5.7 hours to five hours. This success was sustained the following year despite a 10% increase in ED presentations. The average time spent in the ED was reduced by six minutes.

This initial success at FMC shows that concerted effort into process improvement activities is an important step and can yield positive outcomes. The aim of this research was to investigate the applicability of process mining framework in healthcare that would give insight into the underlying process of a complex healthcare setting. This insight will empower hospital management with the knowledge needed to undertake cost effective continuous process improvement.

2.4.3 Clinical Process Redesign

According to Ben-Tovim et al. (2008b), clinical process redesign is the use of process redesign and change management in healthcare. The activities in the redesign process focus on the patient's perspective. The patient is the only entity in the hospital who sees the entire journey as they move from ward to ward or from unit to unit during their treatment. On the contrary, a care provider only sees the process from the perspective of their department and that care provider may be unaware of the processes extant in other departments with which they are not directly involved. As a consequence, a patient's journey is generally not coordinated as well as it could or should be (Ben-Tovim et al., 2008b) .

The aim of clinical process redesign is to harmonise the poorly coordinated patient journey as the patients move between multiple departments. The aim is to make the patient journey simpler while looking at the overall design of the surrounding clinical processes (Ben-Tovim et al., 2008b) .

Clinical Process Redesign (CPR) demands drastic and sometimes unusual changes to the overall processes. These changes might be made in small increments. Businesses undertake Business Process Redesign (BPR) to gain a competitive advantage over others. In regards to healthcare the advantage is centred upon delivering the best care for patients.

CPR strives for a best model for the health system in order to provide safe, efficient and well-coordinated service for all patients resulting in major improvements in QoC (O'Connell et al., 2008). Fundamental to any CPR is the continuous enhancement of processes. It also requires management working hand in hand with the frontline staff. The frontline staff must take ownership of their processes and work environment (O'Connell et al., 2008).

MacLellan et al. (2008) talked about the successes of undertaking CPR activities for the processes of care for planned arrivals at New South Wales hospitals. CPR has allowed hospital management to recognise a significant number of blocks and inefficiencies.

Another success story of CPR has been in the operating theatres. Harders et al. (2006) reported that well harmonised multidisciplinary process redesign reduced non operative time (*room turnover time plus anaesthesia induction and emergence time*) considerably. This was confirmed by a three month multidisciplinary study focussing on reducing non operative time tasks.

CPR takes a holistic approach by looking at a wide area for process redesign. Process improvement is a vital element in any system and requires continuous fine-tuning and adjustment in order to continuously adapt to the ever transient nature of the hospital system. Therefore, what is needed is a framework to ensure that any positive outcome gained from undertaking CPR is not only sustained but a new benchmark is established for the process to continually mature and be accredited at the next level of maturity.

2.4.4 Healthcare modelling

There is plenty of literature on healthcare modelling using various mathematical modelling techniques combined with simulation. Researchers have used these modelling techniques to model various healthcare related scenarios. The results of these modelling are used for forecasting and predicting in order to improve healthcare performance. Health care modelling was also used to evaluate cost effectiveness of certain hospital procedures and to derive treatment costs incurred by hospitals when providing certain services.

Karnon et al. (2009) used compartmental models using the Bayesian Information Criterion in describing bed occupancy in an acute hospital. Results of their healthcare modelling using compartmental models could be understood with ease both by clinicians as well as other decision makers. As a result better strategic decision making could occur. Karnon et al. (2009) demonstrated from their study that modelling and simulation have key roles in improving healthcare systems at all levels.

Jones et al. (2009) concluded that a multivariate time series approach to modelling was more reliable in forecasting ED patient census when compared to standard univariate time series method. However their forecast for diagnostic resources was not reliable hence they suggested that their methods were not practical to be used in a clinical setting. Jones et al. (2009) suggested the use of queuing theory, optimisation and simulation modelling together with the use of analytical methods would give strong results for decision making.

Another study undertaken at Show-Chwan Memorial Hospital in Central Taiwan used simulation techniques and a genetic algorithm to assess the QoC in an ED. Yeh and Lin (2007) concluded that patients' queue times were shortened by making appropriate fine-tuning to nurses' schedules and not by hiring additional staff. Yeh and Lin (2007) stated that their simulation model took into account the complete patient journey through the ED. They used their genetic algorithm to search for a near-optimal nurse schedule that also reduced patients' queue times.

The success of a model is dependent on how well the model represents the system that is being modelled. Representing a real system in a model is far from a trivial exercise. The modelling difficulties multiply when the system being modelled is a complex system such as the hospital. Therefore, in this research the process mining framework was investigated as a potential tool that might ease the difficult task of modelling a complex hospital system. A system that is modelled to adequately reflect reality will be able to do a better job at predicting and forecasting.

2.4.5 Simulation - Discrete Event Simulation (DES) in healthcare

There is a substantial amount of literature on the use of DES in healthcare. According to Sally C. Brailsford (2007) the most extensively used simulation approach in healthcare is DES. The main focus of this section is on DES. Hoot et al. (2008) stated that almost any ED overcrowding can be predicted using a thorough and adequate simulation of ED patient flow. Their findings also revealed the possibility of accurately predicting ambulance bypass. The findings by Hoot et al. (2008) further elaborated on the limitation of their model. Some of the outcomes of their simulation, such as ambulance bypass time and boarding times, were systematically biased and they proposed ways to reduce these biases. Their simulation model was flexible and described the state of the ED in regards to patient presentations (*as opposed to a standard regression model where the model depends on a specific definition of crowding*) (Hoot et al., 2008).

Côté (1999) carried out a study using DES to estimate the effect of potential changes to the environment on an individual physician's family based practice. They stressed that the applicability of this approach to model a simple and uncomplicated operation to generate meaningful analyses with the prospect to revisit the way the operation policy or process was carried out. Incremental process evaluation and improvement are essential when evaluating the positives and the negatives of a current process to form a good foundation to initiate any process change.

To further benefit from DES, Ceglowski et al. (2007) proposed combining data mining techniques for identifying bottlenecks in the patient flow between the ED and a hospital ward. The outcome of such data mining work can provide insight into the complex relationship between patient urgency, treatment and disposal and the occurrence of queues for treatment. Ceglowski et al. (2007) stated that an ED with sufficient data can identify bottlenecks by gathering and preparing data and following some simple data mining techniques. Those techniques might include dividing patients into homogenous groups, non-parametric clustering and so on in order to get an overview of ED workload and the most frequent relationships between ED and the hospital wards. The strength of the method proposed by Ceglowski et al. (2007) was in taking advantage of the data mining results and the benefits of DES. Ceglowski et al. (2007) were able to reduce the ED to ward transfer time for certain patient groups without increasing the number of ED beds.

Jun et al. (1999) concluded that DES is very applicable to healthcare because of multiple performance measures associated with healthcare. DES aids in understanding the relationships between various entities in healthcare systems with diverse input and output and performance measures. The inputs mentioned by Jun et al. (1999) included patient scheduling, admission rules, patient routing rules, facility and resource allocation whereas their outputs were patient throughput, waiting times and resource utilisation.

The success of DES is dependent on the input used for the simulation. Results of DES based on a weak model will be unreliable. Therefore, great care has to be taken to ensure the environment being modelled is fully understood. Traditionally, rigorous statistical analysis is undertaken to analyse and model the historical data. A good understanding of the system will provide a strong foundation for DES. Therefore, combining strategies that will give deeper understanding of the system such as using data mining techniques is important for a successful outcome when undertaking DES projects. In this research the use of process mining as a tool to give the insight needed to depict the complex hospital patient journey and for automating the discovery of a process model for DES was investigated. This is further discussed in Chapter 8.

2.4.6 Decision Support System (DSS) in healthcare

The use of DSS in healthcare is widespread. DSS in healthcare can be divided into two broad categories. One type of DSS can be used to help physicians in their day-to-day decision making processes. For example a DSS based on clinical practice guidelines was used in the management of diabetic patients (Lobach and Hammond, 1997). The other category of DSS can be used by hospital management to make decisions to improve coordination of the use of hospital resources. The following paragraphs discuss the DSSs that are used by the hospital management to make better informed decisions on hospital resource management.

Kusters and Groot (1996) reported on the benefit of a DSS that was used for admission planning in general hospitals. This DSS was based on a statistical model for predicting resource availability that affected patient throughput. The resources concerned were bed availability, operating theatre availability and nursing staff availability. It was concluded that a DSS based on statistical data and prediction model provided accurate information and have seen improvements in the admission planning. This goes to emphasise, the need for deeper understanding of the system is essential for a successful DSS implementation.

Aktas et al. (2007) proposed a DSS based on a Bayesian Belief Network (BBN) for the use of managers. The use of BBNs for healthcare management-oriented decision making lags behind its use in diagnosis and follow-up type analysis. The research by Aktas et al. (2007) further described why BBN was not a popular method upon which to base a DSS. They attempted to reduce the problem by specifying strategies to identify key variables which would facilitate construction of a DSS based on BBN. This system was able to assist management when making complex decisions.

Other DSS in the healthcare have been based on Knowledge Discovery from Databases (KDD) integrated with a Human-Computer Interaction (HCI) model (Ben Ayed et al., 2010). The method used a combination of Software Engineering’s Unified Process (UP) and HCI’s U model as a base model for DSS for a more user-centred approach.

This is another example of a commonly used tool in healthcare, which relies on good understanding of the system; gained from the use of historical data. This emphasises the importance of historical data. A robust methodology used to analyse the historical data and the outcome of the analysis consequently becomes the foundation for many important applications used in healthcare.

2.4.7 Process mining in healthcare

The Process Mining Group (2009) website states "Process mining is closely associated to BAM (Business Activity Monitoring), BOM (Business Operations Management), BPI (Business Process Intelligence), and data/workflow mining.” Process mining techniques are used for discovering the process knowledge based on existing event logs. Event logs are auto-generated computer logs with timestamps which record the execution undertaken by an information system. Investigating and analysing event logs yield real world understanding and a realisation of what actually happened. This information can be evaluated in terms of whether what happened in the process conforms to what should have happened.

An event log will be able to reveal the information below:

- Activity – the action undertaken in the process
- Timestamp – the timestamp / start and end time of the activity
- Originator – the executer of the activity

van der Aalst et al. (2007) and Weijters et al. (2006) stated that process mining aims to construct a process model from behaviour that has been observed from a process perspective, an organisational perspective and/or a case perspective. A process perspective enables the discovery of control flow or the ordering of activities or the ordering of a path. The organisational perspective aims to discover the relationship between the various originators in the organisation. Cases are grouped according to the path in the process or the originators.

van der Aalst et al. (2007) concluded that the most significant output of process mining is the discovery of the main process flow. Rozinat et al. (2008) stated that process mining is normally used to create a static model which can be given to the users of the system to reflect on the process.

Process mining is a Business Intelligence (BI) tool which aims to improve the operational business processes by amalgamating the knowledge from information technology and management science as defined by van der Aalst (2011). Unlike many mainstream BI and data mining tools which are data-centric, process mining is process-centric. It is aimed to gain insight about the processes to which the data refer. The focus of process mining is not on fancy-looking dashboards but rather on deeper analysis of the data (van der Aalst, 2011).

Process mining in healthcare is in its infancy. There were not many documented studies on the application of process mining in healthcare at the inception of this research in March 2010. The literature of interest was articles with direct application of process mining in healthcare setting. The following paragraphs capture the 11 scholarly articles published on the application of process mining in healthcare from 2008 to 2014. PubMed was used for this literature search. The intent of this literature search was to document the application of this newly emerging technology in healthcare rather than the technical details of process mining.

Lang et al. (2008) carried out a detailed hands-on evaluation and analysis using event logs from clinical workflow. They used real-life event data from all ~15.000 Computer Tomography (CT), Magnetic Resonance Imaging (MR), Ultrasound (US) and X-Ray (XR).

Lang et al. (2008) concluded none of the seven process mining approaches were able to construct valid process models. The approaches failed to handle noise, incompleteness, multiple occurrences of activities, richness of process types and variants. However, they also reported good results were produced despite noisy data by the heuristic miner algorithm, the DWS-algorithm, and the genetic miner algorithm (Lang et al., 2008). The analysis was undertaken using the ProM toolkit. There was no mention on the number of records used in this study.

Mans et al. (2008) used process mining techniques on two separate datasets of stroke patients. The first dataset contained clinical course of stroke patients from admission to discharge for 368 subjects. The second dataset contained pre-hospital behaviour of 234 stroke patients. The heuristic miner algorithm within the ProM toolkit was used. Mans et al. (2008) were able to discover different clinical pathways and identified bottlenecks.

Fernandez-Llatas et al. (2010) reported classical process mining event-based algorithms were not fully compatible for clinical pathways paradigm. They therefore developed a pattern recognition algorithm based on the classical process mining paradigm. This new algorithm; an Activity-Based Process Mining algorithm called Parallel Activity Log Inference Algorithm (PALIA) was experimented using a Heart Failure Clinical Pathway data. They used 510 samples and concluded; PALIA achieved better results than classical process mining algorithms in a controlled experiment of Clinical Pathways. The classical process mining event-based algorithms compared with PALIA were heuristic miner, genetic process miner, alpha and alpha++ (Fernandez-Llatas et al., 2010).

Huang et al. (2012) reported in their article that traditional process mining techniques were unable to mine complex clinical pathway patterns. The complexity and the variation embedded in clinical pathways produced spaghetti-like patterns. The patterns were not easy to understand (Huang et al., 2012). Therefore, Huang et al. (2012) developed a new process mining algorithm to effectively mine clinical pathway patterns. They experimented with real-life clinical workflow data sets for six diseases from Zhejiang Huzhou Central Hospital in China. The dataset used were from August 2007 to September 2009. The algorithm were able to discover patterns that were interpretable by the clinicians and reflected the actual clinical pathways (Huang et al., 2012).

Huang et al. (2013) published another article using their algorithm. This article by Huang et al. (2012) was discussed in the previous paragraph. The same experimental dataset as the previous study was used. Huang et al. (2013) concluded their approach; a better algorithm than the traditional process mining algorithm was able to produce interpretable clinical pathway summaries.

Kim et al. (2013) undertook process mining using ProM and Disco (Fluxicon, 2012) on event logs extracted from HIS at a tertiary university hospital. They used one month's event log data accounting for 698,158 records. The aim of the study was to discover the outpatient care process using process mining techniques. Kim et al. (2013) used the heuristic miner and the fuzzy miner algorithms within ProM. The discovered model was compared with an expert-driven model recognised by the clinicians. Kim et al. (2013) concluded that the processes discovered were relatively accurate. Process mining techniques can be useful for real process improvement for outpatient care processes (Kim et al., 2013).

Caron et al. (2013b) published a paper titled "*A process mining-based investigation of adverse events in care processes"*. However, as of 6th March 2014, there was no further published information available on this paper.

Meneu et al. (2013) further discussed their process mining algorithm; Parallel Activity Log Inference Algorithm (PALIA) developed by the same group as discussed previously and published byFernandez-Llatas et al. (2010). This paper further elaborated the data source and once again concluded the traditional process mining algorithms were unable to mine complex clinical workflow. The PALIA algorithm performed better compared to the traditional process mining algorithms and is yet to be tested on a larger data set (Meneu et al., 2013).

Fernandez-Llatas et al. (2013) once again reported that their Parallel Activity Log Inference Algorithm (PALIA) performed better than the traditional process mining algorithms such as the alpha algorithm, heuristic miner algorithm and alpha++ algorithm. In this paper, Fernandez-Llatas et al. (2013) experimented on data captured by an Indoor Location System (ILS). The data from the monitoring system was extracted from nine residents of a Spanish nursing home for a 25-week period. The research was done based on the framework of the PALIA algorithm and Fernandez-Llatas et al. (2013) concluded that this algorithm is suitable to support experts in human behaviour modelling.

Caron et al. (2014) used process mining based approach to successfully discover recurring patterns, analyse, characterise process variants and identify medical events. They used data on the care process of 1,143 patients of the gynaecologic oncology department of a major European academic hospital from January 2005 to January 2008. ProM was used to undertake this research. In one paper, looking at the same population, Caron et al. (2013a) acknowledged the discrepancies in the accuracy of timestamps within their constructed event log. Caron et al. (2014) concluded in yet another paper, that process mining techniques are suitable for continuous quality and performance monitoring of care processes.

Kelleher et al. (2014) used process mining techniques within ProM to detect differences in clinical behaviour and monitor specific processes of paediatric trauma resuscitation patients. The study was done using data from the Children's National Medical Centre in the District of Columbia. The event log contained 222 records. The main outcome of the research done by Kelleher et al. (2014) is from the strength of their statistical analysis. Therefore, in conclusion Kelleher et al. (2014) acknowledged that their data was not specifically collected for process mining analysis but was suitable to be used for process mining.

To summarise; five separate papers by (Fernandez-Llatas et al. 2010, Huang et al. 2012, Huang et al. 2013, Meneu et al. 2013, Fernandez-Llatas et al. 2013) out of the 11 papers were undertaken by two groups. These studies introduced new process mining algorithms for mining clinical workflow. The authors argued; traditional process mining algorithms were unable to produce correct process models. One group extended their PALIA algorithm to analyse logs from ILS. These authors developed another tool to implement and test their algorithms. Another study undertaken by Lang et al. (2008) reported seven of the process mining algorithms within the ProM toolkit were unable to handle the complex clinical workflow data. The same study however, reported the heuristic miner and the genetic miner algorithms were able to produce good results. The rest of the studies were undertaken using the ProM toolkit; reported successful results. However, except for Kim et al. (2013) who used a larger data set around ~600,000 records; the other studies undertaken separately by (Mans et al. 2008, Caron et al. 2013, Caron et al. 2014, Kelleher et al. 2014) used smaller samples ranging from 234 to 1143 records. Another common characteristic observed in all studies were the scale of the study. All studies undertook department or unit specific process mining studies. No hospital-wide initiative reported in these studies; although the study undertaken by Kim et al. (2013) could be assumed to contain outpatient data for the entire hospital.

Rebuge and Ferreira (2012) concluded that although process mining techniques have been proven in some instances to be successful in mining of health data, there is still room for improvement when trying to identify the right algorithm to handle noise in the data, the complexity of the data and the ad hoc nature of health data.

In this research the aim was to undertake hospital-wide patient journey modelling using the algorithms that are already available within the ProM toolkit.

2.4.8 Workflow modelling in healthcare

Malhotra et al. (2007) described their workflow modelling in Intensive Care Unit (ICU) as pieces of puzzle. During their workflow modelling activity, they followed key individuals and recorded all the interaction in the ICU. They used ethnography observation and interview data to build individual pieces of the workflow. They divided the entire workflow into individual activities, then isolating these activities in chronological event based on their critical importance. They integrated and built individual workflows in these critical zones. In the final step they analysed the relationship of these critical zones to one another and eventually formed a complete workflow. The aim of this workflow analysis was to help reduce medical errors as it was perceived if a medical error was encountered in one activity section than similar error could be caught at another section. Ultimately, Malhotra et al. (2007), aimed to use this method in conjunction with developing a taxonomy to discover root causes of medical errors leading the way to error prevention and prediction. Information discovered was perceived to be used for the development of DSS in this area.

Workflow analysis using ethnographic observations and interviewing techniques does not suffice to capture concurrent activities from various angles therefore quantitative way of capturing and analysing workflow is needed (Vankipuram et al., 2010). Vankipuram et al. (2010) also reported the connection between medical errors and the role of workflow analysis in discovering the underlying defects that contributed to these medical errors. They also described the two types of methods used to analyse workflow in healthcare which were qualitative and quantitative methods and discussed the advantages and disadvantages of these methods then proposed for a workflow monitoring solution using both methods. Other methods used in this research were Hidden Markov Modelling (HMM) and Visualization of workflow in a virtual environment.

A fairly new area of workflow study is the concept of workflow mining. Information Technology (IT) has become vital aspect of an organisation. More organisations are tapping into new advances in IT with the view of getting a competitive niche in their product/service offerings. Information Systems (IS) have evolved and have become more process aware. Organisation are relying more on their information system to extract information on the actual work that was undertaken in the company on a day to day basis. Organisations are also using information systems to implement processes identified as enhancing productivity. Information systems of this era are seen as process enabling systems because the system or software is designed around enabling certain processes. A common characteristic of these information systems is the ability to capture events that has taken place in the form of logs. Every action or activity undertaken in the system is tracked and recorded. Workflow mining uses the event logs of an information system to undertake workflow mining. The event logs are mined using various mining techniques to discover the workflow. van der Aalst et al. (2003) discussed the concept and the emergence of workflow mining, the intricacies involved and the techniques used to discover a workflow from runtime data to support workflow design and analysis. This outcome as described in this paper is an “a posteriori” process model which can be compared with the “a priori” model.

Rozinat et al. (2008) stated that the workflow model was configured using some real-world process data based on the organisational model which was designed before the enactment. When the process is simulated, the workflow model or software will record the event logs.

The enactment of workflow model using observed behaviour and by merging the event log information into a simulation model enabled the construction of a more accurate simulation model. Process Mining techniques are then used to analyse the event logs and to generate the simulation model. This approach of using the extracted historic and state information from an operational workflow as a starting point of the analyses is a way to calibrate the simulation model (Rozinat et al., 2008).

Workflow Modelling is the action of modelling a process in terms of sequence of activities. Rozinat et al. (2008) reported that the main aim of workflow modelling was to focus on the enactment and then knowledge discovery from simulation.

Vankipuram et al. (2010) in their report highlighted the role of workflow analysis in complex clinical environments. Unpredictable and complex nature of workflow in healthcare sometimes makes it difficult to capture some aspects of the workflow and suggested the use of both qualitative and quantitative methods in order to provide a complete picture of the workflow. Vankipuram et al. (2010) further explained that visualisation of workflow can be used as an education and training tool in understanding the workflow to limit potential errors.

In this research, workflow mining is not the focus. Clinical workflow process mining has been attempted by various groups. Workflow mining is undertaken from event logs taken from clinical workflow systems and these systems are department or unit specific. The aim of this research was to undertake a hospital-wide patient journey modelling rather than department or unit specific modelling.

2.5 Conclusion

The application of process mining in healthcare is still in its infancy. Process mining relies on event logs from PAIS as the source of its data. Clinical workflow systems are PAISs. However, in healthcare and in many other industries, the availability of enterprise-wide event log is scarce. For example modelling a patient journey from start-to-end to reflect all ward or unit movements require event log that spans across multiple units. On the contrary, modelling patient journey for a Cardiology unit only can rely on event log of a clinical workflow system used in this unit.

To undertake process mining using the ProM toolkit, the event log needs to have certain characteristics. The event log with the characteristics needed by ProM toolkit can only be sourced from certain type of information systems (*e.g. PAISs*). In the absence of an event log from PAIS, process mining is then dependent on derived event log. This is further discussed in Chapter 4.

When using derived event log, the source of data for process mining is the same as any other methodology that relies on stored data. So, this makes process mining susceptible to the guidelines and prerequisites needed for the use of health data. The unavailability of auto generated event log and therefore the need to use derived event log is seen as one of the main challenges of process mining. This challenge has to be addressed first before process mining can be successfully tested and used in healthcare environment. In this research, deriving the event log in the absence of auto-generated event log is one of the research agenda.

As reflected throughout this literature review, data has played a major role. The statistics presented in Section 2.2 and Section 2.3 was the outcome of analysing data collected by the hospital. Healthcare modelling relies on data to undertake sophisticated modelling using statistical science in order to understand and consequently predict future events. DES based on a strong model constructed from the use of historical data has been proven to produce superior outcome. DSS based on sophisticated statistical models gives a strong foundation for a successful system. These examples highlight; firstly the availability of data and secondly the need for insight or deeper understanding of the data as the foundation to develop the system.

Process mining is advocated as a methodology to give deeper understanding of the underlying processes. Process mining framework looks into a process from a case perspective, organisational perspective and the control flow perspective (van der Aalst, 2011). However, all the applications of process mining in healthcare reviewed thus far have only reported on process model discovery which is from a control flow perspective. This research will address the application of process mining in healthcare from all perspectives. The case perspective of process mining will be complemented with the use of the already mature statistical science in healthcare. The organisational perspective of process mining will also be explored in this research.

As reflected previously, current application of process mining in healthcare has been limited to specific units or departments. There is no evidence from the literature reviewed of any systematic framework for deriving an event log; the fundamental element needed for process mining. Therefore, this research will be the first research using a novel way for deriving hospital-wide event log. This enabled all three process mining perspectives to be applied and tested using real-life dataset.

3 Gaining insight from patient journey data using agile process-oriented analysis approach (Methodology)

This chapter discusses the methodology proposed to analyse the patient journey dataset supplied by Flinders Medical Centre (FMC) using an agile process-oriented approach. The methodology proposed in this chapter was accepted by the Australasian Workshop on Health Informatics and Knowledge Management 2012 and was published in the Australian Computer Society in the Conferences in Research and Practice in Information Technology series (Perimal-Lewis et al., 2012b).

3.1 Introduction

A process is defined as "*a series of actions or steps taken in order to achieve a particular end*" (Oxford University Dictionary, 2014). An admitted patient's journey through the hospital system from start-to-end consists of a series of steps or movements. Patients move from ward to ward and/or from unit to unit. Multiple teams of doctors (units) and other services are involved in providing the continuum of care as each patient flows through the hospital system. Each individual service can consist of one or more units. Each unit contains a specialised team of doctors. Each unit is allocated a certain number of beds in a ward. Figure 3-1 shows the range of services offered at FMC. It also shows the range of units under each specialty. The number of beds allocated to a unit is based on historical information. Medicine, Cardiac & Critical Care services and Surgical & Speciality services have the highest number of units. GM has the largest number of units at FMC. All the GM units in Figure 3-1 are enclosed in a red box. There are many patients admitted under the care of GM doctors and these admitted patients are distributed amongst the many GM teams.

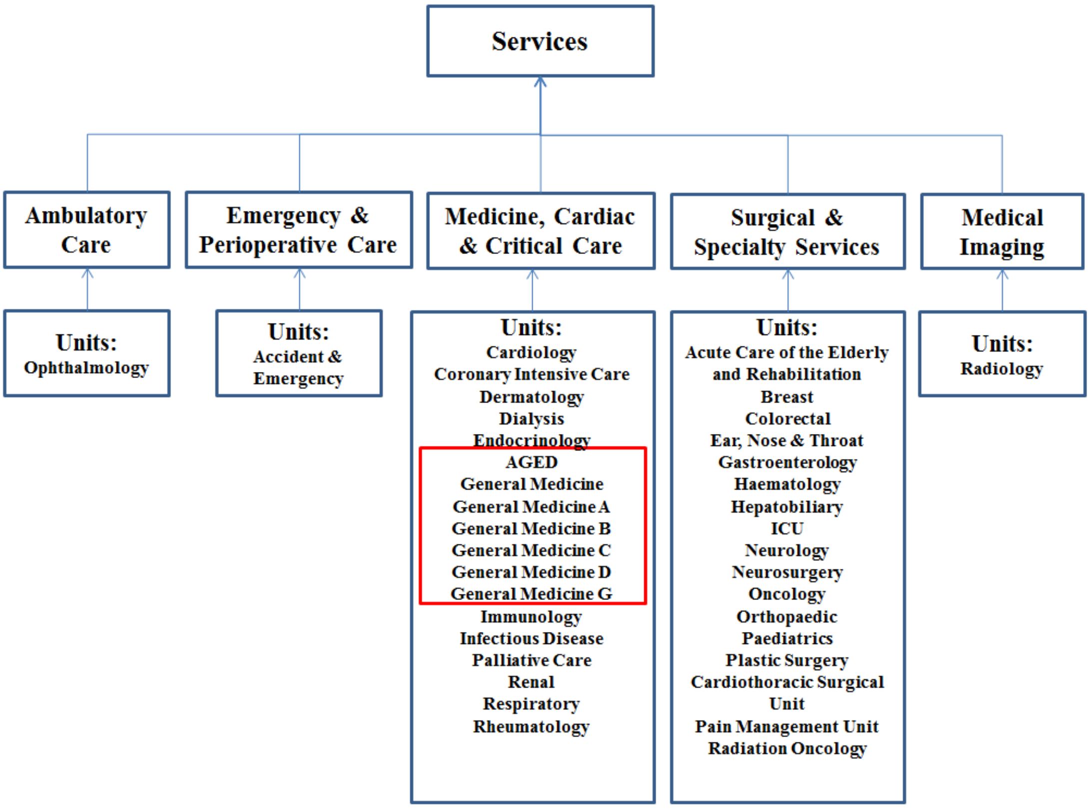


Figure 3-1: The relationship between services and units at FMC

At FMC, there are generally two types of wards: medical wards and surgical wards. Each ward has a team of nurses that look after the admitted patients in their specific wards.

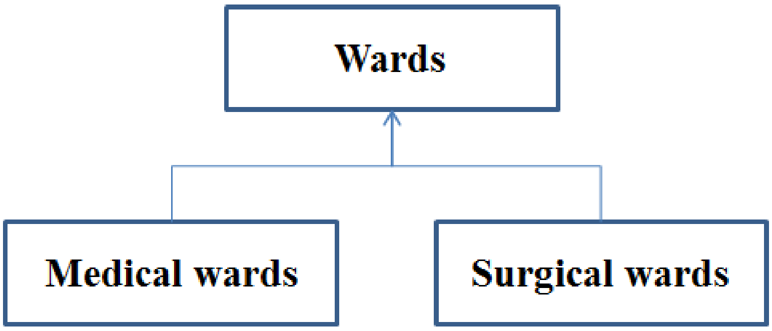


Figure 3-2: Types of wards

Figure 3-2 shows the two types of ward at FMC. Typically medical wards are staffed with medical nurses trained to look after medical patients and surgical wards are staffed with surgical nurses trained to look after surgical patients.

A patient could potentially move from unit to unit to receive the required care from the various services across the hospital (Ben-Tovim et al., 2008b). A medical unit could transfer its patient to a surgical unit and vice versa. Changing a care team (unit) could potentially be accompanied by a physical transfer of ward. The units although working together in the delivery of care for a patient, continue to maintain their independence. The units are self-governing and often practise autonomous decision making. As a result a patient's care could be fragmented. The various units are decentralised, run by doctors who have their own funding models and their goals relate to the delivery of the best possible care for those patients under their care. Contrary to the decentralised structure of units, the nursing structure follows a more centralised model. Each patient's journey on the other hand flows across this decentralised structure of units and each one is cared for by nursing teams who follow that more centralised structure. The combination of these different models of care would naturally cause a disruption in flow. As a result, a coordinated approach for process improvement activities involving a patient's journey is a challenging task. This model is unique to the hospital system making it challenging to incorporate some of the best practices outlined in Business Process Reengineering (BPR) and there is a call for '*radical change*' of the system as championed by Hammer and Champy (1993). Initiating radical change for process improvement needs a unified vision from all stakeholders.

The complex structure of the hospital system poses another challenge when deciding where to start any process improvement activities. This dilemma is however not unique to the hospital system, it is also recognised as one of the main challenges in other industries. Recently Zubizaretta (2013) stated that not knowing where to start is the main challenge in process improvement activities for accounting and financial reporting processes. So, modelling the patient journey through the hospital system for process improvement purposes is a far from trivial exercise especially when deciding the starting point of the project. The decentralised nature of the units meant that the Clinical Indicators (CIs) measuring the performance of each unit differ from unit to unit. Each unit has its own set of KPIs or CIs and these unit specific KPIs are monitored relatively closely as reported in the Clinical Indicators (CIs) report by The Australian Council on Healthcare Standards (2013).

In IT projects one of the main project start-up activities is establishing the boundaries and scope of any project. Many IT projects fail because of a poor project scope which then contributes to projects running over-time and over-budget (Whitney and Daniels, 2013). In the complex structure of the hospital system discussed so far, the inherent challenges in establishing the scope for hospital-wide patient journey modelling are magnified. Therefore innovative methods are needed to define the scope of any project thus avoiding the premature abandonment of projects. The research outlined in this thesis proposed and demonstrated the applicability of a novel way to undertake hospital wide patient journey modelling.

This research takes advantage of the inherent decentralised nature of the hospital system and the KPIs or the CIs used to measure the delivery of patient-centred care. Recognising the decentralised structure of the hospital highlighted the fact that QoC measurement would naturally vary from unit to unit. The measurement of QoC is dependent upon the primary care offered by a particular unit. Therefore the QoC measurement for a GM unit is different from the QoC measurement for a Cardiology unit. The various QoC measurements are described in this report by The Australian Council on Healthcare Standards (2013). This insight enabled an innovative way of deriving event logs using the concept of data plug-in. Using knowledge of the various KPIs or CIs as a foundation; the scope of data collection for hospital-wide patient journey modelling was defined. This is discussed in detail in Chapter 4.

Another recognised feature of most abandoned IT projects is the failure to recruit appropriate stakeholders to champion the project (Whitney and Daniels, 2013) . A top-down approach to process improvement in a decentralised hospital environment is not practical because of the disparate goals in the delivery of patient-centred care. Therefore a framework that enables hospital-wide patient journey modelling from a grassroots level is more likely to succeed. The framework proposed and demonstrated in this thesis for deriving the event logs for process mining takes advantage of the unit specific KPIs and is discussed in Chapter 4. Process improvement is more likely to be championed by units if it is directly relevant to the delivery of patient-centred care specific to each unit’s KPIs. The sustainability of any process improvement project is therefore dependent on the relevance of such a project to each unit.

Acknowledging the complex nature of the hospital systems, the decentralised organisation of doctors, the centralised organisation of nursing team and the impact of these conflicting organisational structures upon the patient journey, the framework proposed and demonstrated in this thesis aimed to adapt an agile process-oriented methodology in order to model the patient journey process from start-to-end for the entire hospital system. The framework is agile because it builds on matters that are important in the delivery of patient-centred care that is different from unit to unit. Therefore the framework moves away from a top-down approach to hospital wide process improvement and away from the rigidity surrounding a specific methodology. The methodology proposed and demonstrated, builds not only on the overarching process mining framework but also takes advantage of process mining as a HI tool. Adaptation of the process mining framework for patient journey modelling through the entire hospital system was investigated from a case perspective, from an organisational perspective and from the control flow perspective. In adapting process mining as HI tool, the focus was on deriving the event logs by assimilating the actual properties of an automatically generated event log of an Information System (IS). The derived event log is a fundamental component required for using process mining as an HI tool. The unit specific KPIs were used as the basis for scope definition and for optimising the sustainability of the project.

The following sections of this Chapter discuss the process mining framework based on these three perspectives.

3.2 Method (Methodology)

3.2.1 Process Mining

Process mining is a newly emerging technology. The fundamental component needed for process mining is an event log. Process mining aims to discover knowledge from event logs of a real process in order to develop a concise assessment of reality. This is contrary to the norm of looking at aggregate data alone for decision making (Song and van der Aalst, 2008). Song and van der Aalst (2008) also categorised process mining activities into process discovery, conformance and extension which, in turn, can be analysed from a process perspective, an organisational perspective and from a case perspective. Therefore process mining uses event logs to gain insight into the actual processes of the organisation (van der Aalst, 2011).

In this thesis, the process mining framework was applied to a complex healthcare environment in order to gain insight into the underlying processes embedded in any patient journey. The methodology modelled the patient journey based on the three process mining perspectives: case perspective, organisational perspective and control flow perspective. Process mining undertaken under the umbrella of a case perspective focusses on the property of each of the cases (van der Aalst, 2011). In this methodology the cases are the patient journeys. Each case or patient journey has a set of attributes that describe a patient journey. These attributes could be patient specific information such as age and gender and/or could be based on information collected as part of KPI reporting. The derived event log has features that enable customisable data plug-in attributes that could be based on unit specific KPIs. Process mining undertaken under the umbrella of an organisational perspective focusses on resource information (van der Aalst, 2011). In this context the methodology focusses on resources in regards to hospital staffing inside and outside of working hours. Process mining undertaken under the umbrella of a control-flow perspective focusses on the ordering of activities (van der Aalst, 2011). In this methodology the control-flow perspective focusses on creating an error-free model for simulation. The process-oriented methodology proposed and demonstrated therefore is based on the foundation of these three process mining perspectives. Figure 3-3 presents an overview of the chapters in this thesis that demonstrate the innovative ways of applying these three process mining perspectives in order to gain insight into the complex patient journey process from start-to-end and thus assess patient QoC.

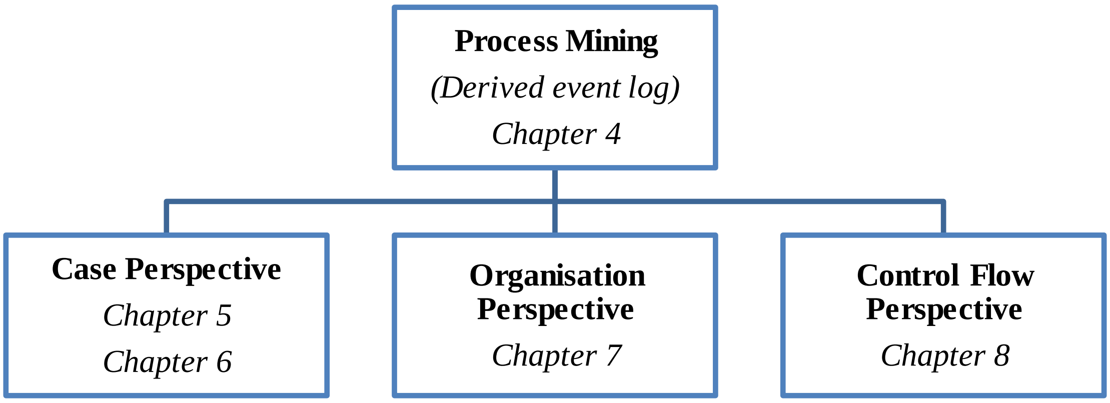


Figure 3-3: Overview of chapters addressing innovative ways of applying the three process mining perspectives

Process-oriented methodology was applied to the patient journey data from FMC which was pre-processed to adhere to the unique properties of event logs that will be discussed in Section 4.3.3. The application of the methodology to real hospital data demonstrated the practicality of the proposed methodology to model the patient journey.

This research utilised the derived event log to investigate the effect of being admitted as a ward inlier as opposed to a ward outlier on the QoC of that patient. At FMC, data on outlier patients is a regularly reported KPI. A complex group of patient journeys were examined; namely the GM cohort at a large public hospital.

In order to gain insight into the characteristics of the inlier and outlier patients, once the derived event log was available, it was important to undertake a series of complex data analyses to explore the GM population and an incremental approach was followed. The clinicians who were also the champions of this project helped with the interpretation of the results. The next step taken was dependent of the outcome of the previous step. The Process

Mining toolkit (ProM) version 5.2 was used for some part of the analysis. The complex nature of the health data being investigated called for advanced data analysis and statistical modelling which was done outside of ProM using Stata®, Microsoft® Access and Microsoft® Excel.

In order to gain insight from the data analysis, close collaboration with the practising clinicians was an essential element in this process-oriented methodology.

3.2.2 Process improvement champions

Regular contact with the clinicians enabled the portrayal of the actual undertakings at the hospital and this enhanced the quality of knowledge discovery and understanding. Clinicians' insight when explaining the processes within the hospital's context and the ways to interpret the result of the analysis was instrumental in deciding the next step to take in this research. The clinicians also played a major role in identifying data entry errors.

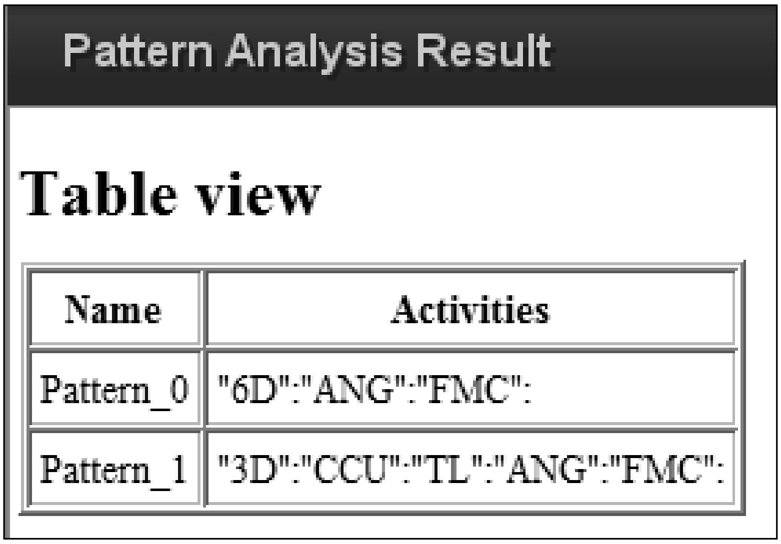
3.2.3 ProM (Process Mining) Toolkit

This section discusses the initial prototype that was created using a small subset of the derived event log using ProM as a tool to discover knowledge hidden within the historical data. The following section discusses the outcome of the process mining activity on this data set using ProM. The clinicians' feedback is also documented following each ProM output.

The derived event log was converted into the Mining eXtensible Markup Language (MXML) format so the data could be read by the ProM toolkit. The paragraphs below discuss some of the output of ProM which was used as a prototype when introducing process mining to the clinicians. The derived event log used for this prototype consists of five patient journeys only. This was deliberately done for the ease of understanding and interpreting the output. The output produced by ProM was cross checked using manual calculations as well.

The *Basic Log Statistics* produced simple statistical information from the data set. The statistical information that was produced comprised aggregate information about the time a patient spent in a ward: the minimum time, the maximum time, the arithmetic mean, standard deviation, geometric mean, sum and the number of times a ward was used. The clinicians responded that such information is readily available from current reports produced by the hospital.

Pattern Analysis output for the five patient journeys is shown in Figure 3-4. The pattern reveals two distinct flows or movements from ward to ward. The clinicians concluded from the output that ward "FMC" which is the Emergency Department (ED), was the source of admission for these patients.



Name	Activities
Pattern_0	"6D":"ANG":"FMC":
Pattern_1	"3D":"CCU":"TL":"ANG":"FMC":

Figure 3-4: Pattern analysis – patient journey flow sequence

Clinicians' feedback also immediately revealed that the ward sequence from the pattern portrayed in Figure 3-4 was not in the correct order. This prompted a closer look at the derived event log which revealed that the timestamp (especially the finer details of minutes and seconds) in the derived event log was also an important property and was needed to enable correct pattern discovery. Despite issues with the sequence, this pattern discovery was of interest to the clinicians. However, no further work was done using this function at the later stages of this research. ProM version 5.2 was not able to process the larger GM population (~20,000 records).

Figure 3-5 shows the frequency of ward usage. It shows the performance of each of the wards involved in the five patient journeys used as input. A quick glance at the diagram shows that one particular ward is being used much more than all the other wards. This information is also readily available to the clinicians by other means.

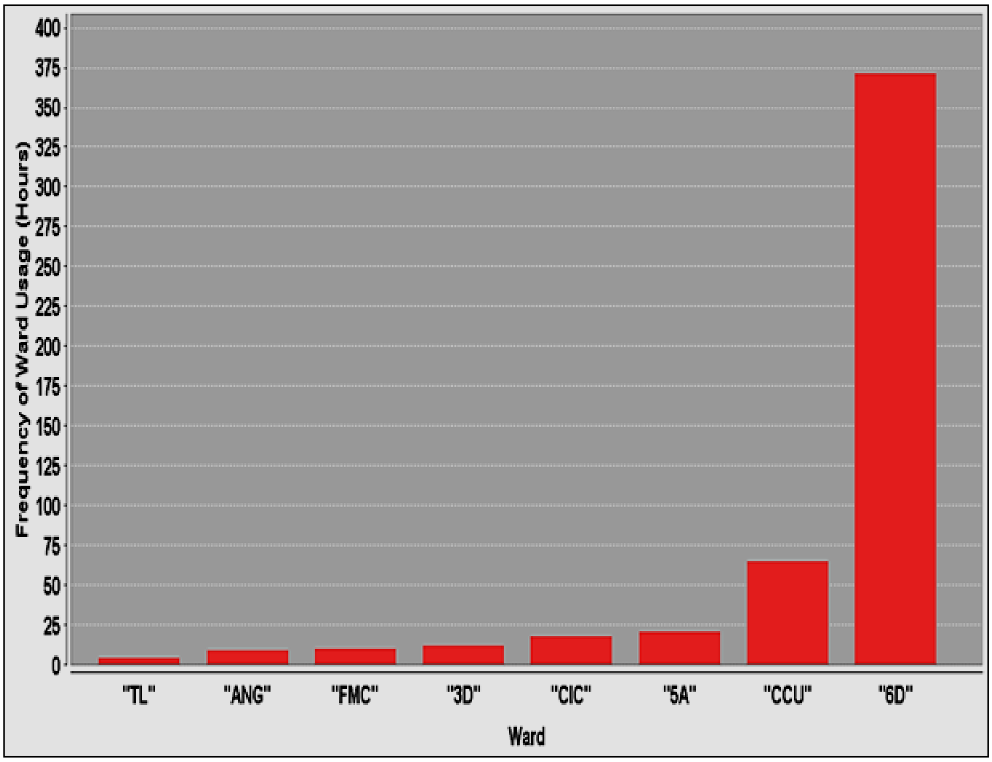


Figure 3-5: Frequency of ward usage

Individual patient journeys could also be further analysed to depict the LOS as per Figure 3-6. Clinicians were not interested in this information because when analysing a large data set, a graph showing LOS for individual patients will not give insight into the delivery of patient-centred care. Such a graph on a large data set where each individual journey will have a unique LOS will not be interpretable.

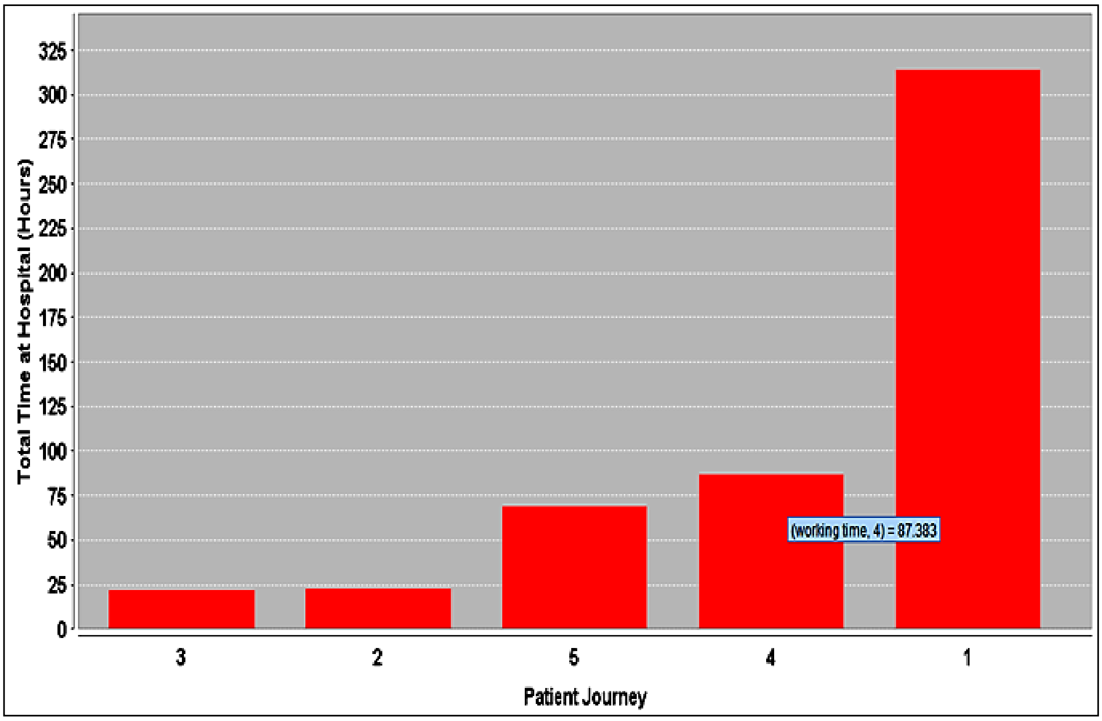


Figure 3-6: Journey Length of Stay (LOS)

Performance Sequence Analysis facilitates the assessment of the journeys as patterns and the 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 the 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 and could show the transfer of work between two doctors, the throughput time for shared patients as well as the frequency of certain behaviour or patterns relating to each doctor. The sequence diagram in Figure 3-7 shows the ward movements of the five patients. The labelled boxes on the top of the diagram are the names of the "wards" involved in these journeys. The numbered scale on the left is the time taken for the journeys as each patient moves from one "ward" to another "ward". The measurement unit shown in the diagram is minutes. The coloured lines represent the five different journeys. It is also apparent from the diagram, that two journeys follow the same path. This was of interest to the clinicians and will be further discussed in Chapter 4.

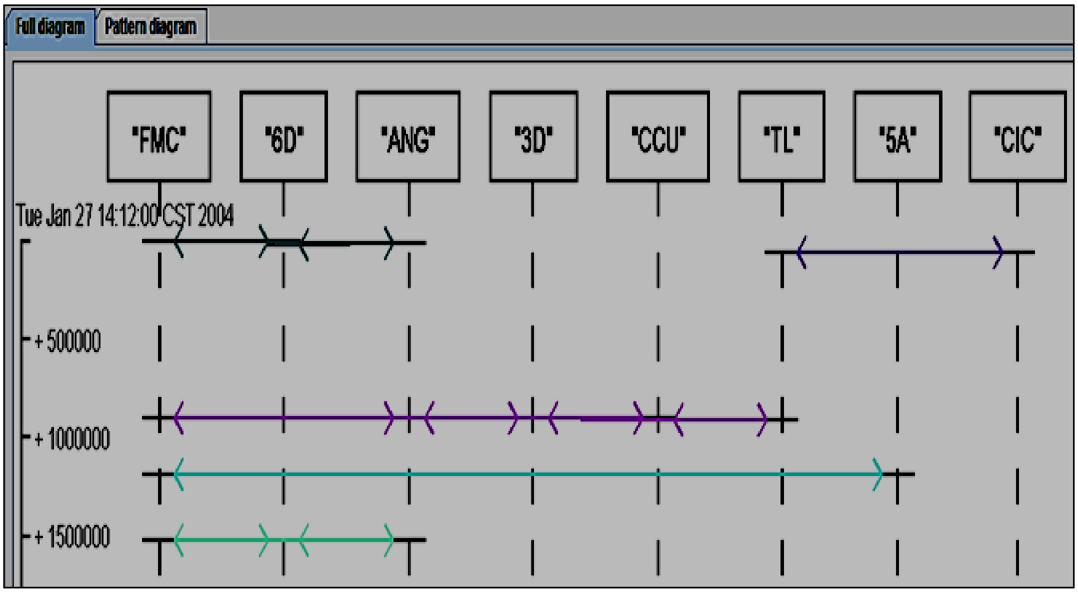


Figure 3-7: Performance sequence diagram

The sequence diagram Figure 3-7 was further processed into a *Pattern Diagram* as per Figure 3-8. Journeys with same path or pattern are grouped together. For example, in this set of data, there were two journeys where the patient moved from ward “FMC” to ward “6D” and then to ward “ANG”. These two journeys are grouped together forming one pattern and the rest of the journeys followed three different paths thus forming the other patterns. The clinicians were also interested in this diagram. This feature is further discussed in Chapter 4.

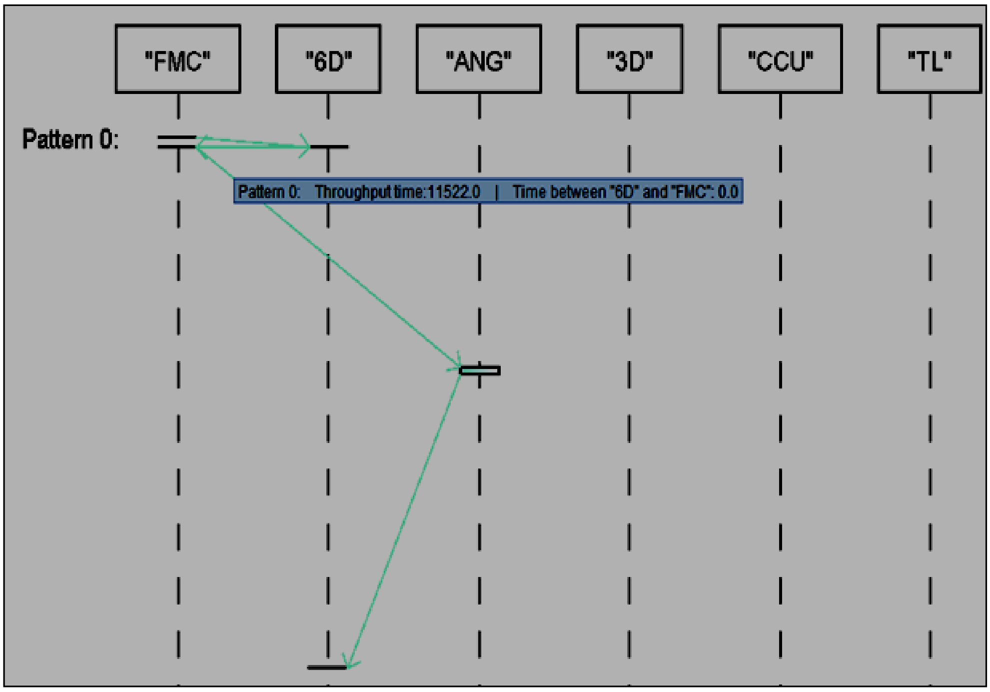


Figure 3-8: Pattern diagram

Further information for the pattern shown in Figure 3-8 is shown in Figure 3-9. The mean throughput time for each pattern is derived. The measurement unit is minutes. The frequency of occurrence for each pattern is also listed. Patterns with higher frequency indicate that such a pattern is common and the behaviours of such a movement could be further analysed. This information could be used to investigate the behaviour of patterns which could be a starting point to analyse "bottlenecks" where queues and inefficiencies of care can arise. The clinicians were interested in exploring this feature with a larger dataset.

Pattern 0:

Frequency: 2

	Throughput time
avg	11522.0
min	4170.0
max	18874.0
stdev	10397.29811

Figure 3-9: Pattern diagram information

Patient Journey flow discovery is as depicted in Figure 3-10. 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 number of journeys that have used that path. The decimal numbers next to the arrows represent the *Dependency Relationship*, between the two wards involved. A decimal number close to one indicates a strong dependency relationship between the two wards. A strong dependency relationship shows the flow in that path is likely to happen. Clinicians were also interested in this feature. The feature will be further explored in Chapter 8.

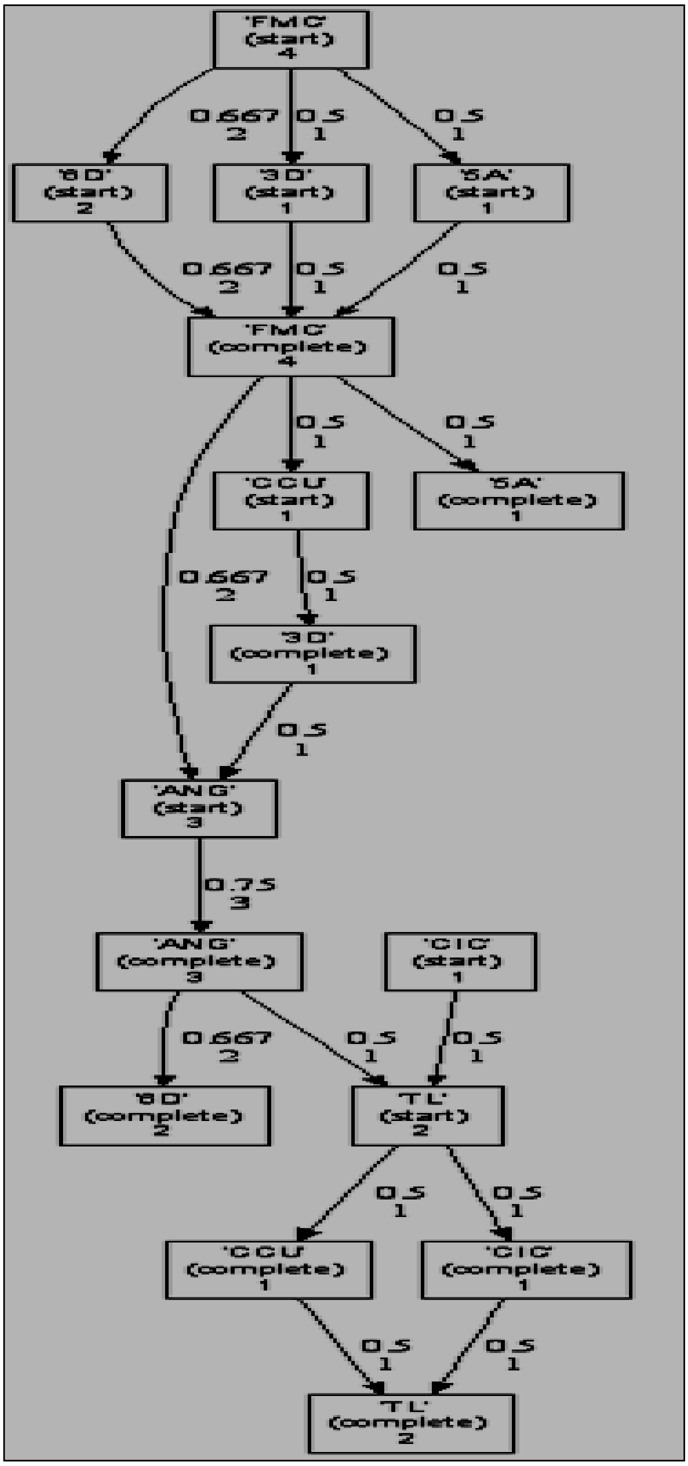


Figure 3-10: Patient journey control flow discovery

Figure 3-10 above shows the complete journey derived for five patient journeys previously discussed. This reinforces the fact that the flow discovery is complex.

3.2.4 Inliers vs. outliers LOS analysis

With little to support such an assumption, clinicians perceive that inliers have a shorter LOS compared to outliers. The testing of this assumption forms the basis of this particular analysis. The purpose was to investigate whether outliers had a longer LOS because their care in the outlier ward was inferior to the care offered to them had they been allocated to an inlier ward. The first cluster of patients included in this analysis was the GM patients. This study is presented in greater depth in Chapter 5.

It is a well-established perception among hospital staff that patients lying in a ward that is not the home ward of the medial team responsible for their care (*e.g. outlier patients*) have a longer LOS compared with those patients on their home ward (ward inliers). This perception is based on the impression of the ward outlying patients receiving a disruption of team-based care their QoC being compromised as a result and this mis-location of patients will contribute to these patients having a longer in-hospital LOS. There has been lack of evidence thus far to support this perception.

3.3 Discussion

The prototype using ProM presented to the clinicians has proven to be a viable approach in modelling patient journey. The collaboration with clinicians has been an invaluable experience in this process. Hospitals are already undertaking statistical data analysis for various reporting purposes to conform to KPIs and the approach taken in this research is to further break down the information and data to discover hidden knowledge.

Inpatient LOS has become one of the many ways used to measure performance of a hospital. Thomas et al. (1997) reported that patient mean LOS has been used to measure QoC and hospital efficiency in terms of resource usage. Thomas et al. (1997) further stressed that lower than normal LOS could indicate that hospitals are discharging patients early possibly sacrificing their QoC. Hospitals react differently to the continuously rising costs of healthcare. One reaction is to reduce the average inpatient LOS. Unfortunately some hospitals reduce the number of beds in the hospital as a direct response to increasing cost of healthcare.

3.4 Conclusion

A smooth patient journey is an important aspect of any patient’s experience in hospital. Process mining techniques will be used for knowledge discovery of the inpatient journey. LOS of inliers versus outliers was modelled using a process-oriented approach. Patient journey control flow was analysed using the Alpha algorithm and the Heuristics miner algorithms that are within ProM version 5.2.

The complex movement of patients shows that modelling patient journeys by following a data analysis approach combined with the use of process mining techniques would give further insights.

The results of process mining in conjunction with the usage of the proposed methods were perceived to offer added benefit to the already successful implementation of Lean Thinking. In fact the approach used in this thesis might possibly enhance those areas where the Lean Thinking approach alone is inadequate such as when investigating the causes and effects of access block.

4 Gaining insight into patient journey from derived event log using process mining

This chapter discusses the process of deriving an event log. Event logs are fundamental elements needed for process mining. A novel framework with the concept of data plug-ins was used to derive an event log which facilitated the hospital-wide patient journey modelling from start-to-end. All work covered from Chapter 4 to Chapter 8 used this derived event log with the appropriate data plug-in.

4.1 Introduction

The healthcare system provides a huge data repository promising many challenges for researchers to unravel the knowledge embedded in it. Health Information Systems are a big part of healthcare systems especially in developed countries. Electronic Patient Records (EPR) also referred to as Electronic Medical Records (EMR) are an indispensable component in modern healthcare delivery. The Australian health system is continually looking at ways to embrace IT to facilitate better healthcare delivery for the Australian public. Some common Hospital Information Systems (HIS) used by the Australian public hospitals are the Emergency Department Information System (EDIS), the Patient Administration Systems (PAS), The Open Patient Administration System (TOPAS), The Electronic Patient Administration System (ePAS) and the Hospital Morbidity Data System (HMDS) to name a few. In addition there are also unit specific Clinical Workflow Systems. These information systems are not necessarily integrated. One hospital could be using multiple information systems. These information systems collect a huge amount of data that is not just limited to patients' records but they also collect other data related to the clinical work practice of a hospital. In addition to these information systems there is also a large amount of information that is maintained using spreadsheets and in other unstructured or semi-structured formats. In a large healthcare setting such as the hospital, multiple services and units work in siloes serving specialised patient needs that differ from specialty to specialty. As a consequence of this need to provide specialised services, the diversity of information systems used should be appreciated. This calls for innovative ways of manipulating the data from these information systems when undertaking hospital-wide modelling; manipulation that is highly data dependent.

Patients move from unit to unit during their visit to the hospital. In a healthcare system that has embraced information technology, patient information is stored electronically in various information systems customised for each unit. In addition each unit could also have its own clinical workflow system. Therefore modelling the patient journey or patient flow from start-to-end through the entire hospital system poses many challenges that are unique to the healthcare system. The variety of information systems introduces unique challenges with regards to sharing information on the same entity (patient). To enable the study of patient flow through the entire hospital, an important ingredient needed is a universal way to share information about a single entity between all the units that work in isolation. The patient flow through the hospital system is subjected to significant variability of practice each specific to an independent department. In principle this horizontal flow is similar to the flow coordination problem in supply chains studied by Sahin and Robinson (2002) which called for a unified way of information sharing to ensure a smooth supply chain flow.

The information sharing of an entity (patient) in the Australian healthcare system is already present. For example, at FMC, patients are uniquely identified by their Unit Identification Number (URN). The same number is used to identify a patient in all the various unit specific information systems. This enables information sharing at a certain level. The experiences carrying out this research revealed that information sharing is still highly dependent on the paper-based patient medical file which has the URN attached. When a patient presents to a particular unit, the paper-based medical record is used by the treating unit to obtain the patient's medical history. Information that has to be recorded in the unit specific information system, if one is present, will usually be keyed in retrospectively. As such, from the data supplied by the hospital, the data types of many crucial fields are inconsistent. This makes it challenging to integrate data when undertaking hospital-wide patient journey modelling using process mining or any other techniques. Usually a considerable amount of time is spent on data pre-processing.

Most scholarly studies on process mining used transaction logs also known as event logs or audit trails extracted from the operational clinical information systems or clinical workflow system. These event logs record activities that were undertaken by the end users when using that particular information system. For example an event log of a workflow application might contain the ID of the end user, the type of task undertaken by the end user, the timestamp of when the task was executed and other relevant information that the application is programmed to capture. These kinds of event logs are application event logs. As described in the paragraphs above, the various units in the healthcare system often work independently. Therefore modelling the patient journey through the hospital presents a fundamental challenge with regards to information sharing. In the current research, the challenge was to recognise an individual patient's flow through the various hospital information systems each of which might have multiple event logs. For example, it would not be possible to find out from an event log where patient A moved after being discharged from an information system used by the Cardiology unit. This kind of patient tracing through the entire hospital system would only be possible to be mined from an event log of a single hospital-wide patient tracking information system. This research demonstrated that such a system was not in place at FMC nor at most public hospitals.

Another concern that needs highlighting is that the event logs of some information systems, especially of process-centric workflow applications, grow very quickly and can take up much needed hard disk space. Experience gained at a large IT service industry proved that the growth rate of these event logs meant that often these logs were only kept for a limited number of days before being deleted to create space for new logs. A patient could potentially stay for months at a hospital so this approach to event log maintenance might not be suitable for patient journey modelling. Some information system’s event logs are archived and cumbersome processes are in place to access these event logs. In some non-critical information systems, the IT support might completely turn off any logging capabilities.

For the reasons elaborated above, the chances are slim of obtaining an event log from a clinical information system that would enable the hospital-wide journey of a patient to be mapped.

There are studies of the application of process mining to healthcare that used derived event logs and these studies have been described in the literature review. All of these studies undertook process mining for a specialised unit or department rather than a hospital-wide patient journey modelling. Undertaking patient journey modelling for the entire hospital system poses challenges that are different to undertaking process mining for a single unit.

This is the first research undertaking patient journey modelling through the entire hospital system using the process mining framework at a large Australian public hospital in the absence of pre-existing event logs of clinical information systems and in the absence of PAISs.

4.1.1 Process Aware Information Systems (PAISs)

Modern information systems are developed to support the processes of an organisation. The emergence of multinational, global businesses with mature and complex business processes has essentially seen the emergence of enterprise information systems. These information systems span across the globe supporting a single universal process. These information systems are process-centric because the systems are designed to enable the execution of the enterprise’s workflow in a standardised manner. Standardisation of processes and having workflow systems that support the execution of these processes help an organisation to optimise the delivery of their core product or service. Therefore, information systems are no longer data-centric where the primary aim is to automate data collection, storage and retrieval. Modern information systems are process-centric. This shift has seen the emergence of PAISs.

PAIS is a software system managing the operational processes of an organisation involving human resource, application and information on the basis of process models (Dumas et al., 2005). All information systems have an automated recording of event logs or audit trails during the execution of the system. Since PAIS supports the business processes of an enterprise, the event log generated by PAIS has the potential to give insight into the actual execution of the process. Information gathered from these audit trails is unaltered and therefore is potentially a true reflection of how the system was executed by the intended users.

For instance the HPs Digital Workflow Enterprise Software (formerly Peregrine) implements an escalation process similar to the Australasian Triage Scale (ATS) categories in its Incident Management module. The helpdesk assigns the business cases to the appropriate teams. Each assigned case is categorised according to a severity index. The severity index attached to each case is an indication of how quickly the problem has to be resolved which is tied to the company's Service Level Agreement (SLA). Monetary penalties are attached to non-conformance to these SLAs. Therefore, there are strict requirements from the employee's perspective for conforming to the Standard Operating Procedures (SOPs) in a timely manner. The information system is developed to capture details in its event log that are of interest. The event log could then be extracted for analysis to gain true insight into how the organisation is performing. One simple analysis could be to check discrepancies between the timestamp of when a case was closed in the event log as opposed to the time recorded by the end user on the front end that is stored in a database.

Although hospital information systems are also PAIS there are fundamental differences in the way the end user (clinical staff) interacts with the system. For example the Emergency Department Information Systems (EDIS) used in most EDs in Australia are implemented around a sound workflow process. However, the main priority of the clinical staff at any point in time is to put the patient's needs first. So the information system naturally cannot be implemented to comply with tight rules. As such the data recorded in the event log, especially those attributes related to timestamps and personal information (*e.g. user identification number*), are less reliable. Although the event log is essentially a true reflection of when the system was executed, it might not be a true reflection of what took place. Of course, there will be times where the information system is executed according to the workflow but the chances of this structured process happening in the ED are slim. This inconsistency in audit trails has to be recognised. A recent report highlighted some limitations in the way EDIS is configured at the front end which raises issues on the integrity of information as reported by the Office of the Auditor General (2013). In a recent review, Callen et al. (2013) reported that clinicians find it difficult to embrace the use of technology with their work because of the significant data entry demand and the time taken to use the integrated ED information system.

Clinicians' focus upon the patient and they deal with processes that ensure that the patient receives appropriate care when needed. In a complex healthcare setting especially in the ED there are many interactions that take place. The frontline ED staff have to manage these interactions and at the same time ensure that the patients are not disadvantaged. Some interactions have been seen as detrimental to patient safety. Certain interruptions in ED have been widely studied and reported by Coiera (2012) and Li et al. (2012). Therefore, it is unrealistic to expect the frontline clinical staff to comply strictly with a workflow system and implementation of a strict workflow system may be counterproductive in that environment.

So, the validity of event logs or audit trails of PAIS systems remain subjective, although they might give insight into the real execution of the underlying processes around a workflow. The understanding of the nature of the work undertaken in a clinical setting (as opposed to other industries) and therefore the validity of the automatically logged event logs or audit trails is important when deciding the appropriate input data needed for process mining. As mentioned previously, the event log is a fundamental element needed for process mining, so a deeper understanding of event logs was needed in order to manipulate the intrinsic attributes of any event log. This understanding would enable the preservation of these attributes when creating a derived event log and thus take advantage of the process mining framework and the ProM toolkit.

4.1.2 Event log properties

The starting point of process mining is the availability of event logs (Song and van der Aalst, 2008). Event logs are also referred to as "history", "audit trail" and "transaction log" (van der Aalst et al., 2007). The fundamental concept of an event log is that it is a historical collection of a well-defined sequence of activities performed by a computerised system (van der Aalst and Giinther, 2007). As discussed in the previous section, there are no predefined prerequisites for the type of data recording to which an event log has to conform. What is programmed to be recorded in an event log is dependent on the purpose of that event log. Event logs are historically used to troubleshoot application faults or for the purpose of undertaking computer forensic investigation. For example when a system fails, the investigators look at the event log in the first instance to determine the last activity executed by the system before it failed. This becomes a starting point to build the troubleshooting effort or to undertake further investigations. In this context, the fundamental property needed in an event log is a code to identify the activity and the timestamp of when the activity took place. van der Aalst (2011) stated that the bare minimum attribute needed in an event log to undertake process mining is the ‘case id’ and the ‘activity id’ as per Table 1. The ‘case id’ represents a unique process instance. A process instance could have from one to many activities. An ‘activity id’ is linked to an event. From the ‘activity id’ the order of events can be established. Each line represents exactly one event. In summary, there are three main prerequisites for an event log suitable for process mining:

i. Each process instance is identifiable. In this instance, there are three process instances which are case 1, case 2 and case 3.
ii. Each activity in the process instance is identifiable. Within case 1, it is possible to identify activity 521 as distinct from activity 522.
iii. The activity within the process instance should be ordered or have attributes that would facilitate this ordering. For example in the Table 1, the ‘activity id’ is designed in such a way that the activities within a case can be ordered.

Table 1: Bare minimum attributes needed in an event log

case id	activity id
1	521
1	522
1	523
1	524
1	529
2	851
2	852
2	859
3	743
3	745
3	746
3	747

4.1.3 Event log for hospital-wide patient journey modelling - challenges

The main challenge posed for healthcare process mining especially when modelling a hospital-wide process is the absence of uniformed event logs originating from one information system. Therefore the event log for process mining has to be sourced from various information systems across the hospital. In this research, the goal was patient journey modelling from start-to-end encompassing the entire hospital system. There was no single hospital-wide system in operation. Therefore, obtaining an event log that would be able to reflect the movement of patients through the various units within the hospital is almost impossible.

In a complex hospital environment although the information system is designed around well-defined processes, the variation in day-to-day patient presentations is un-structured. Therefore the event log might not give a true representation of the exact process as discussed earlier. van der Aalst and Giinther (2007) have also highlighted this fact cautioning against assumptions that exact processes are represented in event logs.

Furthermore, in the multiple event logs from the various units, it will not be possible to identify a particular process instance (patient journey) common to another unit. For example, the ‘case id’ in each event log will be different. Therefore, it will not be possible to trace a particular journey or a particular patient’s movements from unit to unit from this event log. As such when modelling the patient journey through the entire hospital in the absence of a single hospital-wide PAIS, it would not be practical to rely on these automated event logs. To take advantage of the process mining framework, the event logs have to be derived from various data sources. The data collected from the various units should have a common attribute that would facilitate the amalgamation of these various data sources.

4.1.4 Ethics issues for derived event logs

Deriving event logs for healthcare process mining from various data sources will require the process mining initiative to obtain ethics approval for the usage of the data. The abstract nature of the event log is no longer applicable. For example, the ‘case id’ attribute alone in the event log is not enough to reveal the identity of a patient. However to derive an event log, the actual patient record has to be accessed. In Australia, this requires prior ethics approval. The project therefore had to comply with local requirements relating to ethics and guidelines. The sensitive nature of health data and the ethics laws surrounding the use of health data meant that open access to the information systems was not practical. Therefore only the historical raw data approved for release by the hospital were supplied.

Ethics approval for this research was granted by the Southern Adelaide Health Service / Flinders University Human Research Ethics Committee for the use of data from the patient journey database supplied by FMC.

4.2 Aims

The aim of the research was to derive a framework in order to derive an event log in the absence of a hospital-wide PAIS. That derived event log would facilitate the modelling of patient journeys from start-to-end. To demonstrate the applicability of the proposed framework, real patient data supplied by FMC from various data sources were amalgamated and used for knowledge discovery.

4.3 Method

4.3.1 Prerequisites for feature extraction

i. *<u>Identify research question</u>*

Process mining initiatives are carried out in order to gain insight into the hospital's operational processes. Therefore it was necessary to start the initiative with clear objectives and goals for the outcomes of this process mining. This is best done working in collaboration with clinicians involved in the processes being investigated. In the context of the case study presented in the Result Section 4.4, below is the list of the research questions:

- Gain insight into the ward movement pattern of GM patients at the hospital.
- Gain insight into the path followed by ward inliers and ward outliers and the outcome of efficiency of care.
- Identify GM wards and units that provide efficient care (throughput).
- Gain insight into the interaction patterns of the GM units.
- Gain in-sight into the organisational structure of the GM units.
- Confirm existing knowledge of confounding variables and efficiency of care for inlier and outlier GM patients.

ii. *<u>Identify data sources</u>*

Based on these research questions, it was established that the data needed for process mining had to be sourced from various HIS within the hospital. For various ethical restrictions in place for accessing these data, the data were extracted in a combination of native formats of the operational systems used by the hospital staff in compliance with the ethical policies governing the collection and usage of patient data. This occurred before the data were handed over for process mining.

iii. *Create a platform to process the data*

The data obtained for process mining were in .tab format, .xls format, .txt format and .mdb format. It was then necessary to amalgamate these data into one event log. Microsoft Access was chosen as an appropriate platform to be used to convert and amalgamate the data from these various sources. Microsoft Access was chosen because it was readily available and it had sufficient technical capabilities for creating the event logs. The main data set supplied was the patient journey data which had records of the movement of an inpatient from ward to ward from the time of admission to discharge. This dataset also contained time recorded for each movement or time recorded at midnight if the patient stayed in the same ward. It also had a date field. It contained the name of the unit in charge for treating a patient. The ward occupied by the patient was marked as an inlier or an outlier ward. This main dataset was then combined with other required fields extracted from the other dataset provided in order to derive the event log.

iv. *Pre-processing of unstructured or semi-structured data for process mining*

Data can be characterized as structured, un-structured or semi-structured. Structured data are organised in a highly regular way for the entire dataset (Losee, 2006). Structured data can be characterized by data types, have similar domains for a particular attribute and are mostly organised in tuples which comply for storage in Database Management System (DBMS). Semi-structured data share some characteristics of structured data but also contain data that are not organised in a manner complying with database design rules such as normalisation and data type. The attributes might not uniformly conform for the entire dataset. Un-structured data however do not contain precise structuring information (Losee, 2006). Un-structured data also may consist of images, e-mail, medical records and contracts to name a few. Healthcare data predominantly consist of un-structured or semi-structured textual data.

Deriving event logs with a combination of structured, un-structured or semi-structured data from numerous data sources is a challenging exercise because there will be overlapping data and inconsistency in data types. One could consider the exercise of deriving event logs for process mining as building a data repository for a consolidated dataset for the purpose of knowledge discovery or advanced analytical processing.

v. *Variable interpretation*

The meaning of each variable had to be interpreted and coded appropriately to ensure correct information was used and amalgamated. The amalgamation meant losing some records that had missing values.

4.3.2 Feature extraction for the derived event log

Deriving an event log for process mining is far from trivial. Having discussed the prerequisites for feature extraction in the previous section, this section discusses the process involved in deriving the event log for process mining. The concept of an event log as introduced by van der Aalst et al. (2007) shapes the foundation of the derived event log. A process log could be derived from a dataset that contains an order of events which could be used to construct a process model that portrays the activity of the subject matter (van der Aalst et al., 2007). Mans et al. (2013) distinguished four process mining data spectra namely the Administrative systems, Clinical Support systems, Healthcare logistics systems and Medical devices systems which could all be used as a data source for specific process mining activity. van der Aalst (2011) makes the following assumptions about event logs: a process consists of cases, a case consists of events such that each event relates to precisely one case. Events within a case are ordered and events can have attributes.

The goal of this research was to undertake a hospital-wide patient journey modelling from start-to-end to assess patient QoC. The decentralised structure of the hospital as described previously meant that it was necessary to establish a framework in order to derive the event log. The framework developed was flexible and allowed incremental development and amalgamation of data plug-ins as needed. The purpose of the data plug-ins was to satisfy the varied needs of each unit when delivering patient-centred care. The final event log should be agile enough to enable hospital-wide patient journey modelling as well as allow process mining techniques to be used to gain insight into important underlying processes surrounding the KPIs relevant to each unit.

Therefore once the bare minimum attributes discussed previously were established, the concept of data plug-ins was introduced to extend the bare minimum framework to undertake a variety of unit specific knowledge discoveries. This novel approach allowed the coordinated and incremental development of the derived event log. This event log was adaptable and was able to satisfy the data requirements when undertaking a ground-breaking large scale research project on ward inliers and outliers. The derived event log enabled this research to test the suitability of the application of process mining framework. This research took advantage of the open source process mining, the ProM toolkit.

By adapting the process mining framework, the entire patient journey could be seen as a process under investigation. The requirements for the bare minimum event log as discussed in Section 4.1.2 can be fulfilled with a slight variation. The bare minimum requirements for the event log needed for process mining are the 'case id' attribute and the 'activity id' attribute. A 'case id' enables each process instance or case to be uniquely identified. The 'activity id' attribute enables the identification of all the activities undertaken by each case. The way the 'activity id' is numbered enables the ordering of activities. In summary, the activities performed by the unique cases have to be ordered. Therefore a variation to the event log presented in Table 1 could be presented as per Table 2. The variation introduced to the bare minimum event log still fulfils the requirement that enables the activities performed by the unique cases to be ordered. The ordering is facilitated by the 'timestamp' column. This variation reduces the complexity of having a scheme to derive a numbering or coding system for the 'activity'. The 'timestamp' information is an easily available attribute from any database.

Table 2: Bare minimum requirement for event log with a variation

case id	activity name	timestamp
1	activity a	27/01/2004 14:12
1	activity n	27/01/2004 19:30
1	activity q	28/01/2004 0:00
1	activity w	29/01/2004 0:00
1	activity k	30/01/2004 0:00
2	activity y	3/03/2004 16:49
2	activity m	4/03/2004 0:00
2	activity t	4/03/2004 11:29
3	activity v	13/10/2005 21:33
3	activity e	14/10/2005 0:00
3	activity r	14/10/2005 0:36
3	activity k	14/10/2005 14:52

To be able to undertake hospital-wide patient journey modelling, the above variation translated to:

i. A case = a process instance, in this case a 'journey id'.

ii. Activity name = any attribute that relates to the process instance. For example this could be the name of a specific task (*e.g. register, admit, discharge*), it could be the name of a ward, it could be the name of the unit, it could be the name of a drug and so on. The main concept of interest here is that the activity should be something that relates to the process instance.

iii. Timestamp = the date/time recorded in the database when a specific activity took place.

The next step in this process was analysing the data supplied to derive a bare minimum event log specific to the environment under investigation. The raw files supplied were in both comma-separated values and tab-separated flat files. The records did not conform to any database structure. The required features from these flat files were extracted to form the bare minimum event log for hospital-wide patient journey modelling. Figure 4-1 shows a snippet of the patient journey tab-separated flat file supplied.

journey_id	type	urn	date	time1	time2	ward	unit	division
1	"AW"	31	27 Jan 04	852	1170	"FMC"	"CARD"	"Medicine, Cardiac and Critical care"
1	""	31	27 Jan 04	1170	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
1	""	31	28 Jan 04	0	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
1	""	31	29 Jan 04	0	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
5	""	84	30 Jan 04	0	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
5	""	84	31 Jan 04	0	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
5	""	84	01 Feb 04	0	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
5	""	84	02 Feb 04	0	825	"6D"	"CARD"	"Medicine, Cardiac and Critical care"
5	"W"	84	02 Feb 04	825	972	"ANG"	"CARD"	"Medicine, Cardiac and Critical care"
5	""	84	02 Feb 04	972	1439	"6D"	"CARD"	"Medicine, Cardiac and Critical care"

Figure 4-1: Patient journey tab-separated flat file

In this context, the patient journey data set supplied by the hospital contained the 'journey id' as per Figure 4-1. Therefore, in this case, the individual patient journey was comparable to the concept of process instance. Each patient journey is the process instance or the case being studied in relation to the activities on the patient journey. An activity in this case was translated to a ward occupied by the patient. From the 'date', 'time1' and 'time2' attributes, the timestamp could be easily derived. This process is further elaborated in the next section. Although the 'journey id' attribute satisfied the requirements to appropriately identify a process instance, another attribute that would identify a patient under all units is needed. Patients move from unit to unit. The universal identifier of a patient in the hospital is the patient's Unit Record Number (URN). Using the URN a patient's journey through the hospital could be tracked. For example, a patient treated in the ED would have the same URN when that patient is admitted to an inpatient ward. As such, the bare minimum event log needed for the hospital-wide patient journey modelling is the as per Table 3.

Table 3: Bare minimum event log for patient journey modelling

journey id	urn	ward name	timestamp of admission
1	25	FMC	27/01/2004 14:12
1	25	5G	27/01/2004 19:30
1	25	6F	28/01/2004 0:00
1	25	6F	29/01/2004 0:00
1	25	6F	30/01/2004 0:00
2	158	FMC	3/03/2004 16:49
2	158	FMC	4/03/2004 0:00
2	158	4D	4/03/2004 11:29
3	357	FMC	13/10/2005 21:33
3	357	FMC	14/10/2005 0:00
3	357	3G	14/10/2005 0:36
3	357	4D	14/10/2005 14:52

In Table 3, there are three unique patient admissions. A 'journey id' corresponds to an admission. 'urn' represents a patient. A patient could present to the hospital multiple times. Therefore, each time a patient is admitted to the hospital, a new 'journey id' is allocated. For example, if patient with 'urn' 25 is readmitted to the hospital, this patient would be allocated with a new 'journey id'.

The bare minimum event log for patient can now be extended to incorporate the concept of plug-ins to undertake analysis and/or process mining specific to various units or any other hospital wide research. The next section describes the process of data processing and the amalgamation of data sets to create a plug-in. This plug-in can be used for process mining activity to discover the control flow perspective for simulation (Chapter 8). Similar concepts were used for all other plug-ins.

Figure 4-1 shows a snippet of the dataset from the patient journey database. The patient journey dataset only contains data on an inpatient's ward movement from the time of admission to the time of discharge, therefore only data on officially admitted patients are recorded. The patient journey data supplied had to be transformed and merged with another set of raw data which contained patient data whilst in the ED. The ED dataset contained data on activities that took place on a patient's journey in the ED up until the patient was officially admitted to the hospital. The snippet of the ED data comma-separated flat file is shown in Figure 4-2.

```
Triage date,UR,Priority,Doctor,Date Time seen,Outcome Date,Outcome code,UR6,UR7,URNo,Unit
3/07/1993 11:00,4566982,2,KL,3/07/1993 11:02,3/07/1993 14:55,D/C,4566982,4566982,4566982,????
6/07/1993 16:04,4544545,3,PM,6/07/1993 16:17,6/07/1993 19:00,5G,4544545,4544545,4544545,????
9/07/1993 13:44,8220012,3,PO,9/07/1993 14:04,9/07/1993 15:57,5C,8220012,8220012,8220012,????
15/07/1993 18:50,7860436,3,JKL,15/07/1993 19:08,15/07/1993 22:28,D/C,7860436,7860436,7860436,????
24/07/1993 13:22,2576348,3,PSL,24/07/1993 14:04,24/07/1993 16:40,4F,2576348,2576348,2576348,????
29/07/1993 20:10,3698572,3,NY,29/07/1993 20:42,29/07/1993 23:42,4F,3698572,3698572,3698572,????
9/08/1993 15:52,3657921,4,YM,9/08/1993 16:02,9/08/1993 16:49,D/C,3657921,3657921,3657921,????
11/08/1993 8:50,9872357,4,YUL,11/08/1993 9:04,11/08/1993 10:19,D/C,9872357,9872357,9872357,????
11/08/1993 10:19,2479575,3,CR,22/08/1993 16:54,22/08/1993 19:23,D/C,2479575,2479575,2479575,????
```

Figure 4-2: ED data comma-separated flat file

The raw data from both flat files had to be transformed in the first instance and then merged to form the patient journey event log. MS Access 2010 was used to merge both datasets into a single database. Upon merging the dataset, the semantics of each field were established. Each field had to be converted to its appropriate data type. The major challenge for this process was the conversion of the date/time fields. In the patient journey dataset both ‘time1’ and ‘time2’ fields were recorded as minutes past midnight. The timestamp for the event log was derived by concatenating the ‘date’ field and ‘time1’ field to form a new field called ‘DateIn’ field. (*Note: ‘DateIn’ field is the same as the ‘timestamp for admission’ in the bare minimum event log for patient journey modelling shown in Table 3*). The ‘DateOut’ field was formed by concatenating ‘date’ field and ‘time2’ field. Prior to concatenating the fields, the ‘date’ field had to be converted into Date/Time data type with a 24 hour notation. For example:

Format ([time1], "General Number")

After calculating a timestamp from ‘time1’ and ‘time2’, these values were concatenated with the date value as below:

DateIn: [date] & " " & [time1] → *Timestamp admission.*

DateOut: [date] & " " & [time2] → *Timestamp of discharge or timestamp at the end of 24 hours.*

The concatenated fields were then converted into date/time data type. Accurately deriving the timestamp is essential for an accurate discovery of process model using the process mining tool, ProM. Each patient is uniquely identified by the ‘URN’ field in both datasets. The data type for this field had to be converted to the same data type in order to merge the records. In the patient journey dataset, each admission is identified with a unique ‘journey id’. The ‘journey id’ is unique for a particular admission and stays the same until discharge. Multiple admissions by the same patient will have multiple ‘journey ids’. As a result, when merging the two datasets, two fields had to be used as identifying keys. The first key is the patient ‘URN’ in both datasets. The second key is the derived ‘DateIn’ field from the patient journey dataset and the converted ‘Outcome Date’ field from the ED dataset. The ‘Outcome Date’ is the date when the patient is either admitted as an inpatient or discharged from the ED. If the patient is admitted, the patient’s record will be recorded in the patient journey dataset.

The ED dataset contained three date and time fields: 'Triage Date', 'Date Time Seen' and 'Outcome Date'. All fields had to be converted to the date/time data type. The ‘Triage Date’ is the timestamp when a patient enters the ED and is triaged according to treatment priority. The ‘Date Time Seen’ field is when the patient is seen by a doctor. All three timestamps are essential components in depicting the process that takes place in the ED. Each tuple contains information about an event. The snippet of the derived event log is shown in Table 4.

Table 4: Snippet of the derived event log with plug-in for Chapter 8

journey_id	urn	Priority	DateIn	DateOut	WardDerived	UnitDerivedFinal
1	23	3	27/01/2004 9:50	27/01/2004 10:26	FMC-WTS	TBA
1	23	3	27/01/2004 10:26	27/01/2004 15:44	FMC-RT	TBA
1	23	3	27/01/2004 14:12	27/01/2004 19:30	FMC-Boarding	CARD
1	23	3	27/01/2004 19:30	9/02/2004 16:46	6D	CARD
1	23	3	2/02/2004 13:45	2/02/2004 16:12	ANG	CARD
3	56	4	1/05/2006 19:57	1/05/2006 21:13	FMC-WTS	RESP
3	56	4	1/05/2006 21:13	1/05/2006 22:52	FMC-RT	RESP
3	56	4	1/05/2006 22:08	1/05/2006 23:05	FMC-Boarding	RESP
3	56	4	1/05/2006 23:05	2/05/2006 20:17	5A	RESP

Once a systematic method is established to generate the event log of a process of interest, attributes for cases and activities could be easily extracted for further analysis of cases. An event log contains millions of records, therefore for meaningful analysis it will be necessary to filter the event log according to a specific scope or boundary to produce models that are interpretable.

Table 5: Snippet of the derived event log with plug-in for Chapter 5, Chapter 6 and Chapter 7

journey id	urn	DateIn	DateOut	DaysAtHosp	age	charlindex	nature_of_separation	inlier time	outlier time	Total LOS (Hours)	"FMC" Time (Hours)
7	1743	24/02/2007	22/03/2007	26	80	9	"1"	581	42	623	42
10	4305	16/05/2007	17/05/2007	1	90	0	"1"	15	4	19	4
21	14689	6/02/2003	22/02/2003	16	86	2	"7"	360	32	391	6
26	41644	13/02/2003	18/02/2003	5	72	2	"1"	0	120	120	1
86	62399	10/12/2005	15/12/2005	5	84	2	"1"	115	1	116	1
107	72604	23/06/2004	24/06/2004	1	85	1	"1"	26	0	26	0
116	27894	19/09/2008	20/09/2008	1	85	0	"1"	17	1	18	1
117	31870	13/03/2007	20/03/2007	7	86	0	"1"	148	2	150	2
120	31860	3/08/2008	6/08/2008	3	88	2	"1"	44	23	67	23

Table 5 shows another snippet of the derived event log plug-in. From the derived event log, aggregate data were produced to undertake statistical analysis. The attributes used are further described in the Appendix D

Data Dictionary. The derived event log has to be further processed to take advantage of the open source ProM toolkit.

4.3.3 Further processing of the derived event logs for process mining with ProM

The main software used for Process Mining is ProM. ProM is open-source specialised process mining software. The derived event log was pre-processed into the MXML format required as input to ProM. This conversion was done using the Disco software package Fluxicon (2012) another open-source software. The conversion into MXML format is straight forward as such not discussed further.

4.3.4 Create a small sub-set of data

After deriving the event log and converting it to the MXML format, it was necessary to work with a small sub-set of the data in ProM to understand and follow the various concepts and algorithms used in ProM. A smaller data set makes it easier to manually calculate certain output or outcome when using ProM.

Finally after the data are checked for quality and the output of ProM complies with the expected output as described in ProM documentation, the log is ready to be used as input event log for process mining.

4.4 Results

The following sections presents some of the process mining results to demonstrate a tangible outcome addressing the research questions identified in 4.3.1.

Performance Sequence Diagram within ProM was used to discover the most frequent path used by GM patients and the throughput for the frequent path. In total 2542 ward movement patterns were discovered. However majority of these ward movement patterns had the frequency of occurrence of 1. This shows the diversity of patient journeys within the GM service. The most frequent path with the shortest LOS was for patients who were admitted to ward 3D as inliers. These inlier patients admitted to ward 3D were further clustered into two groups according to the time they spent as outliers in ward FMC (which is the ED itself waiting for an inpatient bed to become available). The two clusters of patients had an average LOS of 0.82 days or 19.67 hours and 7.85 days. According to the clinicians, this level of detail was unnecessary because it created greater complexity when analysing a large cohort of patients. Further investigation using organisational mining within ProM confirmed that ward 3D only admitted inlier patients. Figure 4-3 shows part of the pattern diagram and the average LOS for each ward-based on whether the wards treated inlier or outlier patients. Average LOS or throughput time as indicated in the diagram, cannot be equated to efficiency of care because each ward is treating patients with a variety of conditions. Some wards might have patients whose clinical condition demands they stay in the ward longer than patients in other wards.

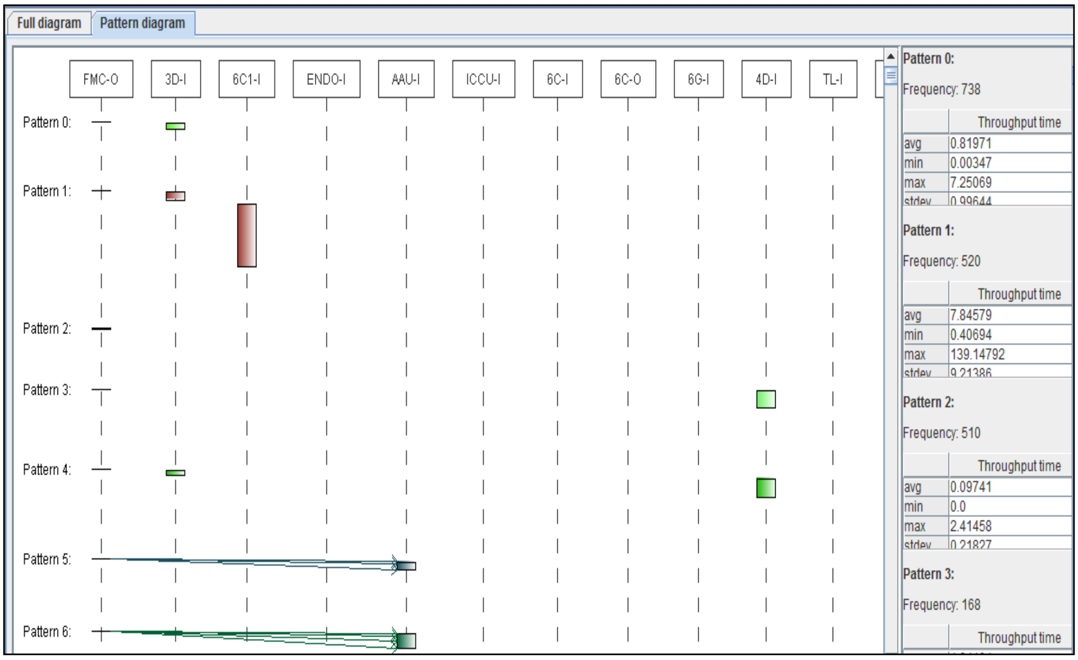


Figure 4-3: Performance sequence diagram

Figure 4-4 shows part of the output from organisational mining. The organisational mining portrayed all the GM units that admit patients in ward 3D. The clinicians were not interested in this output as generally the clinicians are aware of the wards that their patients are admitted.

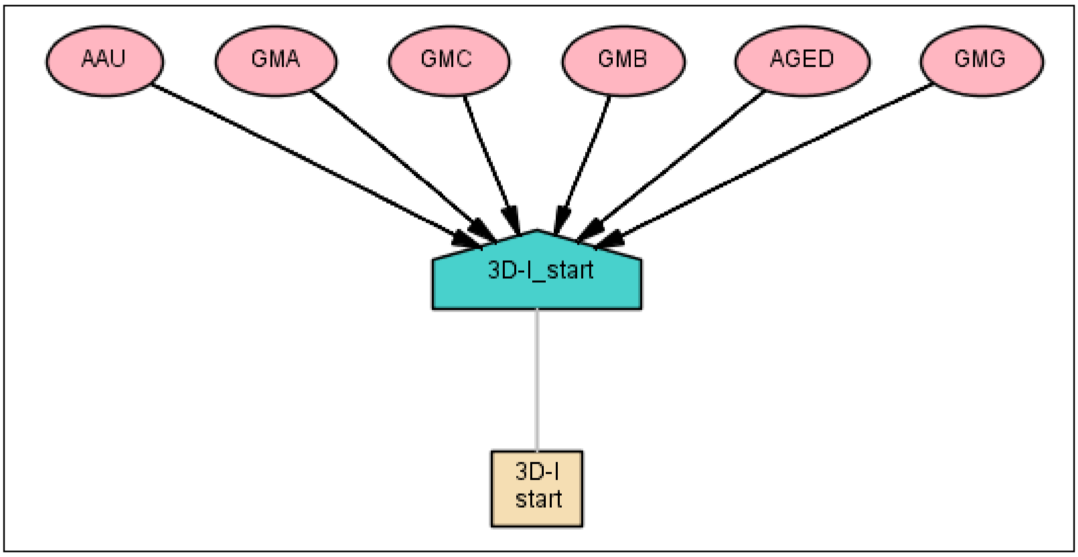


Figure 4-4: Organisational mining

The process mining steps can proceed in many different ways to fulfil knowledge discovery that will aid in giving insight to the process being investigated. For example in this case study the patterns could be analysed by grouping the most frequent pattern according to LOS

KPIs. Another approach is to look at all wards and their corresponding throughput time according to whether the ward accommodated either inlier patients or outlier patients only and the throughput of those wards that accommodated both inlier and outlier patients at the same time. The approach presented in this case study looked at the inpatient LOS of five wards with the highest frequency of patient journeys that treated inlier patients exclusively, five wards with the highest frequency of patient journeys that treated outlier patients exclusively and wards that treated both inlier and outlier patients. Table 6 shows the results of five wards with highest frequency of inliers, five wards with highest frequency of outliers and their throughput time. It also has the results for ward FMC (admitted patients within the ED). All admitted patients in the ED waiting for a bed are considered outliers. Ward 3D as previously mentioned treats the majority of the inliers followed by ward AAU. These two wards have the average throughput of 0.81 days for ward 3D and 1.72 days for ward AAU. As for wards that only treated outliers, apart from ward FMC, ward 5E had the shortest LOS followed by ward 6D and 5A. Ward inliers had shorter inpatient LOS when admitted to their home ward compared to ward outliers. However, a ward could admit both inlier and outlier patients; but this information is not accounted in this output. For example, if a ward admits GM patients that can equally well be classified as inliers or outliers, the LOS of an inlying GM in that ward is not taken into account in this output Therefore the average throughput time shown is not a true reflection of the LOS of the GM patients whether their status be inlier or outlier. A useful insight given by the clinician based on the information presented in is; Ward 3D treats and discharges a GM patient efficiently whereas a GM patient in Ward 5C has a prolonged LOS. Ward FMC (the ED) which had the highest number of patients treated as outliers, had the shortest average LOS of 0.06 days. These presumably are patients who had been admitted to the hospital and had waited in the ED for an inpatient bed but their condition might have improved therefore they were discharged from the ED before an inpatient bed became available. This conforms to the clinicians' experiences that patients discharged from the ED directly should not be included in the sample of patients during inlier or outlier LOS study because these patients will confound the findings.

Table 6: Wards treating inlier and outlier patients exclusively

Ward Name	Inlier/Outlier status	No. of journeys	Average throughput (Days)
3D	Inlier	2490	0.81
AAU	Inlier	1535	1.72
CCMU	Inleir	101	3.70
5F	Inlier	56	2.00
SSW	Inlier	37	1.90
FMC (ED)	Outlier	6550	0.06
6D	Outlier	162	2.24
5A	Outlier	75	2.25
5E	Outlier	57	1.72
5C	Outlier	49	3.58
5B	Outlier	43	2.74

Observations for the five wards with the highest frequency of patient journeys that accommodated both ward inliers and ward outliers, showed that ward 4D had the shortest LOS of 1.91 days for inliers and 1.63 days for outliers respectively. Once again this output did not take into account LOS of patients who were admitted in a ward that only treats GM patients as inliers. Therefore, another method is needed to account for whether the LOS of a GM patient differs in a particular ward according to whether that patient is classified as an inlier or an outlier. In general ward outliers in these wards were discharged quicker than ward inliers. Table 7 shows the results of the five top wards with the highest frequency of journeys and the average LOS for the inliers and outliers respectively.

Table 7: Wards treating both inlier and outlier patients

Ward Name	Inlier/Outlier status	No. of journeys	Average throughput (Days)
6C1	Inlier	766	6.71
6C1	Outlier	244	4.39
6A	Inlier	549	8.67
6A	Outlier	256	2.54
6B	Inlier	455	8.35
6B	Outlier	210	4.91
4D	Inlier	401	1.91
4D	Outlier	30	1.63

Another dimension to this analysis was whether the shorter inpatient LOS for ward outliers in wards that accommodated both inliers and outliers was due to patient mortality during admission. The most common nature of separation (discharge reason) for these ward outliers was "discharged home" for 748 patient journeys followed by "discharged to nursing home or hostel" for 95 patient journeys with average LOS of 2.8 days and 11.11 days respectively. The number of outlier patients who died in the hospital and did not require autopsy was 43 patients with an average LOS of 4.84 days. Outlier patients who died in hospital and needed autopsy accounted for only one patient with a LOS of 1.41 days. Looking at ward 4D which had the shortest average LOS, for both inliers and outliers revealed that the most frequent reason for discharge was "discharged home" with 421 patient journeys followed by "other hospital – up transfer" with 12 patient journeys with average inpatient LOS of 1.96 days and 3.48 days respectively. Three patients died in this ward with an average inpatient LOS of 0.14 days.

Organisational mining for those wards that accommodated both inliers and outliers showed that ward 4D was the homeward for unit GMD only and accepted outlier patients from all other GM units except from the AGED unit. Organizational mining also showed that ward 6B which had the highest average throughput of 8.35 days was a home ward for GMA and accepted outlier patients from all other GM units except from GMD unit. The inlier ward with the longest average LOS was 6A which is a home ward for the AGED unit which is often linked to extended inpatient LOS because that particular unit treated predominantly older patients with multiple comorbidities.

Social network analysis carried out for the GM patients clearly depicted the interaction of these units in regards to the transfer or handover of patients' care between these units. Figure 4-5 shows the sociogram derived for transfer of patients for these GM units. As mentioned previously, it is quite clear that the GMD unit accepts outlier patients from all other GM units except from the AGED unit.

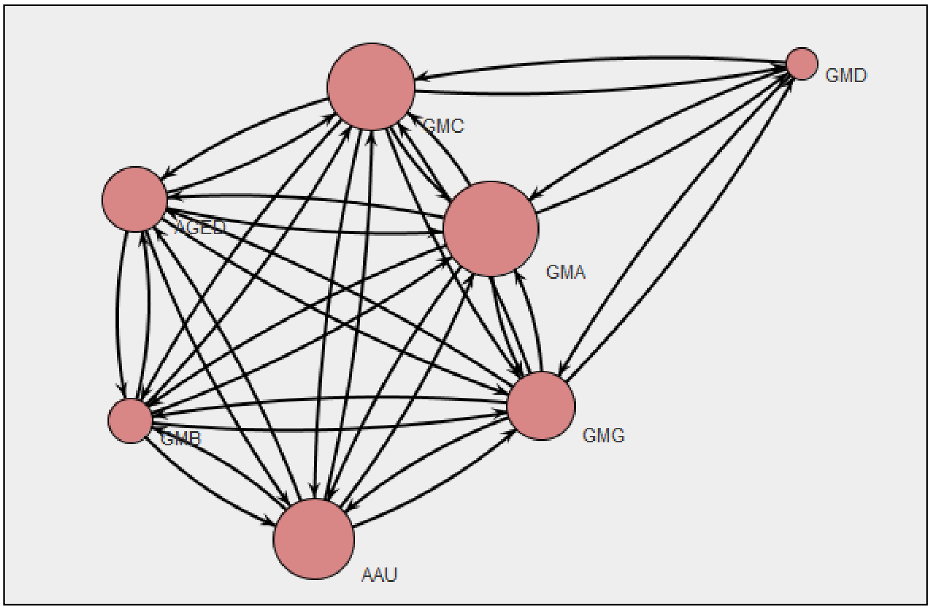


Figure 4-5: Sociogram for GM units showing unit interaction

The possibility of mining and drilling into various aspects of the patient journeys are limitless. This section only discussed a small aspect of the process mining activities of patient journeys in order to demonstrate the research undertaken using this methodology at FMC.

4.5 Discussion

The output also revealed that, for an accurate reflection of reality, the output should be interpreted by clinicians who can give proper insight. One such example is that one of the wards shown in the output as a separate entity became the ED eventually but the model did not depict this. This is because the data set used spanned six years. During this period of time many changes in ward ownership took place. The relative LOS of inliers versus outliers changed according to whether GM patients on those wards were always classified as inliers, always as outliers or as either an inlier or an outlier depending upon the particular GM unit responsible for the patient. However, the average LOS produced is not a true reflection of the LOS of the inlier or outlier population because the LOS of GM patients in a ward that admits GM patients as inliers may differ from the LOS of inlying GM patients admitted to a ward where GM patients can have either inlier or outlier status.

4.6 Conclusions

Gaining insight from un-structured or semi-structured health data is a challenging task which calls for systematic methods in deriving event logs from various health information systems. Deriving event logs is far from a trivial task. It also requires a systematic approach. The un-structured and semi-structured nature of health data makes process mining using ProM a viable approach for knowledge discovery for the purposes of gaining insight into the complex health processes. Such knowledge discovery will assist in better decision making.

As with any knowledge discovery, meaningful insight from data has to be supported by the knowledge offered by the clinicians in order to understand the intricacies behind the complex healthcare decision making processes.

Healthcare analytics are ingrained with extensive hypothesis studies which are different from the approach of process mining. However, clinicians are used to this form of reporting and healthcare KPIs are also governed by statistical reporting. Therefore process mining in healthcare will naturally be accompanied by statistical analysis of some sort especially reporting the p-values (significance tests). Any of the results discussed in the case studies above could be easily extended and reported with statistical testing. Future development in process mining tools should look into the feasibility of easy exporting of data to be used with major statistical software. However, the framework proposed for the derived event log and the generation of clinically-relevant outcomes enables the derived event log to be easily aggregated for further analysis using statistical software. This capability is demonstrated in the studies undertaken in the following chapters.

5 The relationship between in-hospital location and outcomes of care in patients of a large general medical service

This chapter discusses the core of the work done in assessing the outcomes of patients admitted as ward inliers and ward outliers. This work was accepted for publication by the Internal Medicine Journal of the Royal Australasian College of Physicians (Perimal-Lewis et al., 2012a).

5.1 Introduction

Building on the concept of the process-oriented methodology using the process mining framework as discussed in Chapter 3 and the concept of the derived event log as discussed in Chapter 4, the patient journey process is classified as a key process of the hospital. A single patient journey within this process is classified as a process instance. This understanding facilitated further exploration of the patient journey process from a case perspective. In a case perspective, the characteristics of the process instances (*single patient journeys*) are of interest. Building on the knowledge of the movement of patients through the hospital system as discussed in Chapter 3, in this Chapter details of each patient's movements are further explored. An admitted patient could move from a team-based model of care to a ward-based model of care and vice versa. Figure 5-1 shows the patient journey as a key process for the hospital and the adaptation of the process mining framework to investigate patient movements between these models of care using a process-oriented methodology approach. This understanding formed the basis for the structure of the derived event log which was used in this work.

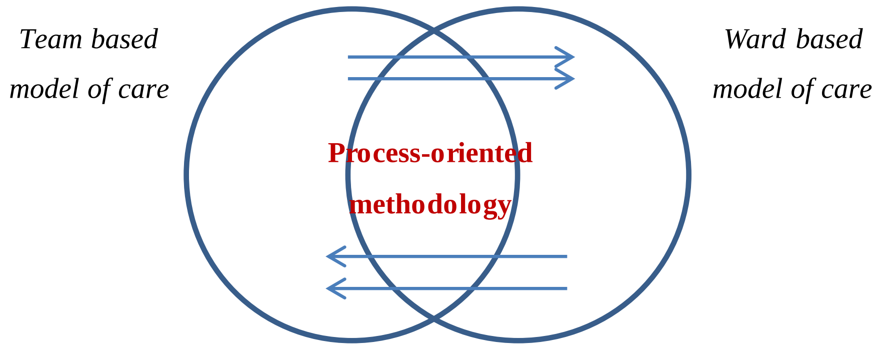


Figure 5-1: Patient journey process

In a team-based model of care, one team of doctors (one unit) is responsible for the care of the patient from admission to discharge. In a ward-based model of care, it depends upon the type of ward to which the patient is admitted that determines the nursing team responsible for looking after the patient.

Some patients could receive 100% team-based care. This could happen when a patient's care has been the responsibility of one team of doctors (one unit) from admission to discharge. Some patients could potentially receive 100% ward-based care. This could happen when the patient's medical condition demands that the patient is admitted to a particular ward. The doctors and the nurses in that particular ward are specialised in the delivery of the right care needed to treat the medical condition of the admitted patient. This type of ward-based care is usually limited to patients suffering from a stroke (Stroke Unit), a heart attack (Coronary Care Unit) or the patient requires resuscitation (Emergency Department and Intensive Care Unit). Generally these patients spend only a period of their admission in that ward-based care environment and the rest of their admission in a team-based care environment.

If the relevant team of doctors (unit) only have a few inpatients and a large ward; the patient can enjoy both ward-based and team-based models of care in their home ward anywhere in the hospital. It is only when the number of inpatients allocated to a team of doctors exceeds the number of their ward beds that a team-based rather than a ward-based model of care operates and as a result the excess of patients have to be admitted to any available ward. Many patients experience both team-based care and ward-based care during their admission.

Figure 5-2 shows a typical high level interaction that takes place in a ward. As previously discussed in Chapter 3, there are generally two types of hospital wards: medical and surgical wards. Each unit in the hospital is allocated a number of beds in an appropriate ward. Therefore, each unit has at least one home ward where patients admitted under their care should be placed.

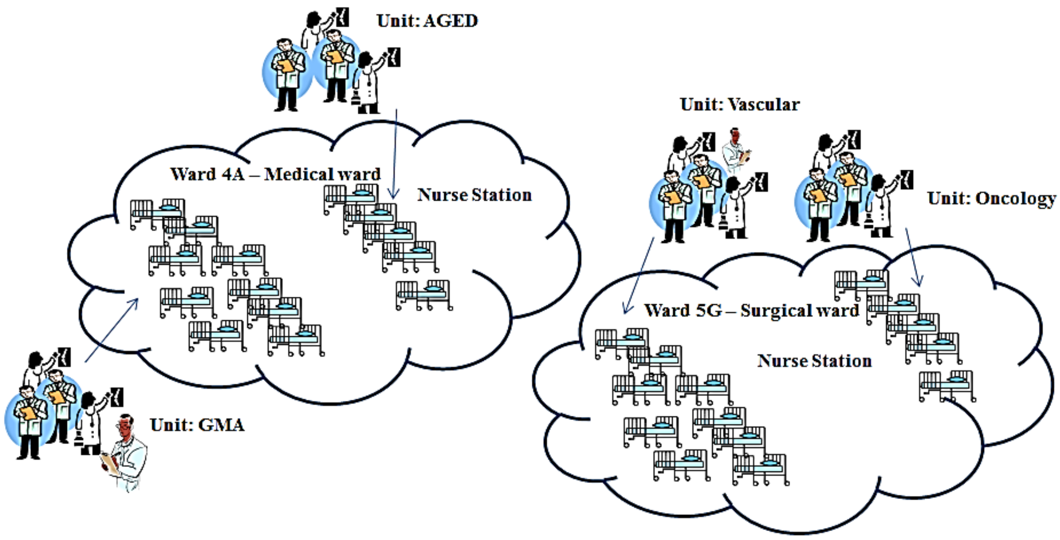


Figure 5-2: Typical high level representation of hospital ward

The home ward of a particular patient is defined as the ward where the multidisciplinary team responsible for their care is located. Each team of doctors (unit) is allocated a finite number of beds. These home wards are generally co-located or local to the workspace of the unit. This co-location facilitates the delivery of efficient and prompt patient care. These home wards are staffed with specialised nursing teams. The doctors and the nurses in home wards usually have an efficient working relationship. The team of doctors and nurses is also conversant with the processes and workings of their ward. Patients admitted to their home ward therefore are expected to receive optimal care. This is because the resources involved in the operation of home wards should combine efficiently. Patients admitted to home wards are referred to as inliers.

The number of beds allocated to each unit is finite; therefore some patients may be admitted to wards distant to the home wards of the team of doctors responsible for their care. These patients are referred to as outliers. Outliers are perceived to receive inferior QoC compared to inliers. This view stems from a belief that often the medical and nursing teams responsible for that patient's care are not co-located and doctor – nurse communication might be inefficient. Accordingly the bed managers at each hospital usually try to source a home ward for every newly-admitted patient whenever possible. The number of ward outliers is one of the measured KPIs for most public hospitals. Figure 5-3 represents the process of ward allocation for an admitted patient.

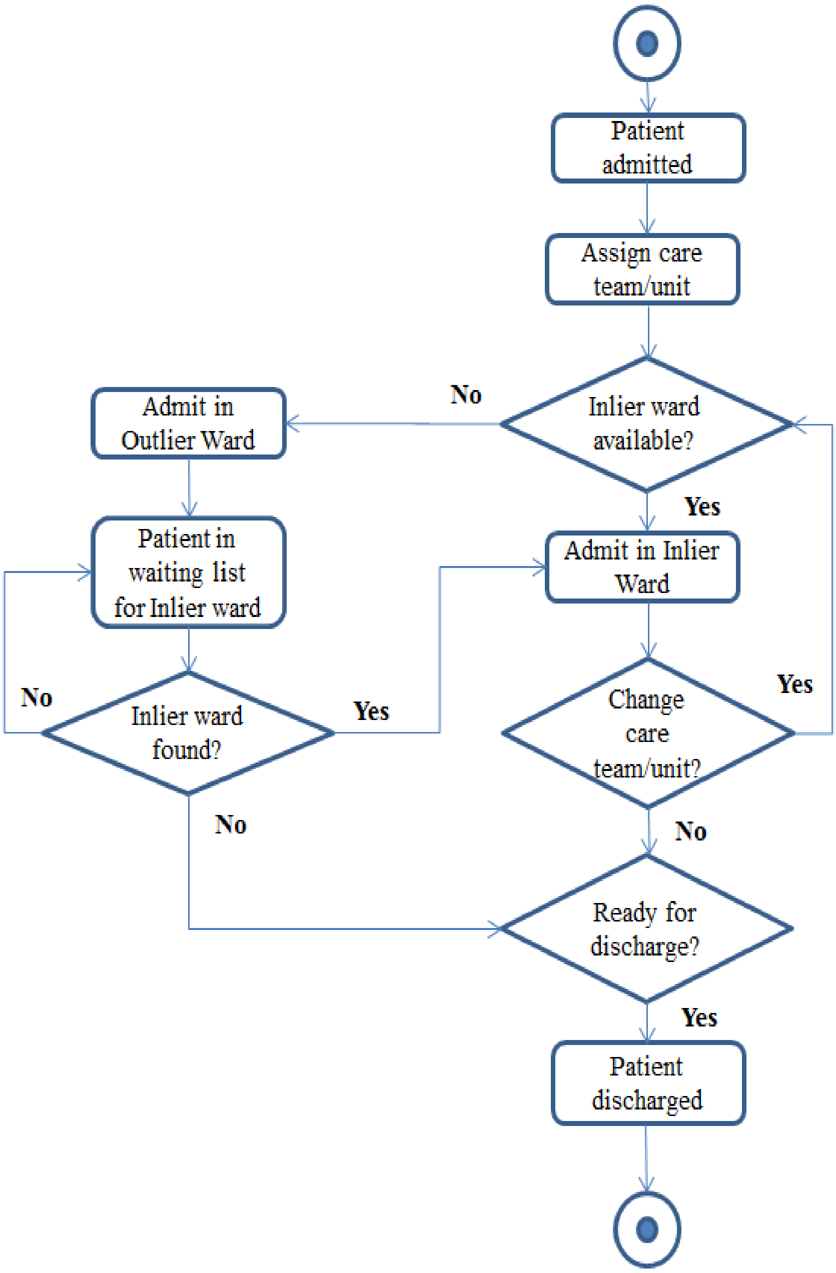


Figure 5-3: Flow chart representing inlier / outlier ward allocation

This complex interaction between the team-based model of care, the ward-based model and the concerns of outliers receiving an inferior QoC are intensified as bed availability diminishes in public hospitals. It is commonplace for Australian general hospitals to work at high levels of bed occupancy. The medical and surgical wards of many of hospitals are often between 95% and 100% occupied. In consequence, it is often difficult to place patients in the home wards of their clinical teams (units). A recently-admitted patient, especially one admitted as an emergency, may have to spend a significant period of their admission as an outlier because a bed is not immediately available in the preferred location. Having groups of patients who are outliers can disrupt both team-based and ward-based models of care, and fragment the care provided for the patients by medical, nursing and allied health clinicians (Ben-Tovim et al., 2008a). Medication errors, for example, are higher in patients housed in an outlier ward (Warne et al., 2010) .

Government expectations that patients will spend less than four hours in the Emergency Department (ED) may put pressure on the management looking for any available bed for a recently-admitted patient and hence increase the frequency with which patients are placed in wards that are not the home wards of in-taking medical or surgical teams. More comprehensive studies are clearly warranted.

5.2 Research on ward outliers

Concerns about outliers remain largely anecdotal because there are only a limited number of scholarly studies on the effects of outlier status either on measures of efficiency of care such as hospital Length of Stay (LOS), or on measures of quality of care such as mortality risk, readmission rates and discharge summary completion rates. Patients with heart failure were reported to stay longer in hospital if they spent time in an outlier ward, though their mortality risk and readmission rates were unaffected (Alameda and Suárez, 2009). A short term observational study of an elderly care unit reported that outlying patients had a longer LOS and higher three month readmission rate than patients accommodated in their home ward despite similar mortality and co-morbidity rates (Xu et al., 2011). As mentioned earlier, acutely ill surgical patients admitted to a medical ward (and hence outliers) were more likely to miss medications (Warne et al., 2010). Santamaria et al. (2014) in a very recent observational study found a strong association between being an outlier and the number of emergency calls made on that patient suggesting that the therapeutic and monitoring needs of outlier patients may be compromised.

This is the first research in this area looking at a large General Medicine (GM) patient population of about 20,000. The strategy adopted was first to understand the process of ward allocation and the fact that the majority of patients move from outlier ward to inlier ward rather than vice versa (*insight gained from process mining discussed in the Chapter 4*). The inlier and outlier status definitions are variables that can only be defined after most careful consideration and this will be discussed in Section 5.4.1.

5.3 Aims

The primary aim of this study was to define the status of a patient as either a ward inlier or a ward outlier and then contrast certain outcomes of their care during and after admission, taking into account a range of clinical characteristics of the patients.

Understanding the consequence of being an inlier or an outlier would inform hospital bed managers whether it is important to hold patients in the ED congesting the ED in an effort to find the right ward for each patient.

5.4 Methods

Using the derived event log discussed in Chapter 4, the GM units' patients at FMC admitted between 1st January 2003 and 20th September 2009 were examined. Only those patients who remained under the care of the GM units throughout their entire stay were included in this study. If, for example, a patient was transferred to the care of the respiratory team they were excluded from the analysis. Similarly, patients who left the care of General Medicine after being transferred to the High Dependency Unit (HDU), the Intensive Care (ICU) or the Coronary Care Unit (CCU) were excluded. The Diagnostic Related Group (DRG) data were added as a plug-in to the main derived event log using patient's Unit Record Number (URN) and the 'Admission Date'.

The derived event log from the patient journey dataset from FMC contains information on inpatients or officially admitted patients only and records detailed information on the journey or movements of a patient from the time of admission to the time of discharge. An individual patient could have multiple admissions at different points in time and each admission will be allocated a unique journey number ('journey id') that remains 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 are known together with the name of the ward. Each ward occupied by a patient is appropriately marked to reflect whether the ward was an inlier or an outlier ward. A patient admitted to an inlier ward is admitted to their home ward. The timestamp for admission is the combination the 'Date' field and the 'Admission Time' field. The timestamp for discharge is the combination of 'Date' field and 'Discharge Time' field. The timestamp is a derived field. Individual patients are not identifiable at any point.

The original data set contained about 1.9 million records spanning January 2003 to September 2009. The final record set which was used for the analysis only consisted of journeys of patients who had been exclusively cared by the GM unit from admission to discharge. If a patient's journey was under the care of a combination of GM unit and non GM unit, the journey was excluded. This level of filtering reduced the record set to about 24, 439 patient journeys.

The patient journey database which was the source of the derived event log is a continuously updated extract from the hospital's basic patient admission and tracking databases. All patient movements between wards and between units are identified, time-stamped and recorded. Inlier or outlier status is automatically altered at the time of the patient movement or unit re-allocation.

The initial extract of 23,439 records from the derived event log was merged with data from other databases, most importantly, the hospital's ED computerised database. This produced the various variables needed for the analysis that were not captured in the patient journey database and gave a new sample size of 23,312. 674 (2.89%) journeys were deleted from the sample due to incomplete time values. The patient journeys were categorised into 'Journeys with LOS > 30 days', 'those discharged from ED' and 'All Other Journeys'. The Register of Births, Deaths and Marriages was accessed to identify those patients who died within 28 days of discharge. This information was de-identified and supplied by the hospital as a Boolean field. Therefore the final derived event log had a plug-in from the ED computerised database, plug-in from the Register of Births, Deaths and Marriages and a plug-in for the Diagnostic Related Group (DRG).

The GM unit comprises of a number of smaller short and long stay units each with a defined home ward location. The GM service controlled about 100 inpatient beds out of about 500 beds in FMC as a whole and typically cared for patients with complex multi-system pathology. The wards that were home wards for this service changed over six years but always were clearly and operationally defined throughout the period of observation. Patients were assigned to inlier or outlier status by the computerised bed management system of the hospital using instantaneously updated information pertaining to the allocated home wards of all inpatient units working within the hospital. If the patient was not treated within a home ward for the GM unit allocated to care for the patient, they were defined as being an outlier. As discussed previously the percentage of patients who were outliers was a regularly reported hospital indicator and considerable care was taken to keep the relevant data systems up to date.

5.4.1 Outlier / Inlier time definition

There are many interacting hospital system specific processes (*e.g. team-based, ward-based models of care*) that influence where the patient is admitted and the impact of this hospital location on the overall QoC for a patient. The explanation in previous paragraphs about the movement of patients between team-based and ward-based models of care and the implications of being admitted to a home ward or otherwise suggest that defining the outlier or the inlier population is not straightforward. Every hospital in principle follows a variation of the models of care described above but each hospital might have other system level interactions that could influence the number of wards or beds allocated to each team of doctors. The overall bed capacity of the hospital, the demography of the population the hospital serves and a range of funding models could all influence the number of wards and beds allocated to each unit. Patients may be located in a ward adjacent to their home ward and be cared for by appropriately-trained nurses or could be allocated to a bed in a distant part of the hospital and be cared for by nursing team unfamiliar with the patient’s illness. Seeking to define a black and white definition of the outlier status might not reflect the reality of the population being studied. On the other hand only investigating patients with 100% outlier time or outlier status and 100% inlier time or inlier status poses the risk of excluding a large part of the patient population who do not meet this criterion. Therefore, where to draw the line as to what makes an outlier or an inlier is negotiable. The preceding sections describe the strategy proposed and applied in this research.

The time an admitted patient spends in an outlier ward is part of their overall inpatient LOS. The proportion of inlier and outlier time for a specific population potentially fluctuates depending on the various hospital system level factors described above. Therefore, once the proportion of the inlier time and the outlier time has been established for each process instance (*individual patient journey*) the distribution of these times can be graphed and analysed in conjunction with insights from the clinicians. In this study, the inlier and the outlier time definition that was chosen was influenced by the desire to study the majority of the population.

The following section outlines the approach taken when classifying the GM population into either inlier or outlier status.

After filtering the derived event log to include the GM patient population and calculating their overall hospital LOS, the proportion of time each patient spent outside of their home ward *(% of outlier time*) was derived for each process instance (*patient journey*). Table 8 summarises the % of outlier time and the number of patients for each proportion.

Table 8: Percentage of outlier hours

Percentage of Outliers Hours		
% of Outliers Hours	Number of Patient Journeys	Percentage of Journey
0	529	2.26
>0 And <10	11412	48.69
>=10 And <20	2945	12.56
>=20 And <30	1459	6.22
>=30 And <40	857	3.66
>=40 And <50	597	2.55
>=50 And <60	446	1.9
>=60 And <70	337	1.44
>=70 And <80	268	1.14
>=80 And <90	255	1.09
>=90 And <100	222	0.95
100	4112	17.54

Figure 5-4 shows the distribution of the percentage of outlier hours for the GM population. It is a graphical representation of Table 8.

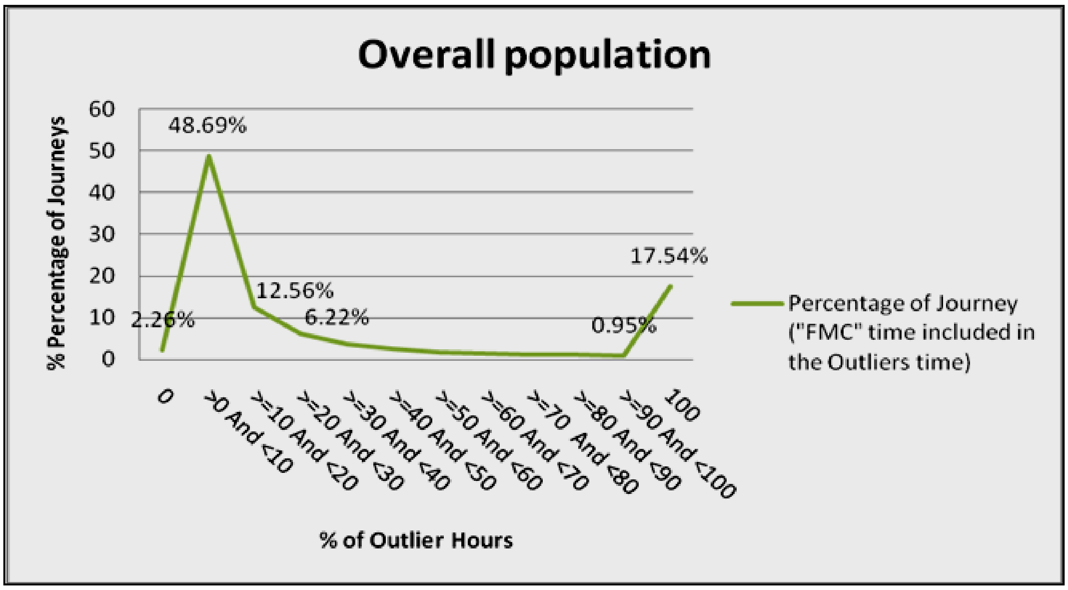


Figure 5-4: Distribution of the outlier hours for the GM population

After leaving the ED, most inpatients (51%) spent almost all of their admission in their home ward. Fewer (17%) spent all their admission in an outlier ward. The remainder were transferred between outlier and inlier locations during their admission. The cut-off point of outlier status was chosen according to where the patients spent the majority of their admission (≥70%). By choosing this cut-off point, most patients (~90%) could be allocated either to a group comprising those who spent ≥70% of their hospital stay in their home ward (inlier status) or a group comprising those who spent ≥70% of their admission outside of their home ward (outlier status).

The cut-off points for the inlier and the outlier status could be refined to >=80% but this would result in the loss of another 6% of the dataset. The cut-off point chosen should be determined by the intent of the study. As discussed previously, choosing 100% inlier and 100% outlier population could also be used as another strategy to classify this population but that would have meant that the majority of GM patients in this data set would have been excluded.

5.4.2 Exclusions

The collaborative nature of this research enabled the recognition of known confounders early in the research. The clinicians identified the known confounders and the reasons for excluding these confounders are discussed below. Outlier LOS in FMC may be affected by a number of administrative hospital processes, including:

1) A patient, having arrived at the ED, can be administratively admitted to the hospital but then discharged from the ED without ever entering the body of the hospital. Under those circumstances, the LOS of that patient is short but they will be viewed as an outlier because they never entered their home ward.

2) The longer a patient stays in hospital, the more likely that patient is to be moved into their home ward. After 30 days, further delay in discharge commonly relates to administrative concerns such as finding placement in an appropriate residential care facility (Victor et al., 2000). Under those circumstances, the placement problems confound the potential differences in LOS outcomes between inliers and outliers.

For these reasons, 2,086 inpatients discharged from the ED (*100% of this cohort were outliers*) and those 629 patients staying in hospital over 30 days (*84% of this cohort were inliers*) were excluded.

5.4.3 Diagnostic Related Group (DRG)

Primary diagnoses were coded at the level of the International Classification of Diseases-10 (ICD-10) and those diagnoses allocated to inliers and outliers were compared. All patients were assigned a Diagnostic Related Group (DRG) for their admission. The patients' DRGs were used to identify a DRG-based average LOS for each DRG, and these LOS formed the predicted LOS for each patient.

The principal admission diagnoses (*based upon ICD codes*) for all patients were assessed. This comprised over 1400 ICD codes. In consultation with the clinicians, the 100 most frequent ICD coded diagnoses were identified; capturing approximately 70% of the inlier and outlier patients. ICD-10 codes were similar between the inlier and the outlier groups. Table 9 summarises the condensed diagnoses and compares their frequencies for both inliers and outliers. The most frequent diagnostic categories were identical between the inlier and outlier groups suggesting that any difference between these two groups was not on the basis of a significant difference in principal diagnoses between the groups.

Table 9: Primary diagnosis for inliers and outliers

	Inliers		Outliers	
Respiratory (*including infection*)	2743	0.18	507	0.20
Cardiac	1634	0.11	319	0.12
Other infection	1106	0.07	186	0.07
Collapse/Hypotension	929	0.06	176	0.07
Renal	934	0.06	130	0.05
Other	899	0.06	98	0.04
Delerium and Dementia	656	0.04	85	0.03
Orthopaedic	577	0.04	52	0.02
Diabetes	355	0.02	56	0.02
Neurological	329	0.02	43	0.02
Vertigo and Nausea	276	0.02	38	0.01
Haematology	273	0.02	36	0.01
	10819	**0.71**	**1730**	**0.67**

5.4.4 Accounting for inlier / outlier population differences

The patient level factors present at the point of admission that are known to be both important and measurable, are a combination of principal and secondary diagnoses and factors such as age, gender and admission status (*emergency/elective*). Admission status is not relevant for this population because all patients were emergency admissions.

The inlier and the outlier LOS differences may be the result of population level differences where the outliers were drawn from a 'predicted' short stay population and inliers from a 'predicted' long stay population. To address this, the predicted LOS data for both groups were analysed.

The DRG system provided an average, or predicted LOS value for each DRG which was used to review the predicted LOS for both the inlier and the outlier group. Figure 5-5 and Figure 5-6 show the proportions of the inlier and the outlier patients in LOS groups, using one-day increments. The histograms show very similar distributions for both groups.

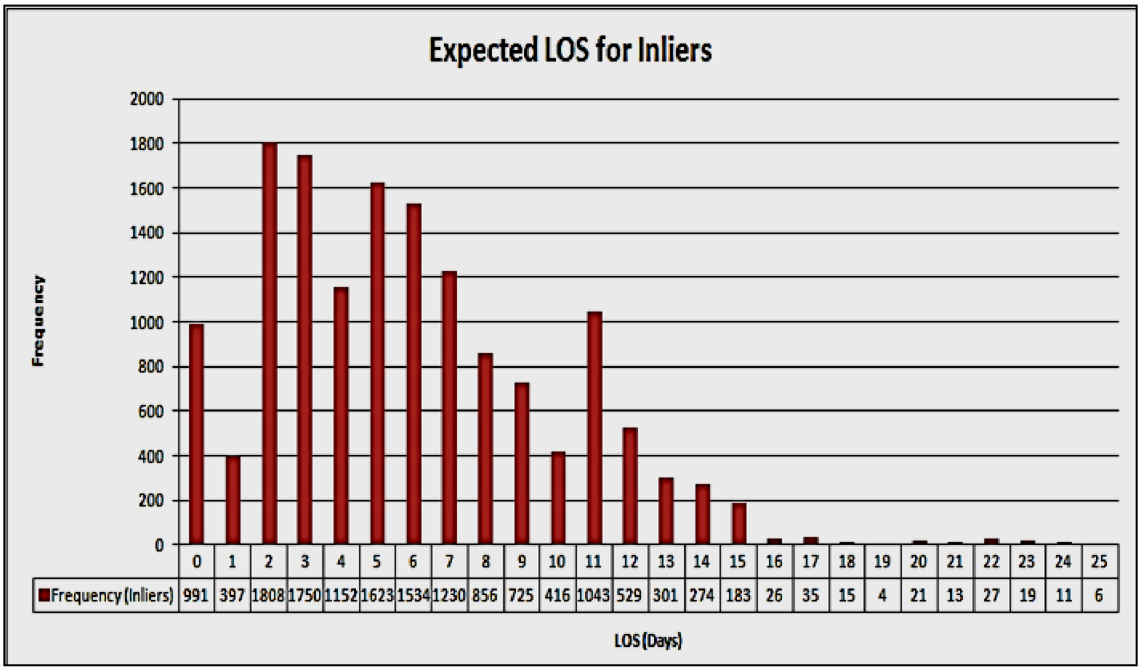


Figure 5-5: Expected LOS for inliers

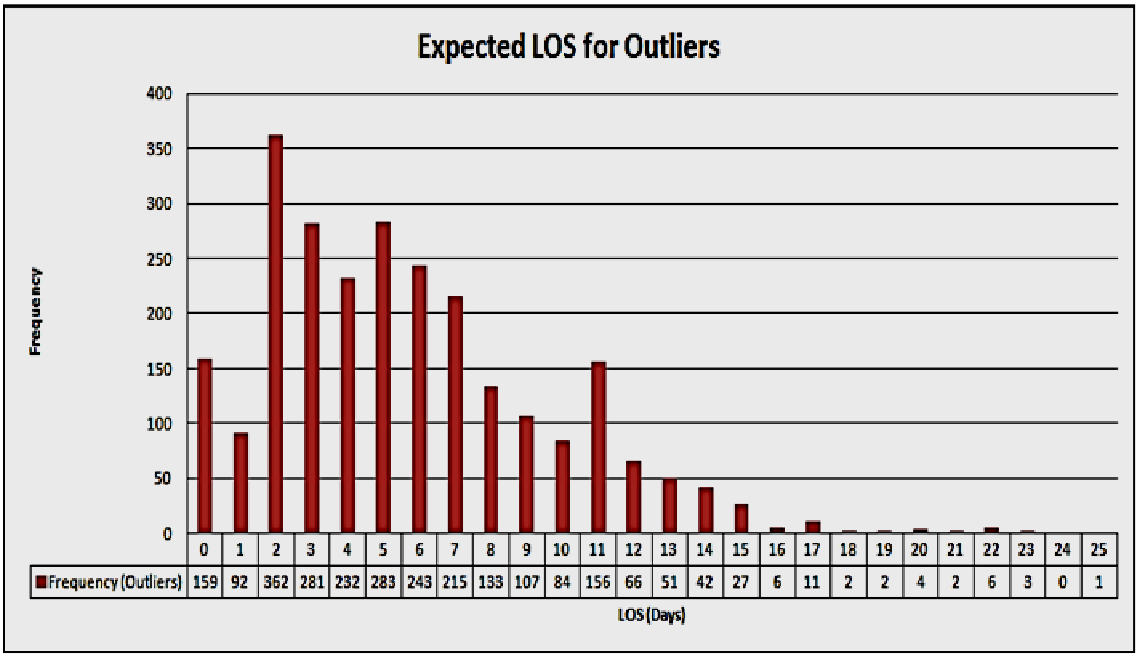


Figure 5-6: Expected LOS for outliers

The predicted LOS for both inliers and the outlier are presented in Table 10 **and** Table 11. The median LOS was 5 days for each group.

Table 10: Predicted LOS for inliers

Inliers	
Sample Size	15013
Mean	5.98
SD	4.12
Median	5
Mode	2

Table 11: Predicted LOS for outliers

Outliers	
Sample Size	2570
Mean	5.72
SD	3.98
Median	5
Mode	2

The histograms and the two tables suggest that the inlier and outlier groups were drawn from similar populations in relation to predicted inpatient LOS.

5.4.5 Statistical analysis

Statistical analysis was performed using Stata® 12.2. Poisson regression analysis was used to assess the effect of outlier status upon mortality, on LOS, on readmission rate and on discharge summary completion adjusting for the patients' age, Charlson co-morbidity index Charlson et al. (1987) as modified by Quan and colleagues (Quan et al., 2005), gender and length of time spent waiting for a bed in the ED for all patients.

The Poisson regression analysis using Stata® 12.2 was undertaken by Mr Paul Hakendorf from FMC's Clinical Epidemiology Unit based on the supplied aggregate data computed from the derived event log.

Patients' LOS in hospital, their risk of in-hospital death and their risk of readmission or death within four weeks of discharge were also assessed relative to the proportion of time each patient spent in their home ward during their index admission.

Information on the patients' age, the time they spent in the ED awaiting a bed, and their co-morbidities, were used to compute another co-morbidity index which was used to control for confounding factors known to affect inpatient LOS.

5.5 Results

After exclusions, there were 19,923 patients of whom 15,213 were classified as inliers, 2,592 were classified as outliers and 2,118 who did not fit into either category.

The characteristics of the inliers and the outliers are presented in Table 13, which also includes the univariate comparisons where relevant. The risk of readmission within seven or twenty eight days was substantially lower in the outlier group. Discharge summaries were less likely to be completed in the outlier population.

On the basis of their DRG, the predicted LOS for each patient was calculated and also expressed relative to their actual LOS. The DRG-based mean predicted LOS for the inlier and the outlier groups were very similar (*5.98 days, SD 4.12 & 5.72 days SD 3.98 respectively*) and the median LOS was five days for both. The actual LOS of outliers was about one day shorter than predicted but actual LOS was similar to predicted in inliers.

Outlier status was a significant predictor of increased in-hospital mortality (*Relative Risk of 1.41; 95% CI 1.16-1.73; $p = 0.001$*) and of reduced LOS (*LOS reduction to 0.77; 95% CI 0.74-0.81; $p < 0.001$*). The likelihood of discharge summary completion within two days was lower for outliers (*Relative Risk of: 0.66; 95% CI 0.62-0.71; $p < 0.001$*). The risk of readmission within seven days was not significantly affected by the outlier status, but the risk of readmission for outliers within twenty eight days of discharge was significantly less than for inliers (*Relative Risk 0.68, 95% CI 0.52 – 0.89; $p = 0.004$*).

The diagnostic profiles of patients who died in hospital were similar for inliers and outliers and not influenced by their outlier or inlier status.

Table 12: Characteristics of excluded patients

Characteristics of excluded patients	
Discharged from the Emergency Department (n=2086)	
Age, years (SD)	69.43 (20.64)
Charlson Index (SD)	0.90 (1.54)
In hospital mortality, n (%)	63 (3.02)
LOS > 30 Days (n=629)	
Age, years (SD)	79.04 (12.35)
Charlson Index (SD)	2.45 (2.57)
In hospital mortality, n (%)	39 (6.20)

The characteristics of inpatients either admitted to GM but discharged from the ED and those admitted to GM but discharged at least 30 days after admission. Where appropriate, data are expressed as mean (SD)

Table 13: Characteristics and outcomes of inlier and outlier patients

Characteristics and outcomes of inlier and outlier patients			
	≥70% Outlier Hrs (n=2592)	≥ 70% Inlier Hrs (n=15213)	P-value
Age, years (SD)	69.7 (19.1)	72.7 (17.2)	<0.001
Charlson Index (SD)	1.5 (1.9)	1.5 (1.9)	0.468
Time spent in the ED, hrs (SD)	6.3 (7.2)	5.3 (5.63)	<0.001
Total in-hospital LOS, hrs (SD)	110.8 (113.3)	141.9 (139.3)	<0.001
In-hospital mortality, n (%)	117 (4.5)	537 (3.5)	0.014
In-hospital mortality within 48 hours of admission, n (%)	59 (50.4)	120 (22.4)	<0.001
Mortality within 28 days of discharge, n (%)	179 (6.9)	953 (6.3)	0.210
Readmitted within 7 days, n (%)	30 (1.2)	308 (2.0)	0.003
Readmitted within 28 days, n (%)	53 (2.1)	746 (4.9)	<0.001
Discharge Summary sent within 2 days of discharge, n (%)	1055 (40.7)	9317 (61.2)	<0.001
Discharge Summary sent within 7 days of discharge, n (%)	1666 (64.3)	11873 (78.0)	<0.001
* Respiratory disease (incl. asthma, infection), n (%)	507 (20)	2743 (18)	
* Cardiac disease (incl. failure, infarction), n (%)	319 (12)	1634 (11)	
* Other sepsis (incl. urinary, cellulitis), n (%)	186 (7)	1106 (7)	
* Collapse/hypotension, (%)	176 (7)	929 (6)	
* Renal disease (incl. failure, nephritis), (%)	130 (5)	934 (6)	

The characteristics and outcomes of patients admitted to general medicine who spent their inpatient time mostly as an inlier or as an outlier. Where appropriate, data are expressed as mean (SD). Chi-square tests were used to compare categorical variables, and t tests for continuous variables.

** Primary diagnosis.*

5.6 Discussion

A tertiary hospital's GM unit embraces a wide variety of patient illnesses and a range of illness acuity, disease complexity and patient ages. This research demonstrates that, over a period of more than six years, the location of care of a GM patient in the hospital carried implications separate to the complexity of their illness and their age.

Unlike findings from earlier, more limited and specialised studies by Xu et al. (2011), the FMC LOS was shorter in outlying patients suggesting efficiency of care was not compromised by having patients housed in outlying wards. Like others (Kossovsky et al. 2002, Thomas et al. 1997), a shorter LOS does not necessarily equate to improved quality of patient care. Readmission within a week is more likely to reflect quality of care and this was not compromised by outlier status. Interestingly, the risk of readmission within 28 days was lower for an outlying patient even after in-hospital deaths were excluded from the analysis. The reduced LOS for outliers was unexpected and occurred despite the exclusion of patients whose LOS was predictably affected by non-clinical processes associated with the allocation of inlier or outlier status. The one day difference between predicted and actual LOS in the outlier and not in the inlier patients suggests outlier ward location affected LOS irrespective of the principal diagnosis of the patient.

The inliers and outliers have differing mean ages but similar predicted LOS, Charlson indices and diagnoses. There was an increase of over 40% in mortality risk in those who were outliers and they were more likely to die within forty eight hours of admission. Neither the clinical condition nor the medical, surgical or intensive care nature of the ward explained these consequences of being an outlier. It remains possible that allocation to an outlier ward was a consequence of importance being given to providing single room accommodation for patients for whom conservative rather than active treatment was deemed appropriate. There was some support for this possibility in a limited review of the discharge summaries (*undertaken by the hospital*) of a sample of patients who had died, but a prospective study is required to properly examine this possibility.

Patient age and co-morbidity separately impose significant effects upon in-hospital mortality, LOS and risk of readmission (Librero et al., 1999). After controlling for these confounders, the analysis confirmed that outlier status remained a significant predictor of both increased inpatient death and reduced inpatient LOS. The risk of death after discharge was unaffected by the patient having been an outlier. Discharge summary completion was slower for the outlying patients resulting in delayed communication between the hospital medical team and clinicians in the community.

This research is retrospective and observational, and has relied on information available in hospital administrative data sets. Most importantly, it is derived from one hospital and one clinical unit, albeit a large unit within that hospital. Results from this setting may not be generalizable, though the use of one service only minimises variation due to differences in medical management and casemix in relation to inlier and outlier status.

The work clearly needs replication elsewhere, with an expanded patient population. Ethical and practical issues preclude a randomised intervention trial of the study of outlier status, but a prospective observational study would allow collection of robust clinical data currently not accessible from the administrative datasets. Issues that could be studied in a prospective study include the range, appropriateness and impact on LOS of investigations performed during hospital stays, and the impact of decision making in relation to issues such as any requirement for single room accommodation. The definition of outlier could also be altered to reflect broader streams of care – medical patients in surgical wards and vice versa.

5.7 Conclusion

This research looked at one aspect of the impact of the organisation of care on the quality of care provided to GM patients in a busy general hospital. The location of a patient's care appears to have a substantial impact on the outcomes of care provided. An outlier medical patient being nursed in a surgical ward may have different outcomes to an outlier medical patient being nursed in a medical ward. This should be grounds for future prospective research. Hospital LOS was significantly reduced for patients who were ward outliers and their risk of readmission within 28 days was also reduced but the in-hospital mortality was significantly increased. Further research is needed to identify those factors that are influencing LOS to ensure that any organisational benefits that might be produced from any reduced LOS are not being obtained at the expense of less than optimal care.

6 Analysing homogenous patient journeys to assess quality of care for patients admitted outside of their 'home ward'

This chapter extends the research done on ward inliers and ward outliers completed in Chapter 5. The case perspective of the process mining framework was further explored. The research undertaken in this chapter was accepted by the Australasian Workshop on Health Informatics and Knowledge Management 2013 and was published in the Australian Computer Society in the Conferences in Research and Practice in Information Technology series (Perimal-Lewis, 2013) .

6.1 Introduction

As discussed in Chapter 5, this research is one of the first comprehensive studies undertaken to assess the impact of ward allocation upon QoC for patients. Patients were admitted to hospital under GM and allocated either as inliers or outliers at a busy public teaching hospital. The complexity and diversity of hospital processes mean that there are also a considerable number of ways to measure the QoC of patients and the ways vary from unit to unit. The research identified common variables or attributes that could be used to measure QoC for inliers and outliers admitted to the GM units. The variables identified were *'discharge summary sent within two days of discharge'*, *'in-hospital mortality'*, '*re-admitted within seven days'*, *'total in-hospital LOS'* and *'time spent in the ED'*. A systematic framework as described previously was used to investigate and explore the GM population in order to define their inlier or outlier status.

Inpatient LOS has become one of the many ways used to measure performance of a hospital. Mean inpatient LOS has been used to measure QoC and hospital efficiency in terms of resource usage (Thomas et al., 1997). Shorter than average LOS could indicate that hospitals are discharging patients early and are possibly sacrificing QoC (Thomas et al., 1997). The hypothesis was that outliers would have a longer (less efficient) overall in-hospital LOS compared to inliers. The analysis undertaken in the previous chapter disproved this hypothesis in that outliers had a shorter LOS.

Out of all the QoC attributes identified, the outlying patients had a higher mortality rate and fewer discharge summaries for these outlying patients were sent within two days of discharge. The circumstances around the higher mortality rate were discussed in the Discussion Section: 5.6 of the previous chapter. Discharge summaries contain relevant information pertinent to a patient's care during a hospital admission. This information should be communicated to the primary health professionals who will continue a patient's care or provide future care for a patient after discharge (Li et al., 2011). The same authors established an association between delayed dissemination or the absence of discharge summary and re-admission rate thus encouraging health professionals to complete discharge summaries promptly. Others have also noted that prompt discharge summary dissemination is associated with decreased hospital re-admission. The hospital re-admission rate within three months decreased when a patient was followed-up by a physician who had received the patient's discharge summary (Van Walraven et al., 2002). Given the importance of discharge summaries for continuity of care and given that a high completion rate of summaries could potentially reduce the re-admission rate, the lower re-admission rate for the outlying patients with delayed discharge summaries was surprising.

To further investigate the importance of appropriate or inappropriate ward location for a patient, this chapter extends on the previous chapter by undertaking an analysis by clustering the GM patients into homogenous groups. Once the homogenous groups were derived, QoC attributes were assessed for both inlier and outlier groups.

Acknowledging the diversity of the GM patients as well as the complexity and variability embedded in each patient journey, it was important to reduce the heterogeneity of the patient journeys in order to gain better insight from the data. Disregarding patient heterogeneity can mask the discovery of meaningful patterns in patient characteristics which can lead to misleading results (Armstrong et al., 2011).

The aim of cluster analysis is to group cases (in this study the patient journeys) into homogenous groups based on the natural structure of the data (Tan et al., 2005). Cluster analysis is an exploratory technique which aims to group cases into clusters based on their similarities and dissimilarities (Luke, 2005). Cases in the same cluster share similar characteristics and are very dissimilar to cases belonging to other clusters (Mooi and Sarstedt, 2011). Applying statistical methods to a homogenous cluster of patients may reveal insights that are otherwise hidden due to pre-existing heterogeneity.

6.2 Method

The analysis was undertaken using the same pre-processed data set used in Chapter 5. After exclusions, there were 19,923 patients of whom 15,213 were classified as inliers, and 2,592 were classified as outliers and 2,118 who did not fit into either of these categories. The patients who did not fit into the inlier or the outlier group were not included in this analysis because the aim was to investigate the outliers and the inliers.

6.2.1 Process Mining – Case Perspective

Process mining in terms of case perspective can be characterised by the values of the corresponding data elements (van der Aalst, 2011). A case in this study is an individual patient journey. Each patient journey has a corresponding attribute that characterises that patient journey. Therefore from the derived event log, it would be possible to identify the attributes of a patient (*age, gender, Charlson Index*) admitted to a particular ward. In this regards, the characteristic of a patient may influence the ward to which the patient is admitted.

van der Aalst (2011) suggested that classical data mining techniques such as decision mining could be used to extend the process model. However, in this research the exploration of the case perspective used the clustering technique to derive homogenous clusters of the GM population for further analysis. Therefore the approach taken in this particular research extended on the conventional method of looking at the case perspective where the attributes of the cases and events are used to determine the best way to route cases in a process model (van der Aalst, 2011).

As such, ProM version 5.2 did not have the capability to undertake process mining in terms of case perspective for the purposes of this particular aim. Therefore IBM SPSS was used. This is further explained in Section 6.2.2

The framework used in constructing the derived event log enabled innovative ways to manipulate the derived event log to gain insight and therefore is software independent.

6.2.2 Statistical - Cluster analysis

In this research, the patient journeys were clustered using the two-step cluster analysis in SPSS. Two-step cluster analysis was chosen because of its ability to handle both continuous and categorical variables (SPSS, 2001). From the automatic number of clusters derived by this clustering procedure, an optimal number of clusters were derived using the exploratory method. At the same time, consideration was made of the practicality of having a large or small number of clusters against the ratio between clusters and the goodness of fit for the derived model.

In the first step, the automatic number of clusters is determined using Clustering Criterion by choosing either Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). The number of clusters derived for this data set using BIC and AIC were similar. SPSS computes the BIC and AIC for J clusters as per Equation 1 and Equation 2 below (IBM, 2011).

$$BIC(J) = -2\sum_{j=1}^{J} \xi_j + m_J \log(N),$$

Equation 1: BIC computation

$$AIC(J) = -2\sum_{j=1}^{J} \xi_j + 2m_J$$

Equation 2: AIC computation

In Equation 1, N stands for the total number of records in the data set. In Equation 2, m_J is calculated as shown in Equation 3. K^A, K^B and L_k in Equation 3 stands for '*total number of continuous variables used in the procedure*', '*total number of categorical variables used in the procedure*' and '*number of categories for the kth categorical variable*' respectively (IBM, 2011).

In the second step, the initial number of clusters derived in the first step is further refined. This is done by finding the largest increase in the distances between the two closest clusters (IBM, 2011). The distance between two clusters is calculated by using log-likelihood distance measure, which is the decrease in log-likelihood as the clusters are combined into one cluster.

$$m_J = J\{2K^A + \sum_{k=1}^{K^B} (L_k - 1)\}$$

Equation 3: m_J computation

6.3 Results

The patient journeys were clustered based on the QoC variables and their outlier or inlier status. Patient journeys with outlier status were journeys with "≥ 70% Outlier Hours" and patient journeys with inlier status were journeys with "≥ 70% Inlier Hours". The variables chosen have been assessed for collinearity between variables to ensure that they were unique in identifying distinct clusters. Table 14 shows the clustering results of the two homogenous clusters.

Table 14: Patient journey composition in the 2 clusters

	Cluster 1 (n=9968)	Cluster 2 (n=7837)
Size, %	9968 (56.6%)	7837 (44%)
Ratio of size between Cluster 1 and Cluster 2	1.27	
Average Silhouette *	0.60 (Good)	

* Measure of cluster cohesion and separation

The Cluster Cohesion measures how closely the cases in the cluster are related to other cases within the cluster (Tan et al., 2005). The Cluster Separation measures how well a cluster is separated or different from other clusters (Tan et al., 2005). The Silhouette Coefficient is the combination of both Cluster Cohesion and Cluster Separation for individual cases and clusters (Tan et al., 2005). Average Silhouette Coefficient ranges from negative one (-1) for a very poor model and one (1) for an excellent model (Kaufman and Rousseeuw, 2005). Based on the measurement as defined by Kaufman and Rousseeuw (2005), the model discussed above indicates a reasonable partitioning of data. The average Silhouette Coefficient is calculated as per Equation 4:, where '*A is the distance from the case to the centroid of every other cluster which the case belongs to*' and '*B is the minimal distance from the case to the centroid of every other cluster*' (IBM, 2012).

$$(B-A) / \max(A,B)$$

Equation 4: The average Silhouette Coefficient computation

The ratio between the smallest and the largest cluster is 1.27 which is a good ratio in that the larger cluster is less than two times larger than the smaller cluster. Between the 0.60 average Silhouette and the ratio, the described model is a good fit for the purpose of this study where all the QoC variables identified had to be included in the model in order to give the insight required.

Table 15: Patient characteristics

Patient Characteristics	Cluster 1 (n=9968)	Cluster 2 (n=7837)
Charlson Index	1.39 (1.88) *	1.55 (2.03) *
Sex, n, (%)	Female, 5831, (58.5%) **	Female, 4474, (57.1%) **
Age, years	72.81 (18.32) *	71.61 (18.47) *

** Mean (SD) for continuous variables; ** Mode, n, (%) for dichotomous variables (indicating the most frequent category)*

The characteristics of patients in both clusters are listed in Table 15. The number of female patients is higher in both the clusters. Charlson co-morbidity Index (CI) is a widely used clinical index for the evaluation of co-morbidities (Simon et al., 2012). CI is a pre-calculated variable for every patient admission and was supplied with the data set. Patients in cluster two had a higher CI score suggesting that these patients were sicker than those in cluster one. Age differences between patients in both clusters were small and clinically insignificant.

Table 16 summarises the QoC attributes and their relative importance in deriving the two clusters. The predictor importance for each QoC attribute is calculated as per Equation 5: where *' Ω is the set of predictor and evaluation fields'* and *'sig_j is the p-value'* (IBM, 2012). The values are relative; therefore the sum of values for all attributes is one. An attribute with a value close to one is the most important attribute in deriving the cluster and a value close to zero is the least important attribute.

$$VI_i = \frac{-\log_{10}(sig_i)}{\max_{j \in \Omega}(-\log_{10}(sig_j))}$$

Equation 5: Predictor importance computation

The most significant QoC attribute for deriving the two homogenous clusters was the *'discharge summary sent within two days of discharge'* with the relative importance of 1.0. Cluster 1 consists of patient journeys where the discharge summaries were sent within two days for the entire (100%) cluster population as opposed to Cluster 2 where 94.8% of patient journeys did not have their discharge summaries sent within two days of discharge. This suggests inferior QoC received by patients in Cluster 2.

The next QoC attribute used to derive the two clusters was *‘in-hospital mortality'* with relative importance value of 0.61. None of the patients in Cluster 1 died during their hospital admission. 8.3% of patients in Cluster 2 died.

The next QoC attribute in order of importance used to derive the two clusters was *'readmission within seven days'* with relative importance value of 0.4. Once again none of the patients in Cluster 1 were re-admitted within seven days; however 5.4% of patients in Cluster 2 were re-admitted within seven days.

The next QoC attribute was *'total in-hospital LOS'* with relative importance value of 0.04. Patients in Cluster 1 had a longer mean LOS (6.14 days) compared to patients in Cluster 2 with mean LOS of 5.54 days.

The final QoC attribute was *'time spent in the ED'* with relative importance value of 0.03. Patients in Cluster 1 spent slightly longer time in the ED (5.7 hours) compared to patients in Cluster 2 with mean time of 5.18 hours.

Table 16: Summary of quality of care variables/attributes

Quality of Care Variables	Cluster 1 (n=9968)		Cluster 2 (n=7837)	
	Mean / Mode *	Predictor Importance **	Mean / Mode *	Predictor Importance **
Discharge Summary sent within 2 days of discharge	Yes (100%)	1	No (94.8%)	1
In-hospital mortality	No (100%)	0.61	No (91.7%)	0.61
Re-admitted within 7 days	No (100%)	0.4	No (94.6%)	0.4
Total in-hospital LOS, days	6.14	0.04	5.54	0.04
Time spent in the ED, hours	5.7	0.03	5.18	0.03

** Mean for continuous variables; Mode for binary variables (indicating the most frequent category);*

*** Relative importance of each quality of care variable/attributes in estimating the model*

The next important step in this study was to investigate whether there were any significant differences between the QoC attributes and patient characteristics in both clusters for those patients in the outlier and the inlier groups defined earlier. Table 17 and Table 18 below summarises the QoC attributes and patient characteristics for the outlier and the inlier group in Cluster 1 (n=9968) and Cluster 2 (n=7837) respectively. The *'Sig.'* column shows the p-value where significance level $\alpha < 0.05$ is considered significant. The Mann-Whitney U test was used as the significance level test for continuous variables. The chi-square test was used as the significance level test for proportions.

In Cluster 1, 10.20% of patient journeys were in the outlier category with the rest of the patient journeys under the inlier category.

The age differences between outliers and inliers were statistically significant (p=0.000; Mann-Whitney U test). According to the clinicians this age difference of just over two years is not of clinical importance. The differences in CI between the inliers and the outliers was not statistically significant (p=0.810; Mann-Whitney U test) suggesting that disease complexity was not an important characteristic in differentiating the patients. Compared with inliers, outliers in Cluster 1 spent a much longer time in the ED waiting for an inpatient bed to become available after a decision to admit them had been made. The difference in time spent in the ED between the outlier and the inlier groups was statistically significant (p=0.000; Mann-Whitney U test). Despite spending a longer time in the ED waiting for an inpatient bed, outliers had overall shorter in-hospital LOS compared with the inliers. The differences in LOS are statistically significant (p=0.000; Mann-Whitney U test). As noted before, there was no in-hospital mortality for patients in Cluster 1. No Cluster 1 patient was re-admitted within seven days and their discharge summaries were all sent within two days of discharge regardless of their outlier or inlier status.

Table 17: Quality of care attributes comparison for inliers and outliers in cluster 1

	≥70% Outlier Hours (n=1017)	≥ 70% Inlier Hours (n-8951)	Sig.
Age, years	70.72 (19.63) *	73.04 (18.15) *	0.000
Charlson Index	1.35 (1.84) *	1.40 (1.88) *	0.810
Time spent in the ED, hours	7.41 (7.45) *	5.50 (5.73) *	0.000
Total in-hospital LOS, days	5.21 (5.38) *	6.25 (6.04) *	0.000
In-hospital mortality, n, %	0	0	n/a
Readmitted within 7 days, n, (%)	0	0	n/a
Discharge Summary sent within 2 days of discharge, n, (%)	1017 (100%)	8951 (100%)	n/a

** Mean (SD) for continuous variables*

In Cluster 2, there were 20.1% outliers and the rest were inliers as per **Table 18**. CI was not statistically significant between the outliers and the inliers. This was similar to patients in Cluster 1. Age differences between the outliers and the inliers were statistically significant (p=0.000; Mann-Whitney U test), however as noted before this was deemed not to be of clinical significance. Contrary to patient journeys in Cluster 1, although outliers in Cluster 2 spent a slightly longer time in the ED compared to the inliers, this was not statistically significant (p=0.778; Mann-Whitney U test) for patients. Similar to outliers in Cluster 1, outliers in Cluster 2 had a shorter overall in-hospital LOS compared with the inliers and this was statistically significant (p=0.000; Mann-Whitney U test).

The main differences between patients in Cluster 1 and Cluster 2 is in relation to the three QoC attributes; '*in-hospital mortality*', '*readmitted within seven days*' and '*discharge summary sent within two days of discharge*'. All patients with inferior QoC in relation to these three attributes were in Cluster 2. In-hospital mortality between outliers and inliers in this cluster was not statistically significant. Outliers were re-admitted more than the inliers and this was statistically significant (p=0.022; chi-square test) suggesting that QoC for outliers were inferior to those who were inliers. Fewer outliers had their discharge summaries sent within two days of discharge compared to the inliers and this was statistically significant (p=0.000; X^2 test). This again suggests an inferior QoC for the outliers.

Table 18: Quality of care attributes comparison for inliers and outliers in cluster 2

	≥70% Outlier Hours (n=1575)	≥ 70% Inlier Hours (n=6262)	Sig.
Age, years	69.13 (18.69) *	72.24 (18.37) *	0.000
Charlson Index	1.57 (2.06) *	1.54 (2.02) *	0.551
Time spent in the ED, hours	5.63 (6.88) *	5.06 (5.48) *	0.778
Total in-hospital LOS, days	4.51 (4.58) *	5.81 (5.77) *	0.000
In-hospital mortality, n, %	117 (7.4)	537 (8.6)	0.141
Readmitted within 7 days, n, (%)	30 (1.90)	40 (0.64)	0.022
Discharge Summary sent within 2 days of discharge, n, (%)	38 (2.4)	366 (5.8)	0.000

* Mean (SD) for continuous variables

Another analysis was carried out to compare the characteristics of patients in Cluster 1 and Cluster 2. Apart from age (p=0.066; Mann-Whitney U test), CI and sex were significantly different between patients in Cluster 1 and Cluster 2 (p=0.000; Mann-Whitney U test). All QoC variables were significantly different between patients in Cluster 1 and patients in Cluster 2 (p=0.000; Mann-Whitney U test).

6.4 Discussion

The main differences between patients in Cluster 1 and Cluster 2 relate to QoC attributes. Patients in Cluster 2 had inferior QoC compared to those in Cluster 1 regardless of whether they were outliers or inliers. In Cluster 1, there were no in-hospital deaths, none was re-admitted within seven days and all discharge summaries were sent within two days of discharge. Analysing patients in Cluster 2 (those who had inferior QoC) in regards to their outlier and inlier status revealed meaningful in-sight because the comparison was done on a cluster of patients with similar characteristics and QoC attributes. One of the major challenges of the study was the considerable effort that went into exploring the data to discover the best way to derive the outlier and the inlier population. Over the period of 6 years, investigating the spread of time spent in an outlier ward and the spread of time spent in an inlier ward led to the grouping of this variable into "≥ 70%" of outlier or inlier time. The method used and the grouping of this variable was believed to be the best approach for this data set to discover the effect of ward allocation on the QoC received by these two groups of patients.

The effect of the QoC attributes on the outliers and inliers status was also investigated using more homogenous groups of patients. The relationships discovered based on analysing the QoC attributes on homogenous clusters were different to those discovered when the patients were not clustered.

This study demonstrated the complexity of analysing hospital data and the need to identify from the raw data groups of patients with more similar characteristics. Although outliers in both clusters were younger and the difference was statistically significant, the age difference was not considered a clinically significant difference. This emphasised the importance of involving clinicians when developing meaningful conclusions.

The patient co-morbidity index, CI, did not significantly differ whether the patient was admitted to a home ward or not. This result was also obtained when the analysis was carried out without clustering the patients into two homogenous clusters (Chapter 5).

There was a linear relationship between being an outlier or inlier and the amount of time spent in the ED. This association is only significant for outliers in Cluster 1. The point-biserial correlation was used to capture the relationship between a dichotomous variable and a continuous variable (DeCoster and Claypool, 2004). Point-biserial correlation showed a high correlation between the time spent in ED and being an outlier ($r = 0.097$, $p = 0.000$).

Contrary to the hypothesis, outliers in both clusters had a shorter in-hospital LOS, a similar association was obtained when analysing the patient population without clustering. According to the clinicians, this is a promising indicator because being an outlier did not compromise the efficiency of care in relation to the overall in-hospital LOS. Interestingly, outliers had inferior QoC in relation to the extended time they spent in the ED.

Point-biserial was used to further analyse the correlation between total in-hospital LOS and in-hospital mortality for patients in Cluster 2. The correlation showed lower in-hospital LOS was associated with patients who did not die whilst in-hospital ($r = -0.139$, $p = 0.000$). The finding reveals that the outliers' short LOS is not associated with in-hospital mortality.

Using point-biserial for patients in Cluster 2, lower in-hospital LOS was associated with patients who were not readmitted within seven days of discharge (r = -0.042, p = 0.000) suggesting that re-admission might not necessarily be linked with shorter LOS or the outliers.

6.5 Conclusion

In conclusion, patients in Cluster 2 had significantly inferior QoC. The most important QoC predictor variable used to separate the GM population into two clusters was '*discharge summary sent within two days of discharge*'. The next most important QoC predictor variable that separated these two clusters was '*in-hospital mortality',* followed by *'re-admitted within 7 days', 'total in-hospital LOS'* and *'time spent in the ED'.*

As the results from Chapter 5 revealed, discharge summary were less likely to be completed in the outlier population. Almost all patients (94.8%) in Cluster 2 did not have their discharge summaries sent promptly. Discharge summaries were less likely to be completed in the outlier population in Cluster 2. So not only patients in Cluster 2 had inferior QoC, but the outliers in Cluster 2 did not have their discharge summaries sent promptly.

The next important predictor variable of inferior QoC was the '*in-hospital mortality'.* Results from Chapter 5 shows, outlier status was a significant predictor of increased mortality. In this study all in-hospital mortality occurred to patients in Cluster 2. Within Cluster 2, the in-hospital mortality for outliers and inliers were not statistically significant; although the inliers in Cluster 2 were more likely to die.

The results in Chapter 5 revealed; the risk of readmission within seven days was not significantly affected by the outlier status. However, outlier patients in Cluster 2 were more likely to be re-admitted within seven days of discharge.

In-hospital LOS for outliers in both Cluster 1 and Cluster 2 has been shorter than the inliers in these clusters. This finding is persistent with the findings from the study undertaken in Chapter 5.

Time spent in the ED was significantly longer for outliers in Cluster 1. This persisted despite the overall superior QoC results for patients in Cluster 1.

The shorter LOS for outliers continued to persist even after clustering the GM patients into homogenous clusters. Therefore it could be concluded that outlier status is a significant predictor of reduced LOS, suggesting that efficiency of care was not compromised for these outlier patients. However, effectiveness of care was compromised for these outlier patients.

Outlier status is a significant predictor of delayed discharge summary and a significant predictor of readmission within 7 days of discharge. Outlier status is also a significant predictor of longer time in the ED. This persisted for patients in Cluster 2 who, otherwise, received superior QoC. In conclusion there is strong evidence; the outlying GM patients received inferior QoC. Once again, the result highlights the need for further research, to identify those factors that are influencing LOS. Further research is needed to ensure that any organisational benefits that might be produced from any reduced LOS are not being obtained at the expense of less than optimal care for the outlying patients.

7 Emergency Department lengths of stay: characteristics favouring a delay to the admission decision as distinct from a delay while awaiting an inpatient bed

This chapter discusses the core of the work done in investigating the Emergency Department (ED) Length of Stay (LOS) in the context of patient journey modelling. This work was accepted for publication by the Internal Medicine Journal of the Royal Australasian College of Physicians (Perimal-Lewis et al., 2014a).

7.1 Introduction

In this chapter, the organisational perspective of the patient journey process was investigated in terms of the resource perspective in the ED. The purpose of undertaking process mining under the organisational perspective is to discover the organisation's structure according to role classification or to portray the social network of the organisation (van der Aalst, 2011). The conventional way of process mining in this perspective is to gain insight into the workings of the resources responsible for undertaking certain tasks, into the transfer of work between resources, into the work pattern of resources and to gain insight into the social structure of the organisation (van der Aalst, 2011).

As discussed and presented in this thesis, the hospital structure is complex and there are many interactions that take place in a patient's journey through the hospital involving many resources. The absence of PAIS in the healthcare environment that can capture the hospital wide process in an event log means that a new way was needed to undertake process mining from the context of an organisational perspective.

Another issue surrounding any investigation of resources is an ethical consideration. Ethical concerns also preclude the possibility of obtaining relevant data and analysing such data. Common resource information that could be easily collected from the hospital system is the identity of the units and wards occupied by the patient. Although the performance, in terms of throughput, of each unit can be easily derived, the usefulness of such information is subjective. Throughput of a Cardiology unit is not comparable with the throughput of the GM unit because of the inherent differences in the nature of the clinical work involved. In a similar context, the QoC measures and the KPIs discussed throughout this thesis are not uniform from unit to unit. The quality of medical care performed by the Cardiology team does not give much information about the quality of care performed by other medical teams and accordingly there is a systemic lack of comparable measurements in healthcare (Wynia, 1999). The complex patient movement through the hospital system discussed in Chapter 3 and Chapter 4 highlights that the transfer of patients from ward to ward or from unit to unit is not only dependent on the capacity of the hospital in terms of bed availability but also on the clinical treatment of each patient. As such, the throughput measurement for units, the work transfer patterns and most performance measures relevant to other industries are not applicable in the healthcare setting.

Therefore, the process mining framework adapted in this research aimed to investigate surrogate measures of resources. The data supplied by the hospital and constructed as an event log contained the variables needed to undertake this research and those patient journeys started in the ED. EDs in public hospitals are put under enormous pressure to increase performance as the need for emergency services increases. The introduction of the National Emergency Access Target discussed previously further puts pressure on EDs to increase throughput. In a complex ED environment there are multiple interacting resources that directly impact upon an ED's efficiency. The decision taken in the ED to admit a patient is usually a multi-stage process where the junior ED doctors discuss a patient's admission with a senior ED doctor. Once the decision is made to admit a patient, the ED doctors then discuss with the relevant inpatient unit on the possibility of them accepting a patient as an inpatient. It is only when the inpatient unit accepts the potential patient that the processes to progress the patient out of the ED are initiated. This is a time consuming as well as a stressful process especially if the inpatient unit refuses to take the patient. Therefore fine detail modelling of an organisational perspective in terms of resources is almost impossible to achieve. As a result, a surrogate measure of resource investigation is appropriate. In this instance the number of patients in the ED at the time of triage and at the time of the decision to admit was used as measures of how busy the ED was. Accounting for whether decision making occurred inside or outside of working hours was another surrogate measure of resources. The hospital is staffed differently inside and outside of working hours.

The research discussed in this chapter investigated two ED KPIs in terms of resource constraint. The first ED KPI relates to triage-to-admit time and the second relates to boarding time.

It was hypothesised that the number of patients in the ED at any given time will influence the efficiency (*how quickly*) of the decision making process for a particular patient. In a similar context of efficiency, it was hypothesised that decision making will be more efficient during working hours for a particular patient because of the factors that could potentially influence this decision making process such as quicker generation of laboratory results (lab turnaround time).

The crowding of an ED and prolonged stays for ED patients are associated with a deterioration in measures of both the efficiency and the quality of subsequent inpatient care (Ackroyd-Stolarz et al. 2011, Liew and Kennedy 2003, Richardson 2006, Sprivulis et al. 2006). The causes of ED overcrowding or congestion are complex but the prolonged stay of patients in ED is clearly one of the major factors involved (Forero et al. 2010, Fatovich et al. 2005, Geelhoed and de Klerk, 2012). Patients, who stay over eight hours in the ED despite having been already admitted to hospital, are called “boarders” (Carr et al. 2010, Henneman et al. 2010, Viccellio et al. 2009). The numbers of boarders rise during periods of high hospital bed occupancy when access to vacant hospital beds are limited (“access block”). Ordering many diagnostic tests for patients may delay the decision for patient admission and that can also lead to ED congestion (Fatovich et al. 2005, Lee-Lewandrowski et al. 2009). Concha et al. (2014) called for improving quality of care for at-risk diagnoses with excess mortality associated with ED patients admitted on the weekend.

In many hospitals in Australia, GM is one of the largest inpatient services. Their inpatients are implicated in access block over and above other medical specialty patients (Carr et al., 2010). The association between ED LOS and patient outcomes may be causal hence effective interventions are sought which will both reduce ED LOS and improve important patient outcomes such as in-hospital mortality Mitra et al. (2012) and inpatient LOS.

At FMC, incoming patients are identified and triaged upon presentation into Australasian Triage Scale (ATS) categories; these scores are used as a measure of the urgency of review necessary for each patient. Patients are also streamed into those who are likely to be discharged home from the ED and those who are likely to be admitted to the hospital. Priority is given to triage category 1&2 patients. Otherwise patients are seen in order of arrival in both streams (King et al., 2006).

7.2 Aims

The aim of this research was to understand better the patient characteristics of those admitted GM patients who stay a long time in ED either awaiting a decision to be admitted or awaiting a bed. It also investigated whether certain characteristics, external to the patient and potentially modifiable, influenced these two phases of care within the ED.

7.3 Methods

This is a retrospective study of all hospital inpatient stays for patients admitted to and discharged by the GM service at FMC between the 1st January 2003 and 20th September 2009 who presented to the ED. The study examined the population of patients admitted to the GM service; approximately 4,000 patients per annum. The patients had complex disease not easily identified as appropriate to be allocated immediately to a sub-specialty unit such as cardiology or neurology.

The derived event log containing the patient journey records were merged with the hospital's ED computerised dataset to obtain the various variables needed for the analysis. Only records with valid values for all variables were included in the analysis so, after 1,121 patient entries with missing values for variables were excluded, a final sample size of 19,476 patient admissions was achieved.

The patient characteristics assessed were age, ATS category and Charlson co-morbidity index (Charlson et al., 1987). The latter index, modified by Quan and colleagues (Quan et al., 2005), was used as a measure of co-morbidity. The ED characteristics that were assessed include the number of patients in the ED both at the time of presentation and of admission decision for each patient and whether the presentation and admission decision happened during working hours (0800-1800 hours from Monday to Friday).

7.3.1 The ED phases

The time a patient spends in ED was divided into two phases. This decision to divide the ED time into two phases was influenced by reported ED KPIs. The first period follows the presentation and triage of the patient and involves their assessment and management by ED staff after which a decision is made for the patient either to be admitted to hospital or be discharged home. This phase is referred as triage-to-admit time phase. It is the elapsed time between triage and the decision to admit the patient.

Once staff decides a patient is to be admitted, a second period begins during which the patient awaits a bed in the hospital itself (Carr et al., 2010). This phase is defined as the boarding time phase. It is the time spent in the ED from the decision to admit the patient until the patient physically leaves the ED to go to an inpatient bed (Solberg et al., 2003).

A long boarding time is associated with worse outcomes for the patient including a longer overall stay in hospital (Singer et al., 2011). However, most studies have not distinguished between triage-to-admit and boarding times. The relative importance of the triage-to-admit time has only recently been addressed in a study restricted to general medical patients Mitra et al. (2012) where a short triage-to-admit time was associated with an increase in mortality if coupled with a prolonged boarding time.

7.3.2 Statistical Analysis

Statistical analysis was performed using Stata® 12.2. The outcome variables were triage-to-admit time and boarding time. Multiple linear regressions were performed to model the relationship between the explanatory variables and the outcome variables. Triage-to-admit time was not normally distributed, therefore was log transformed to satisfy normality assumptions and the exponentiated regression coefficient was interpreted to estimate the expected geometric mean of the original value. Boarding time was not log transformed. The margins command within Stata was used to compute predicted means for triage-to-admit time and boarding time. Independent variables in the regression models were assessed for non-linearity. The F value for all linear models had a p-value of <0.0001 indicating the independent variables are reliable for predicting the dependent variables. Variables in all models were checked for multicollinearity using the Variance Inflation Factor. There were no variables with standard errors for the b coefficient that were greater than 2.0 suggesting there were no numerical problems with any of the independent variables. A P-value of <0.05 was considered to be statistically significant.

7.4 Results

Table 19 is a summary of the patient population describing their total ED LOS, triage-to-admit time and boarding time. For all 19,476 patients, the mean ED LOS was 13.6 hours (SD 8.3) and this was comprised of a triage-to-admit time of 7.5 hours (SD 4.7) and a boarding time of 6.1 hours (SD 6.5). Patients were also grouped into three categories according to whether their total ED LOS, triage-to-admit and boarding times were under four hours, four to eight hours or over eight hours. In terms of entire ED LOS, only 588 admitted GM patients (3%) completed their stay in ED within four hours. 5,422 patients (28%) stayed in ED between four and eight hours and the remaining 13,466 (69%) stayed in ED for a total of at least eight hours. In terms of triage-to-admit times, for 4,137 (21%) patients, their admission decision was made within four hours of arrival. For 8,759 (45%), this decision was made within four to eight hours and for 6,580 (34%) this decision took over eight hours to be made. Once the admission decision had been taken, 9,799 patients (50%) were taken from the ED to their inpatient ward bed within four hours. The boarding time was between four and eight hours for 4,541 (23%) patients and the remaining 5,136 (26%) patients waited in ED for an inpatient bed for over eight hours after their admission decision had been taken.

Table 19: Descriptive statistics for patients in the ED

Descriptive statistics for patients in the Emergency Department (ED)				
	Number of patients	**ED LOS (hours) mean (SD)**	**Triage-to-admit time (hours) mean (SD)**	**Boarding time (hours) mean (SD)**
Presentation to triage inside working hours	8,596	11.76 (7.19)	6.51 (3.78)	5.25 (5.84)
Presentation to triage outside working hours	10,880	15.09 (8.74)	8.35 (5.09)	6.74 (6.88)
Admission decision inside working hours	5,808	17.53 (9.33)	8.33 (5.37)	9.19 (7.77)
Admission decision outside working hours	13,668	11.96 (7.14)	7.21 (4.26)	4.75 (5.33)

Working hours = 0800-1800 hours from Monday - Friday

7.4.1 Triage-to-admit time

Age and Charlson Index were not significant predictors of triage-to-admit times. Significant predictors are shown in Table 20. The impact of triage category on triage-to-admit times was assessed by comparing the mean triage to admit times of ATS categories 2 to 5 against the triage-to-admit times for ATS category 1 patients. The adjusted expected mean triage-to-admit time for patients with ATS 2 is 32% higher than patients with ATS 1 (Coef. 0.28, 95% CI: 0.20 -0.35, p<0.001).The adjusted expected mean triage-to-admit times increased progressively with each ATS category and hence triage-to-admit times for patients with ATS 3 and ATS 4 are extended by 68% and 74% respectively in comparison with patients categorised as ATS 1. The adjusted expected mean triage-to-admit time for patients with ATS category 5 was only 56% higher compared with ATS 1. The small number of patients in this category (112; 0.58% of the cohort) might explain this discrepancy. Every additional patient in the ED at the time of a patient's triage increased the triage-to-admit time by 0.82% (Coef. 0.0082, 95% CI: 0.0075-0.0089, p<0.001). Triage-to-admit time was 29% higher for those who arrived outside of working hours compared to those who arrived inside of working hours (Coef. 0.2556, 95% CI: 0.2388-0.2724, p<0.001).

Table 21 shows the triage-to-admit time categorised according to ATS categories and the predicted impact of extra numbers of patients in the ED at the time of presentation to triage as well as the effect of the time of day of the presentation. The relative influence of these three factors on triage-to-admit time can now be gauged; ATS category being a dominant influence. The diurnal discrepancy in triage-to-admit time is exaggerated as ATS category increases.

Table 20: Linear regression results for triage-to-admit time

Linear regression results for triage-to-admit time					
	Coef.	**Std. Err.**	**t**	**P>|t|**	**[95% Conf. Interval]**
Age	-0.0000782	0.0002446	-0.32	0.749	-0.0005577 – 0.0004013
ATS Categories					
ATS category 2	0.2797479	0.0383987	7.29	0.000	0.2044832 – 0.3550126
ATS category 3	0.5160179	0.0375353	13.75	0.000	0.4424455 – 0.5895903
ATS category 4	0.5541536	0.380669	14.56	0.000	0.4795393 – 0.628768
Charlson Index	0.0035381	0.0021709	1.63	0.103	-0.0007171 – 0.0077932
Presentation to triage outside working hours	0.2555965	0.0085735	29.81	0.000	0.2387916 – 0.2724013
Patient count in ED at time of presentation to triage	0.0082037	0.000336	24.42	0.000	0.0075452 – 0.0088623

Table 21: Estimated means for triage-to-admit time according ATS category and the number of patients in the ED

Estimated means for triage-to-admit time according to Australian Triage Scale (ATS) category and the number of patients in the Emergency Department (ED)			
ATS category	**Number of patients in ED at the time of presentation to triage**	**Adjusted § mean triage-to-admit time inside working hours (hours)**	**Adjusted § mean triage-to-admit time outside working hours (hours)**
ATS 1	40	3.48	4.49
	50	3.77	4.87
	60	4.10	5.29
	70	4.45	5.74
ATS 2	40	4.60	5.94
	50	4.99	6.45
	60	5.42	7.00
	70	5.88	7.60
ATS 3	40	5.83	7.52
	50	6.32	8.17
	60	6.86	8.86
	70	7.45	9.62
ATS 4	40	6.05	7.82
	50	6.57	8.48
	60	7.13	9.21
	70	7.74	10.00

Working hours = 0800-1800 hours from Monday - Friday

§ Adjusted for age, ATS categories, Charlson index, time of presentation to triage and patient count in ED at the time of presentation to triage

7.4.2 Boarding time

Significant predictors of boarding time are presented in Table 22. For every decade increase in a patient's age, the time spent in the ED awaiting a bed increased but only slightly; by just under three minutes (Coef. 0.0048, 95% CI: 0.00022 – 0.0094, $p < 0.040$). ATS categories and Charlson Index were not significant predictors for boarding time.

For every additional patient in the ED at the time of an admission decision, boarding time increased by just under ten minutes (Coef. 0.164, 95% CI: 0.158 - 0.170, $p < 0.001$). Therefore the adjusted mean boarding time is 4.34 hours when there are 30 patients in the ED at the time of admission decision and 5.98 hours when there are 40 patients in the ED at the time of admission. Table 23 shows the predicted impact on the boarding time of additional patients in the ED at the time of the admission decision and the effect of time of day on boarding time. The diurnal difference in boarding times is in the opposite direction to that of the triage-to-admit times and seems independent of the number of patients in the ED. The boarding time itself increases with the number of patients in the ED at the time of the admission decision.

Table 22: Linear regression results for boarding time

Linear regression results for boarding time					
	Coef.	**Std. Err.**	**t**	**P>|t|**	**[95% Conf. Interval]**
Age	0.004793	0.0023309	2.06	0.040	0.0002242 - 0.0093618
ATS Categories					
ATS category 2	0.2761263	0.2889099	0.96	0.339	-0.2901619 - 0.8424144
ATS category 3	0.4923439	0.2805417	1.75	0.079	-0.0575419 – 1.04223
ATS category 4	0.3904473	0.2884975	1.35	0.176	-0.1750326 - 0.9559272
Charlson Index	0.0181628	0.0220462	0.82	0.410	-0.0250497 - 0.0613753
Presentation to triage outside working hours	-3.139183	0.0941635	-33.34	0.000	-3.323751 - -2.954614
Patient count in ED at time of admission decision	0.1640488	0.0032322	50.75	0.000	0.1577134 - 0.1703842

Table 23: Estimated means for boarding time according to the number of patients in the ED

Estimated means for boarding time according to the number of patients in the Emergency Department (ED)		
Number of patients in ED at the time of admission decision	**Adjusted § mean boarding time (hours) if decision to admit inside working hours**	**Adjusted § mean boarding time (hours) if decision to admit outside working hours**
40	8.18	5.05
50	9.83	6.69
60	11.47	8.33
70	13.11	9.97

Inside working hours = 0800-1800 hours from Monday - Friday

§ *Adjusted for age, Australian Triage Scale (ATS) categories, Charlson index, time of admission decision and patient count in ED at the time of admission decision*

7.5 Discussion

This research investigated the two phases of ED time: the triage-to-admit time: a time during which the patient is assessed, treatment commenced and disposition decided and the boarding time: a time when the patient awaits an inpatient bed while their treatment continues. Supporting their differing aetiologies, the proportional distribution of times spent in the ED by the whole population clearly differs. Strategies to reduce the time patients spend in ED should therefore differ depending upon whether a decision to admit has been reached.

A slower decision to admit a patient may have direct effects upon that patient's care but also may be detrimental to others' care due to that delayed patient adding to any congestion existing in the ED. It can take a longer time to decide to admit a patient to hospital if the patient presents to ED outside of working hours or at a time of ED congestion and this triage-to-admit time also lengthens as the ATS category becomes less urgent. There may be limited out-of-hours access to investigations or limited access to senior decision makers out-of-hours; both factors might contribute to these observations. When the ED is congested, clinical resources to manage the increased workload can be stretched, also delaying the admission decision. The less urgent the ATS category, the longer the admission decision took. This finding may be partly explained by a delay in the initiation of clinical assessment of these possibly lower acuity patients.

ED congestion and an admission decision that is taken within working hours both predicted a longer boarding time for the patient. Prolonged boarding times for patients will produce ED congestion so the former association should come as no surprise but, of note, ED congestion affects boarding times to a greater extent than it affects triage-to-admit times. A lengthening of boarding times during working hours points to the problems of exactly matching the timing of bed availability to the demand for beds by incoming patients. Hospital inpatient beds are usually vacated inside, rather than outside, working hours and if they are vacated late in the working day, beds only become available after hours. A bed is sourced for a patient once the decision is taken to admit that patient therefore diurnal variation in bed availability might explain the relation of boarding time to working hours. The best strategy for achieving a short boarding time might be to have beds available for patients at the times the admission decisions are made. Currently there is better synchronisation of these processes outside than inside working hours. Alternatively, one could transfer all boarders to a separate in-hospital location outside the ED for medical care while waiting an inpatient bed. A focus upon discharge of hospital inpatients before 10 am may also help to reduce ED boarding time during working hours. This study confirms that older patients board for longer than young patients but the influence of age on boarding time was trivial and the study was not designed to explain such an association.

The study discovered considerable diurnal variation in the patients' transit through the ED of a large hospital. At the same time of day, those patients awaiting a bed might experience delays leaving the ED yet those patients awaiting an admission decision will experience relatively accelerated care. The opposite will occur at another time of day. The strategies to reduce ED LOS could be better focussed and more accurately measured. To reduce the triage-to-admit time, one could increase the available medical resources (*e.g. the seniority and number of clinical staff on duty*) and increase the out-of-hours access to investigations and specialist consultants. To reduce boarding time, one could increase the bed availability within the hospital – especially at times of peak demand. When appraising the effect of a specific intervention within the ED, assessing ED LOS as a whole may not reveal significant efficiencies achieved in one part of the patient journey in the ED or during one period of the day.

This research has several limitations. First, this study is retrospective and restricted to GM patients so may not be generalizable to other groups of patients. Secondly, the triage-to-admit time has not been broken down into even shorter components of ED work. Finally, patients admitted to the hospital's intensive care facility directly from ED are not included in this study.

7.6 Conclusion

ED congestion is associated both with delays in deciding to admit patients and even longer delays for admitted patients leaving ED for an inpatient bed. Outside of working hours, admission decisions are delayed yet egress of patient from ED is faster, producing opposing effects upon ED LOS at the same time of day. ATS categories are significant predictors for triage-to-admit time whereas these categories are not significant predictors for boarding time. Age is not a significant predictor of triage-to-admit time and had only a slight effect upon boarding time. Strategies to reduce the time patients spend in ED should take into account the importance of service availability for the time of day and should differ depending upon whether a decision to admit the patients has been reached.

8 Health intelligence: Discovering the process model using process mining by constructing Start-to-End patient journeys

This chapter builds and extends on the research done on Emergency Department (ED) phases completed in Chapter 7. This chapter demonstrates the applicability of the control flow perspective of the process mining framework. The research undertaken in this chapter was accepted by the Australasian Workshop on Health Informatics and Knowledge Management 2014 and was published in the Australian Computer Society in the Conferences in Research and Practice in Information Technology series (Perimal-Lewis et al., 2014b).

8.1 Introduction

Hospitals have mature processes that collect data on various quality measures. The performance of public hospitals is compared and judged publically by certain KPIs. One such KPI is any improvement the hospital might make in ED throughput measures such as reducing the time for a patient to be first seen by a doctor, reducing rates of patients who did-not-wait and reducing ED LOS (Shetty et al., 2012). The length of time a patient stays in hospital from admission to discharge is one of the criteria used to measure a hospital's performance in general. Performance can even be measured as the percentage of patients who stayed beyond an established LOS target (Kolker, 2008).

Widespread public reporting will encourage transparency and accountability as reported by McClelland et al. (2012) and Cameron et al. (2002) although dysfunctional processes might be identified that compromise QoC (Dellifraine et al., 2010). Performance measurements in ED have been reported to contribute to process improvements when unrecognised gaps were identified and measures taken to improve the underlying process (McClelland et al., 2012). In Australia, the proportion of patients 'seen on time' is one of the KPIs used as a measure of ED efficiency. In ED, measures relating to care timeliness including ED throughput and measures of care outcome such as mortality form the majority of quality measures of performance of this department (McClelland et al., 2012).

O'Connell et al. (2008) reported that ED congestion is intensified by regular failure to manage those processes responsible for progressing patients through the hospital. The lack of shared understanding between staff, patients and carers of the probable patient path was a contributing factor. The authors called for better inpatient management to ensure a better patient "flow" which would in turn ease the issues faced within ED.

In order to facilitate better flow for patients as they travel from ED to their eventual discharge from the hospital, it is necessary to understand the complex patient journey from start-to-end. Traditionally healthcare modelling has been done using various mathematical modelling techniques focused on forecasting and predicting in order to improve healthcare performances (Perimal-Lewis et al., 2012b). New approaches which use a combination of techniques to complement the strengths and weaknesses of any one technique are now emerging. For example, Ceglowski 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. This strategy could provide insight into the complex relationship between patient urgency, treatment and disposal and the occurrence of queues for treatment.

Simulation is widely used in healthcare as a basis to understand processes, for decision making and for prediction. There are many simulation projects which have been discontinued because these models fail to improve the underlying processes and/or engage the process owners (McHaney and Cronan, 2000). The decentralised structure of the hospital system discussed in Chapter 3 means many of these simulation initiatives are service- or unit-specific. There are many reports of successful service- or unit-specific simulation projects that have been used by doctors for clinical decision making. What is lacking is a hospital wide simulation project to model patient journeys from start-to-end. The complexity of the hospital structure could be seen as one explanation for this deficiency. Gunal (2012) suggested the complexity of the hospital system could be reduced if the modelling were done by building a unit-by-unit model. In order to undertake a simulation project involving the complex patient journey from start-to-end, it is essential to understand the underlying process model. Process models are the core component of any process improvement simulation project; therefore it is essential for the models used for simulation to give a close reflection of reality. Creating a simulation model to depict reality by hand is a challenging task especially in a complex healthcare environment where the system is prone to numerous process variations. Conceptualising the model is the blueprint for simulation projects and this should be independent of the simulation software (Gunal, 2012). The conventional ways of creating process models by hand often ignore event data and are error prone. This can lead to false conclusions. Process mining can resolve this issue because models are extracted from events that have already taken place giving a close reflection of reality (van der Aalst, 2011).

The proposition made in this research was to use process mining techniques for the discovery of process model/s using historical event data and thus the foundation for a simulation model can be formed. This model would give a closer reflection of reality.

This research introduces a novel way of deriving the event logs for process mining as discussed in Chapter 4. The derived event log is used to discover the process model that is validated by the clinicians as an authentic representation of the patient journey from start-to-end. The discovered process model could then be used with any simulation software either as an input or as the conceptual model.

This is also the first research to apply the control flow perspective of the process mining framework using the ProM tool for hospital-wide patient journey modelling. A process model/conceptual model of patient journey from start-to-end was discovered which could be used as an input for a simulation model.

8.2 Aims

The aim of the research was to automate the discovery of the conceptual process model by applying the control-flow perspective of the process mining framework. The historical, derived event log was used to develop a petri net based simulation model.

8.3 Method

The analysis was undertaken on inpatient records for patients admitted and discharged by the GM unit at FMC and a specialty unit, the Cardiology Unit. The two largest medical inpatient specialties at FMC are GM and Cardiology. The typical patients admitted to each of these units differ in terms of their age, complexities of disease and diagnosis. The GM service looks after a wide variety of diagnoses. Cardiology is a specialty unit looking after a limited number of specific diagnoses although it treats more patients than other medical specialty units.

As ED overcrowding intensifies so do the varied strategies proposed to ease ED overcrowding Paul and Lin (2012). Many studies have looked at factors that could reduce inpatient ED boarding time. If this boarding time exceeds eight hours, "access block" is present (Levin et al. 2008, Harris and Sharma 2010). Similarly, there are many studies that have explored waiting times (*time from triage to time seen by doctor*) and overall ED LOS (*time from triage to the time of discharge*) as reported by (O'Brien et al., 2006). There has only been limited study on the interaction between the various ED time intervals (Casalino et al., 2013). A recent study, subdivided the ED LOS into 5 categories (*bed allocation, doctor assessment, doctor decision time, consult time, emergency inpatient time*) and concluded that consult time contributed significantly to ED LOS (Brick et al., 2013). The moves towards more efficient processes, practices and behaviours within and outside the ED are warranted to deal with ED overcrowding (Dellifraine et al., 2010).

8.3.1 Process mining – control flow perspective

The process mining perspective presented in this Chapter is the process perspective. The goal of a process perspective is to focus on the control flow or the ordering of activities with the intention of discovering all possible paths (Weijters et al., 2006) . The use of process mining techniques in the healthcare industry is becoming increasingly widespread. Mans et al. (2008) used process mining techniques to identify bottlenecks and to better understand the different clinical pathways taken by various groups of patients. The key to producing a good simulation model is to first understand the model with all possible variations which cater for all scenarios. This approach produces a model that will be as close to reality as possible. A model accounting for all variations would help the clinicians perceive the entire process as it has taken place. Once the entire model is produced, it can then be manipulated to represent major behaviours of the system.

For a complex process such as the healthcare environment, formulating a process model close to reality is far from trivial. Using historical data to derive a process model for simulation is advocated. The main software used for Process Mining is ProM. The constructed event log was pre-processed into the MXML format required as input to ProM. This conversion was done using the Disco software package (Fluxicon, 2012).

8.3.2 FMC's admission process

FMC offers both inpatient and outpatient services. Outpatients are seen during business hours in the outpatient clinics and sometimes these patients might be admitted, urgently or otherwise, to hospital as inpatients. Inpatients can be categorised into two streams: those who enter the hospital as an emergency admission (usually via the ED) and those whose admissions are pre-planned (usually as an elective surgery stream). Both streams of inpatients affect the hospital occupancy.

The time patients spend in the ED can be categorised into three distinct processes. The phases related to these processes are categorised as waiting to be seen by an ED clinician (FMC-WTS), assessment and treatment time (FMC-RT) and boarding time, where admitted patients await a bed within the hospital proper (FMC-Boarding). The time patients spend in the FMC-WTS phase and overall in the ED (= FMC-WTS + FMC-RT + FMC-Boarding) are measured and reported by the hospital because these times are part of the hospital's KPI register. All the three phases of time take place while the patient is within the ED.

The flow chart in Figure 5-3, as illustrated by the clinicians, is the reflection of the three ED phases portraying how patients flow through the ED and either end up as an inpatient or are discharged from the ED. The flow chart was used to verify the discovered process model/s and could also be used for conformance checking.

8.3.3 Process information from event log

van der Aalst (2011) makes the following assumptions about event logs: a process consists of cases, a case consists of events such that each event relates to precisely one case, events within a case are ordered, events can have attributes. The derived event log of the patient journey process conforms to this assumption.

The cases in the patient journey process are individual journeys. A patient is identified using the URN field which is one of the attributes of the case. A patient could have multiple journeys or cases (multiple admissions). As this particular event log is derived to construct the inpatient journey from ED to discharge, the journey starts from the point of triage. Once the ED processes are completed, patients are then moved to either an appropriate ward or to any available ward. In the context of this event log, an activity relates to the patient moving within the predefined structured process (e.g. progressing from one activity/phase to another within the ED followed by the process of ward movement). Each event or activity is uniquely identifiable by the EventID field. Each ward within the hospital has a predefined activity with regard to the patient's care. Generally, an inpatient changes ward if he/she is originally transferred to an outlier ward. Each event or activity (e.g. ward movement) is recorded with a timestamp (the DateIn and DateOut) field. The timestamp of the event reflects the order of patient movement.

8.3.4 Process mining – Heuristics Miner - algorithm

The GM patient journey flows from the ED to the various wards outside of the ED. This is a structured process. Based on this knowledge, the patient journey flow process was characterised as a "Lasagne Process". In a "Lasagne Process" most cases are handled in a structured and pre-arranged manner (van der Aalst, 2011). For example, certain pre-conditions have to be satisfied before the patient can move to the next activity/phase. The process mining algorithm within ProM, namely the Heuristics Miner was used to discover the control flow of the patient journey process from admission to discharge. It is important to note that the order of activities within each case (*each admission to discharge*) is important because this information is used to calculate the order of activities (*the order of ward movement*). In other words, the algorithm relies heavily on the timestamps (*DateIn and DateOut*). The Heuristics Miner-algorithm is briefly described to set the context for the models presented in Section 8.3.4. The finer details of the workings of this algorithm are addressed in “*Process Mining with the Heuristics Miner-algorithm*” Weijters et al. (2006) and are outside the scope of this thesis.

The control flow process model is constructed by analysing for causal dependency. In this context the event log is analysed to see whether a patient staying in a particular ward always moves to another particular ward. If this movement frequently occurs then there is a causal dependency between these two wards. As described by Weijters et al. (2006), the dependency graph is constructed by:

Deriving a frequency based matrix to indicate the certainty of dependent relationship between event A and B (notation $A =>_W B$). The result of this is used to build the correct dependency relation. The value of the dependency relationship is always between -1 and 1. A value close to 1 indicates that there is a dependency relation between event A and event B (ward A and ward B).

Example 1: In a log where there are five traces, where activity A is directly followed by activity B, the value $A =>_W B = 5/6 = 0.833$. This indicates that the dependency relationship is not too strong. This is also assuming that the opposite direction where activity B followed by activity A will never occur.

Example 2: In a log where there are fifty traces, where activity A is directly followed by activity B, the value $A =>_W B = 50/51 = 0.980$. This indicates that the dependency relationship is very strong. This is also assuming that the opposite direction where activity B followed by activity A will never occur. Where there is noise when activity B follows activity A once, the $A =>_W B = 49/52 = 0.94$ indicating that there is also a strong dependent relationship.

8.4 Results

Process mining is an iterative technique. The main challenge was establishing boundaries for the underlying processes surrounding important hospital KPIs. Identification of these KPIs was helped by the clinicians identifying the relevant data set needed to be extracted from various systems used by the hospital. Using these data as a plug-in to the derived event log, a specific event log was constructed.

Once the event log, the most fundamental element in a process mining activity, is available, this event log could be used for knowledge discovery by applying various process mining algorithms available within ProM. The result presented below is specific and limited to the discovery of a simulation process model and its corresponding analysis. Section 8.4.1 below presents the descriptive statistics in relation to the FMC ED and the processes under investigation.

8.4.1 Descriptive Statistics

Descriptive statistical analysis of the overall FMC's patient dataset shows predictable patterns. Various statistical analyses are already being carried out to improve the efficiency and quality of patient care in general. Previous work relating to QoC received by inlier and outlier patients has been addressed in (Perimal-Lewis, 2013). Figure 8-1 and Figure 8-2 show the daily differences in waiting time to be seen by an ED clinician and in the number of patients within the ED at the time of any patient's triage.

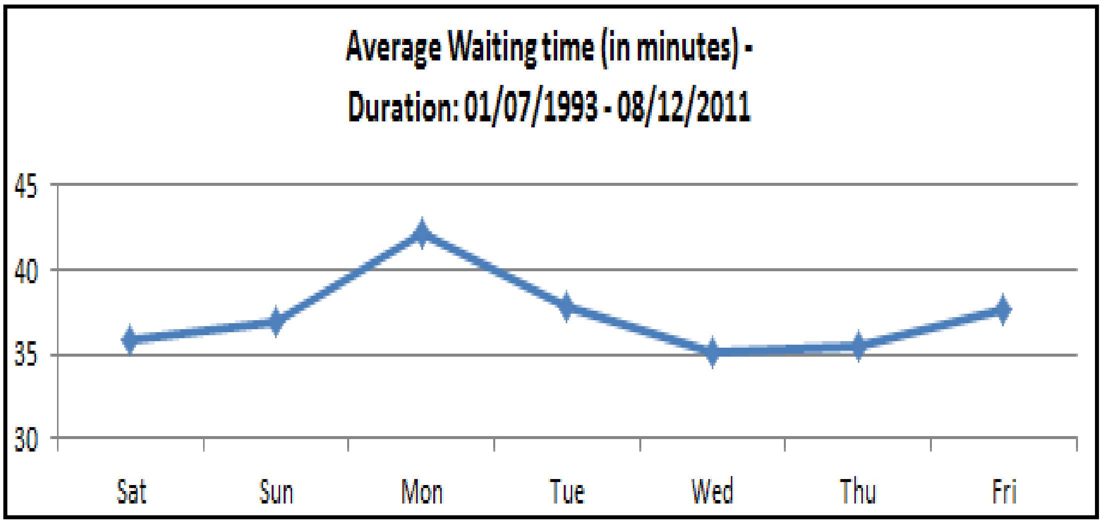


Figure 8-1: Trend in average waiting time (FMC-WTS)

Average waiting time is the time patients spent in the FMC-WTS phase. This shows an association with the patient count in the ED at the time of triage. On Mondays, when there are more patients in the ED, the average waiting time also increases for patients.

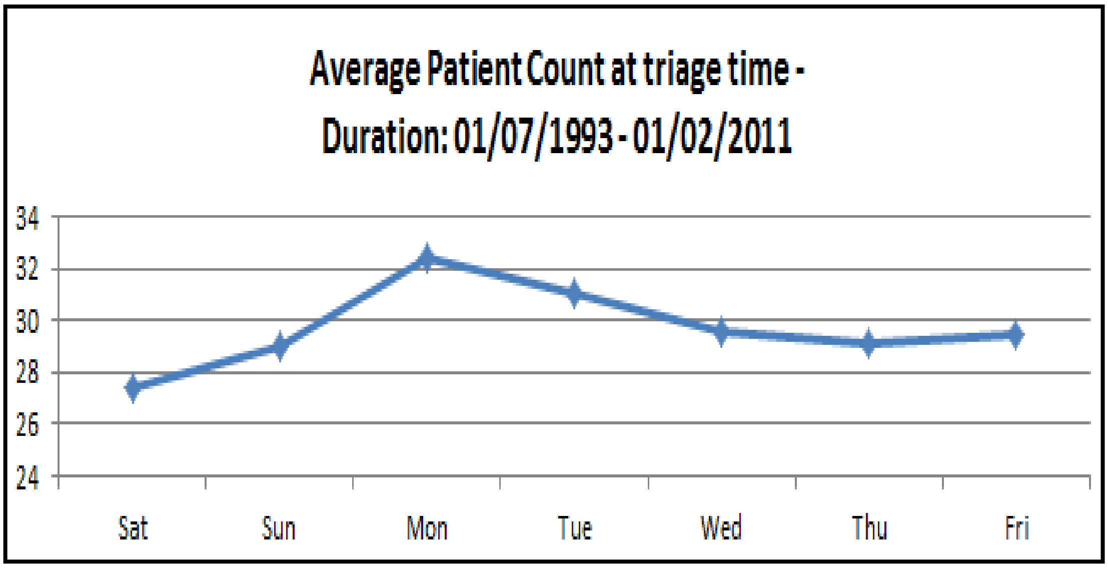


Figure 8-2: Average patient count at triage time

Based on the data captured for the various ED KPI reports, FMC is already undertaking sophisticated statistical analysis for different prediction models. However, FMC is still experiencing ED congestion and access block. Process mining could be used to complement the already existent mature statistical analysis. This might improve one's understanding of these complex ED processes beyond the interpretation of an aggregate statistical analysis. By undertaking process mining, it is possible to dive deeper into the processes underlying the inpatient journey or patient flow from admission to discharge. The next section discusses the heuristic models for GM patient and the Cardiology unit. Similar models could be built for other units as required.

8.4.2 Control flow perspective – heuristic models

Because the models discovered are the base for simulation model/s, it was essential to choose a timeframe within the data set that reflected relative stability of all processes within FMC. Deriving this timeframe was done following close consultation with clinicians. The timeframe used for these models was between 01/01/2007 and 31/12/2009.

Figure 8-3 shows the heuristic model for the cardiology unit. The model is a less complicated model compared to the GM patients' model which will be discussed next. For the purpose of simplicity the model is a reflection of cardiology patient flow from 01/01/2007 – 21/12/2007 only, with further filtering of records to show only those patients who received 100% of their care from the Cardiology team. This means that these patients would have received care throughout admission from the same team of doctors. This model of care is preferred because fewer unit (*team of doctors providing care*) changes imply the patients receive less interrupted care. The model was verified by clinicians to be an acceptable reflection of patient journeys within the Cardiology unit. The weightings next to the arcs between the wards indicate whether there is a strong or weak dependency between the wards as described in the previous section. As reflected in Figure 8-3 the dependency relationship is not too strong because only a small subset of patients are modelled.

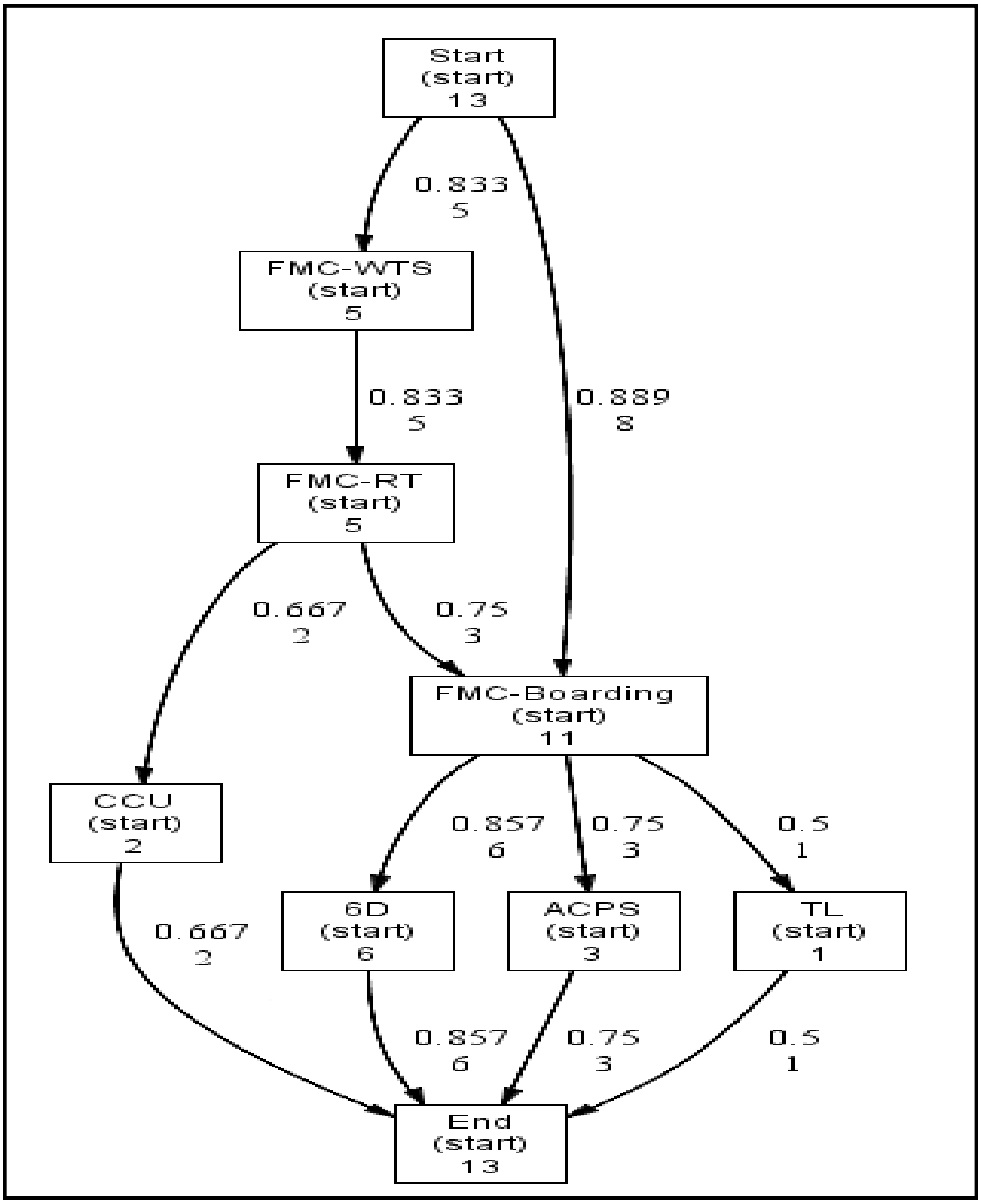


Figure 8-3: Cardiology patient journey

The next model shown in Figure 8-4 is the first model discovered for the overall GM inpatients' journey. As stated before, the model building exercise was an iterative process. The first model derived is as represented in Figure 8-4 which had high variation. Although the processes are well defined, there were high variances in the event log. This was acknowledged and explained by the clinicians as the nature of flow for the GM patients. There will always be patients presenting to the hospital with a unique characteristic that would require the patient to follow a unique path. The discovery of the complicated model confirms the perception of the complex nature of GM patient journeys. Revealing the complexity of the GM patient journey was an important exercise. However using a model with such high variation will not be beneficial when deriving a simulation process model. The GM patients' journeys will always have high variation. Therefore, in this situation a model that portrays the majority of the patient journeys will be preferred. Further analysis of this model using a 'Performance Sequence Diagram' within ProM revealed that there were over 2000 path patterns and many one-off paths. One-off paths do not show the main behaviour of the system. The domain experts verified the model and confirmed the validity of the variations. However modelling the paths that reflected the common behaviour of the system was deemed important in order to identify paths or patterns with high throughput and paths that could contribute to bottlenecks in the system. With this notion, a second model for the GM inpatients was developed as shown in Figure 8-5 and for better readability a small section is of the model is shown in Figure 8-6.

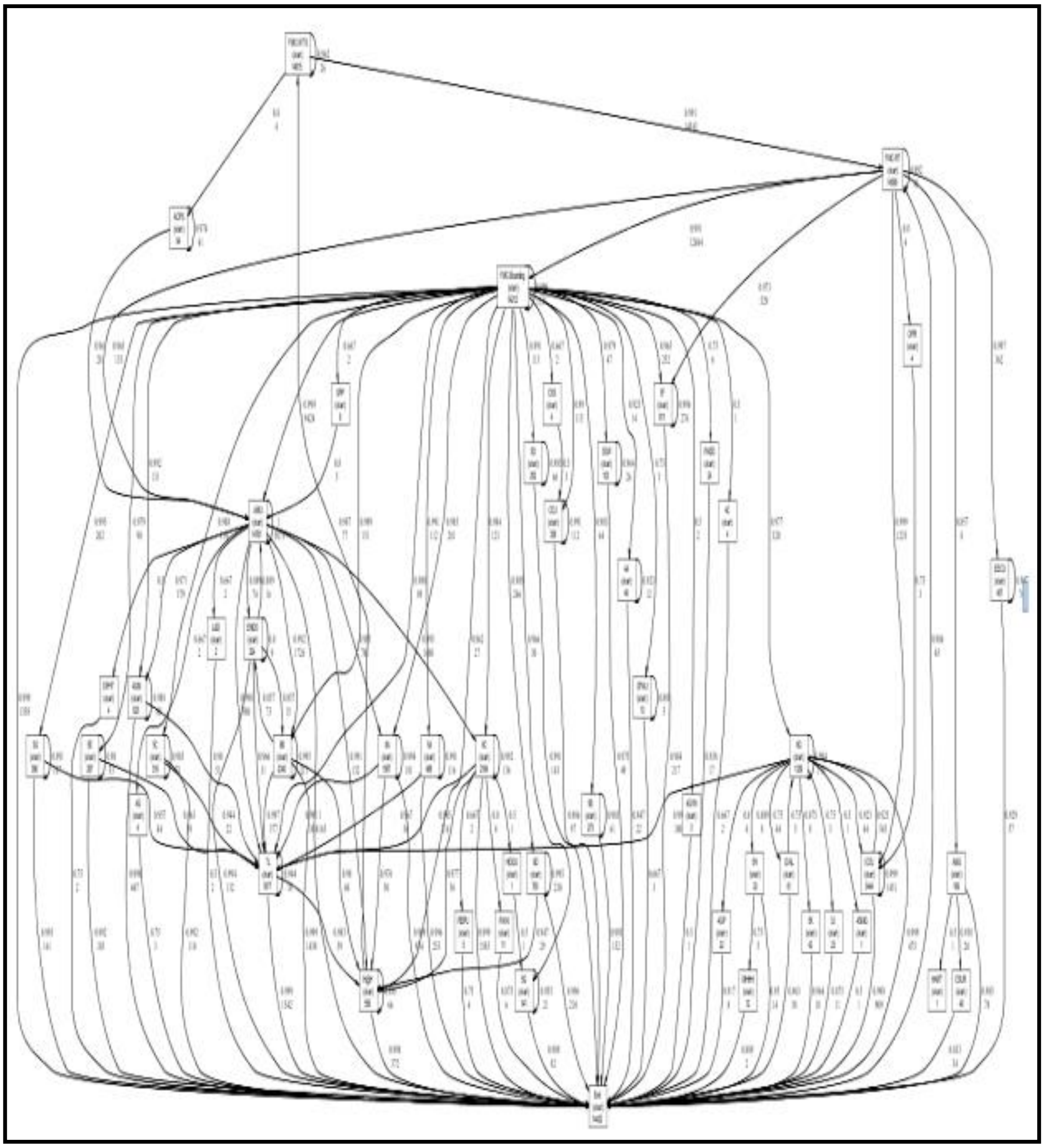


Figure 8-4: Complexity of first patient journey process model for GM patients

The model in Figure 8-5 shows the second heuristic model for the GM inpatients similar to the previous model. However this is a less complex model. The model presented is the final and most representative model. The model is based on GM inpatients where the sequence of activities (the path) is shared by at least 10 cases. This means that journeys with a complete path from start to end that appeared less than 10 times were filtered out in order to produce a model that is interpretable in a complex setting such as the hospital. The new model accounted for 75% of the GM inpatients and shows 113 patterns rather than over 2,000 as in the previous model. The model was verified to be a good reflection of the GM patient journeys by the domain experts. Similarly, a model representing 80% of the population could be derived and validated by the domain experts if necessary.

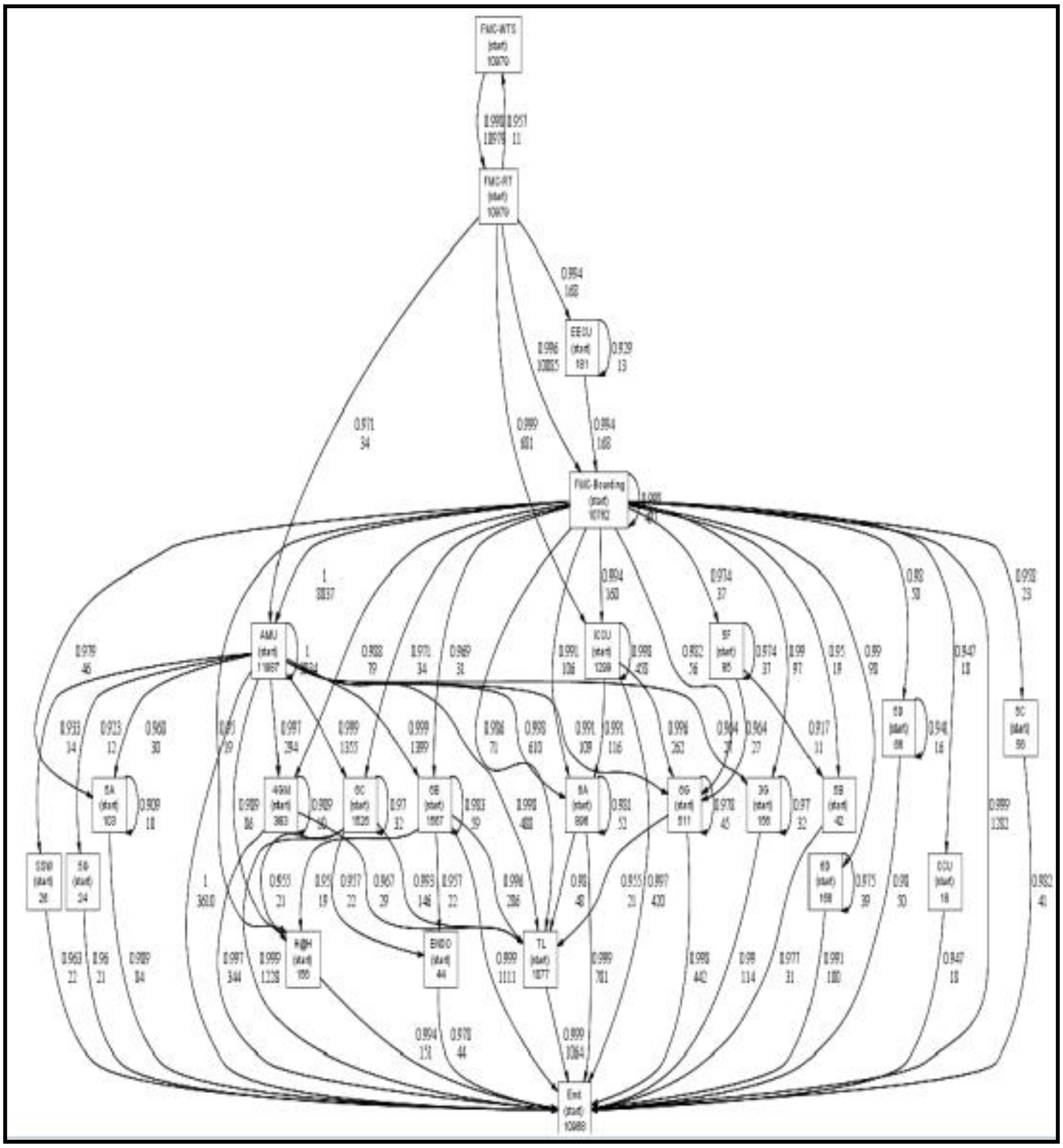


Figure 8-5: Complexity of the second patient journey process model for GM patients

The second heuristic model in Figure 8-5 and the snippet of the model in Figure 8-6 are also close representations in conformance with the GM patient journey reflected by the clinicians. This knowledge helps enforce the validity of the discovered model. As well as validating the process model as depicted by the clinicians, the discovered model also revealed other ward movement patterns which were not expected by the clinicians. The model also helped identify potential deviations in the process which could be attributed to data entry errors. For example patients should not be moving back to FMC-WTS phase from FMC-RT phase as reflected in the models. However this only accounted for a very small percentage of patients.

All the causal dependency values for the first and second GM patients' models were more than 0.9 indicating that the dependency relationships between wards are strong, inspiring more confidence in a particular pathway as being a common feature of the GM patient journey. Other movements were verified correct by the clinicians as valid GM patient journeys. Based on the discovered process model, it was also possible to verify the inlier and outlier wards where GM patients were admitted. Further analysis, which is beyond the scope of this research, could be carried out to analyse the characteristics and outcome of patients following a path consisting of mainly outlier wards as opposed to paths consisting of mainly inlier wards.

The model also depicted wards with a high percentage of unit changes. These wards were verified to be wards where the care of a patient might be transferred to another team. Such care transfer might occur because either the patients were wrongly diagnosed initially and hence "sorted" into that more appropriate second unit. Alternatively the second unit depicted often offered a higher acuity of care and, following a significant deterioration in patients' condition, a change in the team of doctors looking after those patients may have been required.

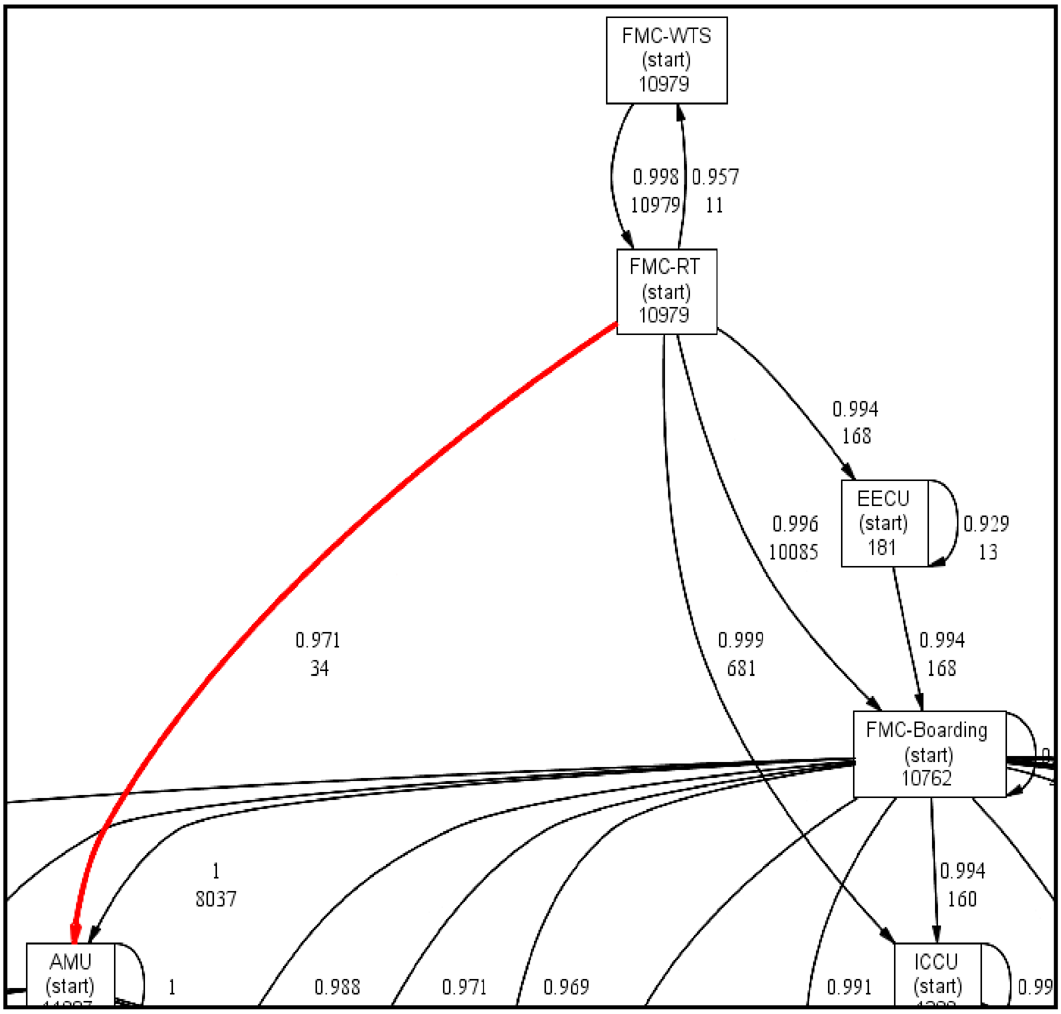


Figure 8-6: Snippet of the second patient journey process model for GM patients

Finally, once the verification of the models had occurred to the satisfaction of all concerned, these models were converted into a Petri Net model. Figure 8-7 is a snippet of a Petri Net model which was derived from the second patient journey model for GM patients shown in Figure 8-5. The Petri Net model can now be exported into a simulation tool such as the Coloured Petri Net (CPN) tool for simulation.

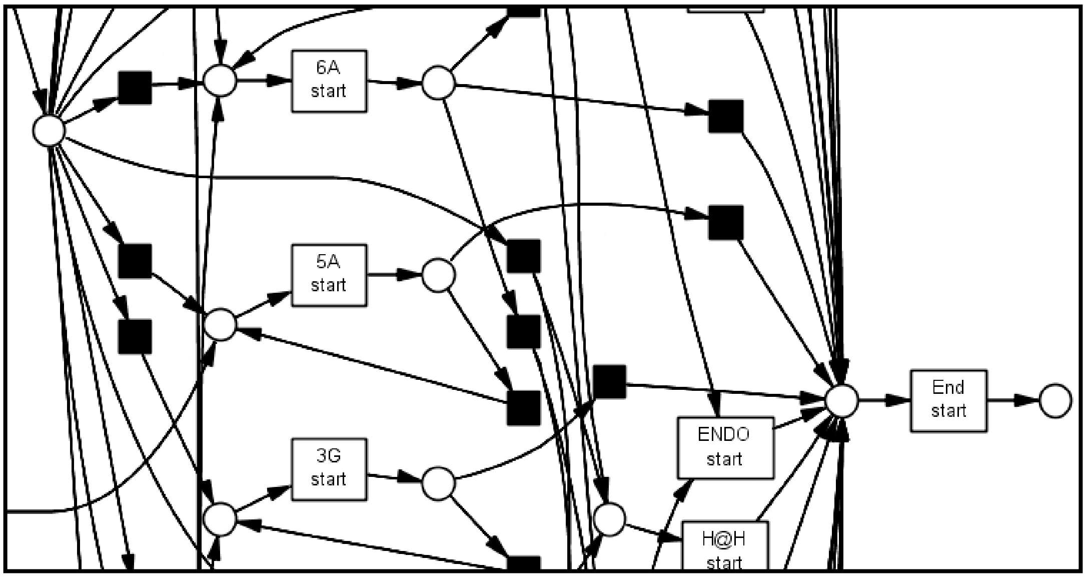


Figure 8-7: Snippet of Petri Net for GM patients

Coloured Petri Nets (CPNs) are a discrete-event modelling language for modelling systems where concurrency, communication and synchronisation play major roles (Jensen et al., 2007). The CPN process model that will be used for simulation is a sound model derived from historical event logs which, in turn, had been validated as a close reflection of reality by clinicians. As a result, the discovered patient journey process models inspire great confidence in any resulting outputs of a simulation exercise. After all, these models are based upon event data that has already taken place. Also the process model reflects the main behaviour of the system and reduces the chances of excluding certain activities by mistake as might occur when constructing these models by 'hand'.

8.5 Discussion

It is important to be mindful of the scope and boundaries needed for the data because otherwise big data files are involved. These files, widely available in health sector, could pose not only technical difficulties requiring high end computer processing power but also could produce models that are not interpretable. Undefined scope for process mining in healthcare could lead to discontinuation of such projects. Therefore collaboration with the clinicians should start at the very inception of the project and continue at every stage of the project. Hospital managers may also have an equally important role to play in the execution of these projects.

Compared to inpatients of a single specialty unit, the inpatient characteristics of GM patients are complex and are non-deterministic. Therefore, studying and understanding the underlying processes of the GM patients, although challenging, will reveal insight to a wider spectrum of behaviours within the patient journey process. Insight from domain experts is critical for successful application of process mining in healthcare settings. Choosing a diverse patient group and then separately focussing on single specialised unit with a less diverse patient clientele is also an effective strategy for interpreting large hospital datasets.

A distinction must be made to differentiate the characteristics of the event log and the characteristics of the process being mined. The event log used is characterised as unstructured or semi-structured, however the process being mined is characterised as being structured. This distinction is important in deciding the appropriate control flow algorithm to use within ProM. The process is structured because there are pre-defined activities and criteria that take place under each phase, all following a structured process. For example, patients within the FMC-WTS phase of their care in the ED are treated according to the Australasian Triage Scale (ATS) which is a measure of urgency for that patient to be seen by a clinician.

The simulation model discovered was for a specific group of inpatients. Similar models could be discovered for other groups of patients.

8.6 Conclusion

Process mining in the healthcare domain is an extensive and time consuming exercise. For process mining activities to be successful, the stakeholders involved need to perceive the advantages of using process mining for gaining health intelligence. Health intelligence is gained by diving deep into a process for knowledge discovery beyond what is offered by statistical analysis alone. The complex movement of patients shows that patient journey analysis using a statistical approach combined with process mining techniques will give better insight into the intricacies of a complex healthcare system.

The next challenge is to work closely with the clinicians, and possibly with managers as well, to identify key process improvement areas. For healthcare, these areas are normally areas where performance is measured by pre-agreed KPIs. This will lead to the identification of appropriate data which will be used as an input for process mining activities. The identification of appropriate data and then the process of pre-processing the data to derive the event log are most crucial and time consuming activities. In most healthcare settings within Australia the absence of Process Aware Information Systems (PAIS) means that a resourceful way of deriving these event logs is needed such that process mining activities can succeed. This thesis has presented one such method for deriving event logs in the absence of a data warehouse or PAIS. In such instances, using the available data destined for

KPI reporting could be a starting point. Since similar KPIs are used in all public hospitals in Australia, this method is generalizable. A similar approach could be used for other hospitals that are keen to embrace process mining and to dive deeper into their processes when embarking upon process improvements.

At FMC, process mining offers added benefit to the already successful implementation of “lean thinking” and enhances the areas where a “lean thinking” approach alone is inadequate. Process mining can reveal previously hidden insights into access block and bottlenecks. Therefore constructing a process model for patient journeys from start-to-end as a base for a simulation model derived from historical event logs and having that model validated by the clinicians is advocated as a sound starting ground for future simulation projects.

9 Conclusion

9.1 Introduction

Hospitals have to comply with strict measures of operational efficiency and effectiveness by conforming to KPIs. Hospitals have mature processes that collect data on various quality measures in order to report and adhere to these KPIs. As an example, hospitals aim to improve ED throughput measures by reducing the time a patient waits until first seen by a doctor, reducing did-not-wait rates and by reducing ED LOS for all patients (Shetty et al., 2012). Patient ED LOS or total inpatient LOS is one criterion that can be used to measure ED performance and a hospital's performance in general. Performance is usually measured as the percentage of patients who stay beyond an established LOS target (Kolker, 2008). Inpatient LOS has become one of the many ways used to measure performance of a hospital. Patient mean LOS has been used to measure QoC and hospital efficiency in terms of resource usage (Thomas et al., 1997). A LOS shorter than a defined “normal” LOS could indicate that hospitals are discharging patients early possibly sacrificing QoC (Thomas et al., 1997). Aiming to comply with an ED LOS target could possibly contribute to the streaming of patients to any available ward rather than the home ward for each patient.

The complexity and diversity of hospital processes mean there are also diverse ways to measure the quality of patient care. Quantifying the quality of patient care varies based on the characteristics of the process area being studied. This research investigated the QoC received by patients who were admitted to their home ward (referred to as inliers) and patients who were admitted outside of their home ward (referred to as outliers). It is a common, albeit poorly substantiated, perception amongst clinicians that outliers have longer overall in-hospital LOS compared to inliers. It is also perceived that the QoC received by outliers is inferior to that of inliers. At FMC where this study was undertaken, the percentage of outlier patients is a regularly reported hospital performance indicator and therefore substantial effort is taken to collect accurate data regarding the inlier or outlier status of every admitted patient.

Several studies have established an association between ED overcrowding and in-hospital mortality. Richardson (2006) reported increased in-hospital mortality at ten days amongst patients presenting to the ED during high ED occupancy. The hypothesis was that more patients would end up in outlier wards during ED overcrowding due to managerial pressure to reduce ED congestion by transferring patients to any available bed rather than wait for a bed in the home ward to appear. The hypothesis also proposed that outliers would have longer overall LOS because their stay was prolonged as a consequence of being admitted outside of their home ward and therefore not receiving the required level of care. Moreover, as a consequence of inferior QoC received by outliers, the outlier group might also have a higher in-hospital mortality rate.

This research was undertaken at a large public hospital in Australia and investigated the consequence of streaming patients to any available wards. The research looked at the QoC of patients admitted either as outliers or as inliers. It was established that a clear cut definition of outliers and inliers was not available. Previous studies that have looked into outliers and inliers were limited; looking at small populations. These studies classified any patients who stayed in an outlier ward as outliers regardless of the proportion of the whole admission time spent on those outlier wards. All of these studies, including a recent study undertaken by Santamaria et al. (2014) showed that outliers had a worse outcome. However, the complex structure of the hospital and its administration means that most patients move eventually from an outlier ward to an inlier ward if the duration of the admission is long enough. Therefore the time a patient spends in an outlier ward and then in an inlier ward could vary from patient to patient. The outcome of a patient who spent the majority of their stay in an outlier ward could be different to a patient who only spent a small proportion of their stay in an outlier ward. A blanket definition of outliers based on any time spent in an outlier ward can be considered and might be relevant when undertaking the analysis on a specialty unit. However this definition needed further tightening for a diverse group of GM patients. By diving deep into the real patient data supplied by the hospital, it was established that the GM population is not only diverse but also has many ward movements. The derived event log especially the timestamp field used for process mining enabled a holistic view of the patient journey depicting ward movements between inlier and outlier wards. Therefore in this research the definition of an outlier and inlier was based on the distribution of time each patient spent in an outlier ward and in an inlier ward. The ward inlier and ward outlier work could be extended by studying the outcome of medical outliers in surgical wards, the outcome of medical outliers in 2nd best medical wards, the outcome of surgical outliers in medical wards and the outcome of surgical outliers in 2nd best surgical wards.

Having looked at the QoC of inliers and outliers, both in terms of the efficiency of care and effectiveness of care, this research examined the attributes relating to efficiency of care. Efficiency of care is one of the overall QoC attributes established in this research. The impact of the time spent in the ED was related to the efficiency of care for the whole hospital admission of each patient. Most ED studies investigating on the effect of a prolonged ED time on QoC concentrate upon the boarding time (the time an admitted patient waits in the ED for a bed to become available). The effect on QoC on the rest of the time a patient stays in the ED has not been widely studied. The effect of 'disposition time' in the ED on mortality has only recently been addressed by Mitra et al. (2012). In this research, the ED time was compartmentalised into triage-to-admit (disposition) time and boarding time. The influence of these different times on overall efficiency of care was assessed. Once again the derived event log especially the timestamp field used for process mining revealed the value of dividing ED processes into phases or compartments. The independent influences of each of these phases could then be assessed using statistical techniques. The time spent in each of these ED phases could be applied when studying the influence of delays during each ED phase on other QoC attributes such as the inpatient LOS and mortality.

Finally, to enable the hospital to take advantage of the well-documented benefit of simulation, in particular the advantages of DES for decision making, this research looked at automating the discovery of process models based on historical data. It modelled the patient journey from start-to-end for the discovery of a process model. The process model could then be used as an input for future simulation projects. This research strongly advocates the use of automated discovery of process models for simulation to avoid using handmade models which are prone to errors. As reflected in this research, a patient journey through the hospital is complex; therefore creating a model from historical data is strongly advocated. The patient journey control flow models, discovered using process mining, should be used as a starting point for a simulation project. Models from the organisational perspective and from the control flow perspective could be merged to form an all-encompassing simulation model.

All the above achievements in this research were undertaken by using an innovative way of applying the process mining framework to healthcare. This same framework, having been successfully applied to gain insight into complex hospital processes, could be adapted for less intricate healthcare settings such as primary care, aged care and public health. This is the first systematic, large scale research of the application of process mining in healthcare. Most scholarly research in process mining has only looked at a specific unit or department in the hospital. This research looked at hospital-wide patient journey modelling. To undertake this hospital-wide patient journey modelling an innovative framework for deriving the event log was established using the concept of plug-in. The derived event log for all the case studies initially consisted of the bare minimum data requirement and then plug-ins were added as needed. Hospital-wide patient journey modelling to assess QoC was undertaken by building on the three perspectives of the process mining framework: a case perspective, the organisational perspective and finally the control flow perspective. The research was undertaken by adapting the novel and relatively young process mining framework and by complementing this framework with a mature statistical modelling approach.

An event log is the cornerstone of process mining so a novel way of data collection around KPIs was used and is advocated. This simplifies the tasks of scoping a large project within a complex healthcare setting. Another advantage of using KPI data is that the data quality is more reliable. These KPI data are used for government reporting and government funding requires these data to be at certain compliance levels. The developed framework will be generalizable in all public hospital settings because it uses the measured hospital KPIs as a starting point.

9.2 Summary of contribution

This is first study using an innovative way of applying the process mining framework to a complex healthcare environment in order to undertake hospital-wide patient journey modelling. Data or an event log is the most fundamental element needed for process mining. Most applications of process mining have been unit- or department-specific projects. Therefore it is feasible to acquire a useful event log from a PAIS already used within the unit. The event log, because of its abstract nature, does not necessarily have sensitive patient information. However, when undertaking hospital-wide patient journey modelling, the unit specific event logs do not have the capability needed to trace each patient journey outside of one unit or department. This is when a derived event log can be used as a substitute for an automated event log. The derived event log enables the research to take advantage of the existing process mining algorithms embedded in ProM. The next challenge is to develop a framework that can facilitate adequate data collection needed for hospital-wide patient journey modelling. This research therefore developed an innovative way to derive the event log. It introduced the concept of data plug-ins. The data plug-in is added to the bare minimum event log. The concept of data plug-in enables unit-specific QoC to be added and mined. The bare minimum event log together with the added plug-in makes the outcome of the modelling activity relevant for units as well as the hospital in general. The bare minimum event log could be derived from a patient tracking or logistics system if one is available. If a patient tracking system is not available, the bare minimum event log could be easily derived from stored patient data using a patient identification number, admission timestamp, ward name and a derived unique patient journey identification counter. The unique patient journey identification counter should be unique for each admission. Despite the complexity of scoping the patient journey modelling project in a complex healthcare environment, data collection was facilitated by innovatively scoping the project around hospital KPIs. Using KPI data does not guarantee 100% accuracy but it does guarantee consistency of reporting and analysis outcomes. Therefore, the knowledge discovered could be used to undertake process improvement activities for a specific KPI.

The derived event log was used to undertake the following studies:

The first systematic evidence-based research was performed on hospital team-based and ward-based models of care and the impact on QoC of patients being either ward inliers or ward outliers. This research looked at a large number of complex GM patients spanning over six years. A novel framework was established to classify these ward inliers and ward outliers based on the distribution of time each patient spent in a home or outlier ward.

The research then investigated ED time by breaking it into two distinct phases in order to study the effect of each ED phase of care upon efficiency of care. The research introduced a surrogate measure to investigate the influence of the organisation on delivering efficient care. The surrogate outcome measured the efficiency of care depending on whether the patient was admitted inside or outside working hours.

The research also established a framework for automating the discovery of a patient journey process model from start-to-end. The process model can be used in simulation studies when modelling a complex hospital environment. This will minimise errors prone when applying handmade models. The research demonstrated the possibilities of applying the control flow perspective of the process mining framework to discover a hospital wide process model. The innovative way of deriving the event log necessary for undertaking process mining contributed to the success of this study.

The research conducted a practical evaluation of the applicability of a current ProM tool to a real healthcare setting. The clinicians were given the opportunity to reflect upon the usefulness of the ProM output. In some cases, the feedback and the insight offered by the clinicians were the driving force giving insight to the next aspect of the data investigation. Some ProM outputs were elementary and were not seen to be adding value to the already mature results produced by the clinical epidemiology unit at the hospital. The ProM output that was most beneficial was the insight gained from the ward movement pattern and the discovery of a process model for simulation. The case perspective and the organisational perspective of process mining framework were undertaken using statistical tools outside of ProM.

Finally, this research contributed to the advancement of health informatics by applying process mining techniques, which are still in their infancy, to intricate healthcare scenarios within complex hospital processes.

9.3 Hospital process accreditation

Literature on hospital accreditation seems to be based of Health Standard Quality Accreditations. In Australia, the Australian Council for Healthcare Standards International (ACHSI) and the Australian Commission on Safety and Quality in Health Care undertake hospital accreditations. The quality assurances in the healthcare industry are mainly patient-centred and are aimed to improve QoC. Brand et al. (2008) stated that the best model for healthcare clinical governance is yet to be identified. Hospitals work very hard to keep up-to-date with KPIs and meet the established healthcare standards. In Australia, hospitals' efforts to maintain and improve on their performances are commendable. The Australian Government Department of Health and Ageing (2010) points out that, as of 30th June 2009, 87% of public hospitals were accredited. This amounts to 97% of available beds.

Although many hospitals are accredited for healthcare standards, there is no relevant literature to support how processes within the hospitals are accredited for their maturity. A process becomes mature when the process is continually being challenged and tested and, having demonstrated its full potential at the present level, the process is then accredited to progress to the next level. CPR takes a holistic approach by looking at a wide area for process redesign. Process improvement is a vital element in any system and requires continuous fine-tuning and adjustments in order to adapt continuously to the ever-transient nature of the hospital system. Therefore, what is needed is a framework to ensure that any positive outcome gained from undertaking CPR is not only sustained but a benchmark is established for the process to continually mature. Hospitals are similar to any other organisation where there are many interconnected processes and departments working together towards one goal. Hospital processes could be benchmarked learning from the successes of the IT / software industry in implementing and using benchmarks such as the Capability Maturity Model (CMM) and the Information Technology Infrastructure Library (ITIL). It should also be noted that, although these benchmarks have a set framework to assess maturity of the process, the concept behind the framework is for organisations to derive a quality model. The derived quality model should best fit the hospital's goals. This will ensure the improvement process is focussed on what would work best for the problem at hand rather than adopting someone else's practices. In this case this requires formulation of a framework that will work for the healthcare industry which can be tailored to suit different hospital setting. A focus on measurements and continuous improvement will promote the sustainability of any process improvement initiative (Andersen et al., 2014).

Process mining is a promising technology and this research is an innovative approach of applying process mining to a complex healthcare environment.

9.4 Collaboration with clinicians

The strength of this work was from working in close collaboration with the domain experts. This gave clinical relevance to this research and is one reason the research is breaking new grounds in improving evidence-based patient-centred clinical care. A multi-disciplinary team approach to modelling is an essential ingredient for the success of any modelling research. Modelling cannot depict everything in a complex system such as healthcare but a systematic approach to modelling can depict the main behaviours of the system.

Mah (2009) asserted that providing good quality, efficient healthcare can only be accomplished if time and effort are given towards bringing staff together to examine the process of care delivery. The staff need to see the patient journey as a whole. Work done using the approach of Lean Thinking is focussed on the patient. Further analysis of this approach in conjunction with the insight gained from process mining (as demonstrated in this research) can offer added benefit to the already successful implementation of Lean Thinking.

9.5 Process mining in healthcare – final remarks

The most important contribution from process mining was the development of a novel way of adopting the underlying three perspectives: the case perspective, the organisational perspective and the control flow perspective. An understanding of this concept enabled process mining to be adapted in healthcare. This adaptation was independent of the process mining, ProM tool. The most mature process mining algorithm implementation within ProM is the control flow perspective. Once the derived event log was available, the research was able to take advantage of this feature.

Many authors have argued that the process models produced by the existing algorithms are not interpretable. Therefore, they conclude, the existing algorithms are probably not suitable for undertaking process mining in healthcare. These other authors have also argued that the existing models are not interpretable even by clinicians. This research confirms the findings from these previous studies. However, by practising an agile, iterative development of process models, this research was able to produce interpretable process models using the existing algorithms. The reason for this success came from the insight rendered by the clinicians.

Process mining can be a powerful tool when used to gain insight into processes surrounding a KPI. Otherwise the abundance of data and the complexity of healthcare data will produce complicated results. Process mining is not a stand-alone strategy. Its strength is in the knowledge discovery gained from pattern and process model discovery. These patterns and process models, once interpreted by the clinicians, can be further explored using pre-existing mature statistical modelling. In evidence-based clinical research, it is a common practice to use data spanning over many years. This practice is well suited for statistical modelling which relies upon aggregate data. Using an extended timeframe for process mining will produce patterns and process models that do not reflect the changes that have occurred in the hospital over an extended time period. This insight was revealed when working in close collaboration with the clinicians.

For process mining to be successful it has to be based around process improvement activities. This will allow the process mining project to scope data collection around KPIs. The insight gained from process mining, once it has been correctly interpreted by domain experts, can be directly used to improve hospital KPIs.

Finally, in addition to the advancement of health informatics, this research demonstrated the use of process mining to reveal new, previously concealed, complex hospital processes. The discovery of this abstract knowledge complements the aggregate results of stochastic analysis commonly used in epidemiology. Health intelligence therefore is the ability to develop insight from complex health processes. This intelligence, used in conjunction with mature stochastic modelling techniques, provides strong evidence to revolutionize health service delivery by initiating process improvement activities.

Appendices

Appendix A

Publications Resulting From This Thesis

The following publications resulted from this research.

Perimal-Lewis, L., Ben-Tovim, D. I., Li, J. Y., Hakendorf, P. H. and Thompson, C. H. (2014). "Relationship Emergency Department lengths of stay: characteristics favouring a delay to the admission decision as distinct from a delay while awaiting an inpatient bed." Internal Medicine Journal. (In Press).

Abstract:

Background: A prolonged stay for a patient within the Emergency Department (ED) can adversely affect the outcome of their ensuing hospital admission.

Aims: To investigate the characteristics of those eventual general medical hospital inpatients who stay in the ED awaiting a decision to be admitted and then await a bed.

Methods: Data from Flinders Medical Centre's patient journey database were analysed. The analysis was carried out on 19,476 patients admitted as an emergency under the General Medicine units.

Results: A less urgent Australian Triage Scale category significantly prolonged triage-to-admit time but did not affect boarding time. The decision to admit a patient took 29% longer for patients who presented to the ED outside of working hours. However a decision to admit taken outside working hours meant the boarding time was over three hours shorter than if the decision had been taken inside working hours. For every additional patient in the ED at the time of presentation, the admission decision was delayed by about half a minute. Every additional patient in the ED at the time of an admission decision increased boarding time by almost ten minutes.

Conclusion: Outside of working hours, patients presenting to ED have longer triage-to-admit times while patients for admission have shorter boarding times. ED congestion delays admission decisions only slightly and prolongs patients' boarding times to a greater extent. Strategies to reduce the time patients spend in ED should differ depending upon whether a decision to admit the patient has been reached.

Perimal-Lewis, L., Vries, D. D. and Thompson, C. H. (2014). Health intelligence: Discovering the process model using process mining by constructing Start-to-End patient journeys. Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand, ACS.

Abstract:

Australian Public Hospitals are continually engaged in various process improvement activities to improve patient care and to improve hospital efficiency as the demand for service intensifies. As a consequence there are many initiatives within the health sector focusing on gaining insight into the underlying health processes which are assessed for compliance with specified Key Performance Indicators (KPIs). Process Mining is classified as a Business Intelligence (BI) tool. The aim of process mining activities is to gain insight into the underlying process or processes. The fundamental element needed for process mining is a historical event log of a process. Generally, these event logs are easily sourced from Process Aware Information Systems (PAIS). Simulation is widely used by hospitals as a tool to study the complex hospital setting and for prediction. Generally, simulation models are constructed by 'hand'. This paper presents a novel way of deriving event logs for health data in the absence of PAIS. The constructed event log is then used as an input for process mining activities taking advantage of existing process mining algorithms aiding the discovery of knowledge of the underlying processes which leads to Health Intelligence (HI). One such output of process mining activity, presented in this paper, is the discovery of process model for simulation using the derived event log as an input for process mining by constructing start-to-end patient journey. The study was undertaken using data from Flinders Medical Centre to gain insight into patient journeys from the point of admission to the Emergency Department (ED) until the patient is discharged from the hospital.

Perimal-Lewis, L., Qin, S. and Thompson, C. H. (2013). Analysing homogenous patient journeys to assess quality of care for patients admitted outside of their 'home ward'. Seventh Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014). Gray, K. and Koronios, A. Adelaide, Australia, ACS. 142: 29-38.

Abstract:

This study is the first to explore the quality of care based on the outlier or the inlier status of patients for a large heterogeneous General Medicine (GM) service at a busy public hospital. The study compared the quality of care between ward outliers and ward inliers based on a homogenous group of patients using Two-step clustering method. Contrary to common perception, ward outliers had overall shorter Length of Stay (LOS) than ward inliers. The study also was unable to support the perception of shorter LOS in the outlier group being associated with higher in-hospital mortality. The study confirmed that overall the outliers received inferior quality of care as discharge summaries for the outliers were delayed and more outliers were re-admitted within 7 days of discharge in comparison to the inliers.

Perimal-Lewis, L., Li, J. Y., Hakendorf, P. H., Ben-Tovim, D. I., Qin, S. and Thompson, C. H. (2012). "The relationship between in-hospital location and outcomes of care in patients of a large general medical service." Internal Medicine Journal.

Abstract:

Background: The discrepancy between the number of admissions and the allocation of hospital beds means that many patients admitted under the care of a general medical service can be placed in other departments' wards. These patients are called 'outliers', and their outcomes are unknown.

Aims: To examine the relation between the proportion of time each patient spent in their 'home ward' during an index admission and the outcomes of that hospital stay.

Methods: Data from Flinders Medical Centre's patient journey database were extracted and analysed. The analysis was carried out on the patient journeys of patients admitted under the general medicine units.

Results: Outlier patients' length of stay was significantly shorter than that of the inlier patients (110.7 h cf 141.9 h; $P < 0.001$).They had a reduced risk of readmission within 28 days of discharge from hospital. Outlier patients' discharge summaries were less likely to be completed within a week (64.3% cf 78.0%; $P < 0.001$). Being an outlier patient increased the risk-adjusted risk of in-hospital mortality by over 40%. Fifty per cent of deaths in the outlier group occurred within 48 h of admission. Outlier patients had spent longer in the emergency department waiting for a bed (6.3 h cf 5.3 h; $P < 0.001$) but duration of emergency department stay was not an independent predictor of mortality risk.

Conclusion: Outlier patients had significantly shorter length of stay in hospital but significantly greater inpatient death rates. Surviving outlier patients had lower rates of readmission but lower rates of discharge summary completion.

Perimal-Lewis, L., Qin, S., Thompson, C. and Hakendorf, P. (2012). "Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach." Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012) Melbourne, Australia 129(ACS): 59-66

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.

Appendix B

List of Abbreviations

ATS - Australian Triage Scale

BI - Business Intelligence

BPM - Business Process Management

BRP - Business Process Re-engineering

CI - Clinical Indicators

CPR - Clinical Process Redesign

DES - Discrete Event Simulation

ED - Emergency Department

EDIS - Emergency Department Information Systems

EHR - Electronic Health Record

EMR - Electronic Medical Record

FMC - Flinders Medical Centre

GM - General Medicine

GP – General Practice

HI - Health Intelligence

HIS - Hospital Information Systems

ILS - Indoor Location System

IT - Information Technology

IS - Information Systems

KPI - Key Performance Indicator

NEAT - National Emergency Access Target

PAIS - Process Aware Information System

QoC - Quality of Care

Unit - A team of doctors

URN - Unit Reference Number

Appendix C

Glossary

Inside working hours: Between 0800-1800 hours from Monday – Friday.

Access block: Access block is when a patient in the ED is not able to access on-going inpatient care because of the lack of an inpatient bed and as a result these patients occupy and crowd the ED.

ED overcrowding: ED to be "overcrowded" when the number of patients waiting to be seen, the number of patients being treated and assessed and the number of patients waiting to leave the ED exceed either the bed capacity or staffing capacity of that ED

Ambulance bypass: When the ED is overcrowded, ambulances are instructed to divert to another facility because of a lack of capacity to safely attend to a newly-arrived patient at the original facility.

Domain experts: Clinicians.

Home ward: Home ward of a particular patient is defined as the ward where the multidisciplinary team responsible for their care is located.

Inliers: Patients admitted to their home ward.

Outliers: Patients admitted outside of their home ward.

Appendix D

Data Dictionary

Variable Name	Description	Data Type	Codes
journey_id	Unique journey identification for each admission	Long integer	
urn	Unique patient identification number	journey_id	
DateIn	Admission Date to a ward DDMMYYYY and HH:MM	Date/Time	
DateOut	Discharge Date from a ward DDMMYYYY and HH:MM	Date/Time	
ward	Name of ward	Text	
unit	Name of unit	Text	
division	Name of division	Text	
status	Inlier / outlier status	Text	
adm_cat	Admission category	Text	1 = Elective 2 = Emergency 3 = Elective Booking List 4 = Not applicable

Variable Name	Description	Data Type	Codes
nos	Nature of separation	Text	0 = Discharge on leave 1 = Home 2 = Other hospital - up transfer 3 = Nursing Home or hostel 4 = Other HC accom. 5 = Died - no autopsy 6 = Died - autopsy 7 = Other hospital - down transfer 8 = Self discharge 9 = Unknown A = Administrative separation E = End of Quater reporting
ageinyears	The age of the patient in years		
sex	Gender of the patient		1 = Male 2 = Female 3 = Indeterminant
Date Time seen	Date/time seen DDMMYYYY and HH:MM	Date / time	
Arrival date and time	Date/time of arrival DDMMYYYY and HH:MM	Date / time	
Triage date and time	Triage date/time DDMMYYYY and HH:MM	Date / time	

Variable Name	Description	Data Type	Codes
Triage category	Australian Triage Scale (ATS)		1 = Resuscitation 2 = Emergency 3 = Urgent 4 = Semi urgent 5 = Non urgent
Discharge summary sent within 2 days	Discharge summary was sent 2 days	Text	
Discharge summary sent within 7 days	Discharge summary was sent 7 days	Text	
Readmission within 7 days	Patient readmitted within seven days	Text	
Readmission within 28 days	Patient readmitted within seven days	Text	

Bibliography

Abraham, J. and Reddy, M. C. (2010). "Challenges to inter-departmental coordination of patient transfers: A workflow perspective." International Journal of Medical Informatics **79**(2): 112-122.

Ackroyd-Stolarz, S., Read Guernsey, J., Mackinnon, N. J. and Kovacs, G. (2011). "The association between a prolonged stay in the emergency department and adverse events in older patients admitted to hospital: a retrospective cohort study." BMJ Quality & Safety **20**(7): 564-569.

Aktas, E., Ülengin, F. and Önsel Sahin, S. (2007). "A decision support system to improve the efficiency of resource allocation in healthcare management." Socio-Economic Planning Sciences **41**(2): 130-146.

Alameda, C. and Suárez, C. (2009). "Clinical outcomes in medical outliers admitted to hospital with heart failure." European Journal of Internal Medicine **20**(8): 764-767.

Andersen, H., Rovik, K. A. and Ingebrigtsen, T. (2014). "Lean thinking in hospitals: is there a cure for the absence of evidence? A systematic review of reviews." British Medical Journal Open **4**(1): e003873.

Armstrong, J. J., Zhu, M., Hirdes, J. P. and Stolee, P. (2011). "K-Means Cluster Analysis of Rehabilitation Service Users in the Home Health Care System of Ontario: Examining the Heterogeneity of a Complex Geriatric Population." Archives of Physical Medicine and Rehabilitation(0).

Australasian College for Emergency Medicine. (2011). "Statement on Emergency Department Overcrowding." Retrieved 15 December, 2013, from http://www.acem.org.au/getattachment/0789ef2f-d814-4e86-af81-aad8b9e57c6d/Statement-on-Emergency-Department-Overcrowding.aspx.

Australian Bureau of Statistics. (2013). "Population Projections, Australia, 2006 to 2101 " Population Projections, Australia, 2006 to 2101 Retrieved 4 August, 2013, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3222.0Main+Features12006%20to%202101?OpenDocument.

Australian Government Department of Health and Ageing (2010). The State of Our Public Hospitals - June 2010 Report. Ageing, D. o. H. a., Commonwealth of Australia 2010.

Australian Institute of Health and Welfare (2011-2012). Australian Hospital Statistics 2011-12. Welfare, A. I. o. H. a., Australian Government: Australian Institute of Health and Welfare.

Banham, D., Woollacott, T. and Lynch, J. (2011). "Healthy life gains in South Australia 1999-2008: analysis of a local Burden of Disease series." Population Health Metrics **9**: 13.

Ben-Tovim, D. I., Bassham, J. E., Bennett, D. M., Dougherty, M. L., Martin, M. A., O'Neill, S. J., Sincock, J. L. and Szwarcbord, M. G. (2008a). "Redesigning care at the Flinders Medical Centre: clinical process redesign using "lean thinking"." Medical Journal of Australia **188**(6): S27-31.

Ben-Tovim, D. I., Bassham, J. E., Bolch, D., Martin, M., Dougherty, M. A. and Szwarcbord, M. (2007). "Lean thinking across a hospital: redesigning care at the: Flinders Medical Centre." Australian Health Review **31**(1): 10-15.

Ben-Tovim, D. I., Dougherty, M. L., O'Connell, T. J. and McGrath, K. M. (2008b). "Patient journeys: the process of clinical redesign." Medical Journal of Australia **188**(6): S14-17.

Ben Ayed, M., Ltifi, H., Kolski, C. and Alimi, A. M. (2010). "A user-centered approach for the design and implementation of KDD-based DSS: A case study in the healthcare domain." Decision Support Systems **50**(1): 64-78.

Bennett, C. C. (2010). "Taking the first step toward a healthier future." Medical Journal of Australia **192**(9): 510.

Braitberg, G. (2007). "Emergency department overcrowding: dying to get in?" Medical Journal of Australia **187**(11/12): 624-625.

Brand, C. A., Ibrahim, J. E., Cameron, P. A. and Scott, I. A. (2008). "Standards for health care: a necessary but unknown quantity." Medical Journal of Australia **189**(5): 257–260.

Brick, C., Lowes, J., Lovstrom, L., Kokotilo, A., Villa-Roel, C., Lee, P., Lang, E. and Rowe, B. H. (2013). "The impact of consultation on length of stay in tertiary care emergency departments." Emergency Medicine Journal.

Callen, J., Paoloni, R., Li, J., Stewart, M., Gibson, K., Georgiou, A., Braithwaite, J. and Westbrook, J. (2013). "Perceptions of the Effect of Information and Communication Technology on the Quality of Care Delivered in Emergency Departments: A Cross-Site Qualitative Study." Annals of Emergency Medicine **61**(2): 131-144.

Cameron, P., Scown, P. and Campbell, D. (2002). "Managing access block." Australian Health Review **25**(4): 59-68.

Cameron, P. A., Joseph, A. P. and McCarthy, S. M. (2009). "Access block can be managed." Medical Journal of Australia **190**(7): 364-368.

Caron, F., Vanthienen, J. and Baesens, B. (2013a). "Healthcare Analytics: Examining the Diagnosis–treatment Cycle." Procedia Technology **9**(0): 996-1004.

Caron, F., Vanthienen, J., Vanhaecht, K., Limbergen, E. V., De Weerdt, J. and Baesens, B. (2014). "Monitoring care processes in the gynecologic oncology department." Computers in Biology and Medicine **44**: 88-96.

Caron, F., Vanthienen, J., Vanhaecht, K., Van Limbergen, E., Deweerdt, J. and Baesens, B. (2013b). "A process mining-based investigation of adverse events in care processes." HIM J.

Carr, B. G., Hollander, J. E., Baxt, W. G., Datner, E. M. and Pines, J. M. (2010). "Trends in boarding of admitted patients in US Emergency Departments 2003-2005." Journal of Emergency Medicine **39**(4): 506-511.

Casalino, E., Choquet, C., Bernard, J., Debit, A., Doumenc, B., Berthoumieu, A. and Wargon, M. (2013). "Predictive variables of an emergency department quality and performance indicator: a 1-year prospective, observational, cohort study evaluating hospital and emergency census variables and emergency department time interval measurements." Emergency Medicine Journal **30**(8): 638-645.

Ceglowski, R., Churilov, L. and Wasserthiel., J. (2007). "Combining Data Mining and Discrete Event Simulation for a value-added view of a hospital emergency department." The Journal of the Operational Research Society **58**(2): 246-254.

Charlson, M. E., Pompei, P., Ales, K. L. and MacKenzie, C. R. (1987). "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation." Journal of Chronic Diseases **40**(5): 373-383.

Coiera, E. (2012). "The science of interruption." BMJ Quality & Safety **21**(5): 357-360.

Concha, O. P., Gallego, B., Hillman, K., Delaney, G. P. and Coiera, E. (2014). "Do variations in hospital mortality patterns after weekend admission reflect reduced quality of care or different patient cohorts? A population-based study." British Medical Journal - Quality & Safety **23**(3): 215-222.

Cooper, R. G. and Mohabeersingh, C. (2008). "Lean thinking for medical practices " Journal of Pre-Clinical and Clinical Research **Vol 2**(No. 1): 088-089.

Côté, M. J. (1999). "Patient flow and resource utilization in an outpatient clinic." Socio-Economic Planning Sciences **33**(3): 231-245.

DeCoster, J. and Claypool, H. (2004). "Data Analysis in SPSS." Retrieved 25 August, 2012, from http://www.stat-help.com/spss.pdf.

Delgado, M. K., Meng, L. J., Mercer, M. P., Pines, J. M., Owens, D. K. and Zaric, G. S. (2013). "Reducing Ambulance Diversion at Hospital and Regional Levels: Systemic Review of Insights from Simulation Models." Western Journal of Emergency Medicine **14**(5): 489-498.

Dellifraine, J., Langabeer, J. and King, B. (2010). "Quality improvement practices in academic emergency medicine: perspectives from the chairs." Western Journal of Emergency Medicine **11**(5): 479-485.

Derlet, R. W. (2002). "Overcrowding in emergency departments: Increased demand and decreased capacity." Annals Of Emergency Medicine **39**(4): 430-432.

Dumas, M., Aalst, W. M. v. d. and Hofstede, A. H. t. (2005). Process-aware information systems: bridging people and software through process technology, John Wiley Sons, Inc.

Dunn, R. (2003). "Reduced access block causes shorter emergency department waiting times: An historical control observational study." Emergency Medicine (Fremantle, W.A.) **15**(3): 232-238.

Fatovich, D. M. (2002). "Emergency medicine." British Medical Journal **324**: 958-962.

Fatovich, D. M., Nagree, Y. and Sprivulis, P. (2005). "Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia." Emergency Medicine Journal **22**(5): 351-354.

Fernandez-Llatas, C., Benedi, J. M., Garcia-Gomez, J. M. and Traver, V. (2013). "Process mining for individualized behavior modeling using wireless tracking in nursing homes." Sensors (Basel) **13**(11): 15434-15451.

Fernandez-Llatas, C., Meneu, T., Benedi, J. M. and Traver, V. (2010). "Activity-based process mining for clinical pathways computer aided design." Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society **2010**: 6178-6181.

Flinders Medical Centre. (2012). "Emergency & Perioperative Medicine." 2012, from http://www.flinders.sa.gov.au/emergency/pages/emerg/ABK_PCAz_z/.

Flinders Medical Centre. (2013, 13 March 2013). "Flinders Medical Centre - Profile." Retrieved 10 March, 2014, from http://www.flinders.sa.gov.au/aboutfmc/.

Fluxicon. (2012). "Disco." from http://fluxicon.com/disco/.

Forero, R., Hillman, K. M., McCarthy, S., Fatovich, D. M., Joseph, A. P. and Richardson, D. B. (2010). "Access block and ED overcrowding." Emergency Medicine Australasia **22**(2): 119-135.

Geelhoed, G. C. and de Klerk, N. H. (2012). "Emergency department overcrowding, mortality and the 4-hour rule in Western Australia." Medical Journal of Australia **196**: 122-126.

Government of South Australia. (2014, 12/02/2014). "Enterprise Patient Administration System (EPAS)." Retrieved 1 March, 2014, from http://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/health+reform/ehealth/enterprise+patient+administration+system+epas.

Gunal, M. M. (2012). "A guide for building hospital simulation models." Operational Research Society Limited **1**(1): 17-25.

Hammer, M. and Champy, J. (1993). "Reengineering the corporation." Small Business Reports **18**(11): 65.

Harders, M., Malangoni, M. A., Weight, S. and Sidhu, T. (2006). "Improving operating room efficiency through process redesign." International Journal of Surgery **140**(4).

Harris, A. and Sharma, A. (2010). "Access block and overcrowding in emergency departments: an empirical analysis." Emergency Medicine Journal **27**(7): 508-511.

Henneman, P. L., Nathanson, B. H., Li, H., Smithline, H. A., Blank, F. S., Santoro, J. P., Maynard, A. M., Provost, D. A. and Henneman, E. A. (2010). "Emergency department patients who stay more than 6 hours contribute to crowding." Journal of Emergency Medicine **39**(1): 105-112.

Higginson, I. (2012). "Emergency department crowding." Emergency Medicine Journal **29**(6): 437-443.

Hilberts, S. and Gray, K. (2014). "Education as ehealth infrastructure: considerations in advancing a national agenda for ehealth." Advances in Health Sciences Education: Theory and Practice **19**(1): 115-127.

Hoot, N. R., LeBlanc, L. J., Jones, I., Levin, S. R., Zhou, C., Gadd, C. S. and Aronsky, D. (2008). "Forecasting Emergency Department Crowding: A Discrete Event Simulation." Annals Of Emergency Medicine **52**(2): 116-125.

Howell, E., Bessman, E., Kravet, S., Kolodner, K., Marshall, R. and Wright, S. (2008). "Active bed management by hospitalists and emergency department throughput." Annals Of Internal Medicine **149**(11): 804-811.

Huang, Z., Lu, X. and Duan, H. (2012). "On mining clinical pathway patterns from medical behaviors." Artificial Intelligence in Medicine **56**(1): 35-50.

Huang, Z., Lu, X., Duan, H. and Fan, W. (2013). "Summarizing clinical pathways from event logs." Journal of Biomedical Informatics **46**(1): 111-127.

Hwang, U., McCarthy, M. L., Aronsky, D., Asplin, B., Crane, P. W., Craven, C. K., Epstein, S. K., Fee, C., Handel, D. A., Pines, J. M., Rathlev, N. K., Schafermeyer, R. W., Zwemer, F. L., Jr. and Bernstein, S. L. (2011). "Measures of crowding in the emergency department: a systematic review." Academic Emergency Medicine **18**(5): 527-538.

IBM. (2011). "Number of Clusters (auto-clustering) (TwoStep clustering algorithms)." Retrieved 27 October, 2012, from http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.cs%2Ftwostepcluster_table.htm.

IBM. (2012). "Goodness Measures (cluster evaluation algorithms)." Retrieved 20 August, 2012, from http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp.

Jensen, K., Kristensen, L. and Wells, L. (2007). "Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems." International Journal on Software Tools for Technology Transfer **9**(3-4): 213-254.

Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., Welch, S. J. and Snow, G. L. (2009). "A multivariate time series approach to modeling and forecasting demand in the emergency department." Journal of Biomedical Informatics **42**(1): 123-139.

Jun, J. B., Jacobson, S. H. and Swisher, J. R. (1999). "Application of discrete-event simulation in health care clinics: A survey." The Journal of the Operational Research Society **50**(2): 109-123.

Karnon, J., Mackay, M. and Mills, T. M. (2009). Mathematical modelling in health care. 18th World IMACS / MODSIM Congress, Cairns, Australia.

Kaufman, L. and Rousseeuw, P. J. (2005). Finding Groups in Data: Introduction to Cluster Analysis.

Kelleher, D. C., Jagadeesh Chandra Bose, R. P., Waterhouse, L. J., Carter, E. A. and Burd, R. S. (2014). "Effect of a Checklist on Advanced Trauma Life Support Workflow Deviations during Trauma Resuscitations without Pre-Arrival Notification." Journal of the American College of Surgeons **218**(3): 459-466.

Kim, E., Kim, S., Song, M., Kim, S., Yoo, D., Hwang, H. and Yoo, S. (2013). "Discovery of outpatient care process of a tertiary university hospital using process mining." Healthcare Informatics Research **19**(1): 42-49.

King, D. L., Ben-Tovim, D. I. and Bassham, J. (2006). "Redesigning emergency department patient flows: application of Lean Thinking to health care." Emergency Medicine Australasia **18**(4): 391-397.

Kolker, A. (2008). "Process modeling of emergency department patient flow: effect of patient length of stay on ED diversion." J Med Syst **32**(5): 389-401.

Kossovsky, M. P., Sarasin, F. P., Chopard, P., Louis-Simonet, M., Sigaud, P., Perneger, T. V. and Gaspoz, J.-M. (2002). "Relationship between hospital length of stay and quality of care in patients with congestive heart failure." Quality and Safety in Health Care **11**(3): 219-223.

Kusters, R. J. and Groot, P. M. A. (1996). "Modelling resource availability in general hospitals design and implementation of a decision support model." European Journal of Operational Research **88**(3): 428-445.

Lang, M., Burkle, T., Laumann, S. and Prokosch, H. U. (2008). "Process mining for clinical workflows: challenges and current limitations." Studies in Health Technology and Informatics **136**: 229-234.

Lee-Lewandrowski, E., Nichols, J., Van Cott, E., Grisson, R., Louissaint, A., Benzer, T. and Lewandrowski, K. (2009). "Implementation of a rapid whole blood D-dimer test in the emergency department of an urban academic medical center: impact on ED length of stay and ancillary test utilization." American Journal of Clinical Pathology **132**(3): 326-331.

Levin, S. R., Dittus, R., Aronsky, D., Weinger, M. B., Han, J., Boord, J. and France, D. (2008). "Optimizing cardiology capacity to reduce emergency department boarding: a systems engineering approach." American Heart Journal **156**(6): 1202-1209.

Li, J. Y. Z., Yong, T. Y., Hakendorf, P., Ben-Tovim, D. and Thompson, C. H. (2011). "Timeliness in discharge summary dissemination is associated with patients' clinical outcomes." Journal of Evaluation in Clinical Practice.

Li, S. Y., Magrabi, F. and Coiera, E. (2012). "A systematic review of the psychological literature on interruption and its patient safety implications." Journal of the American Medical Informatics Association **19**(1): 6-12.

Librero, J., Peiró, S. and Ordiñana, R. (1999). "Chronic Comorbidity and Outcomes of Hospital Care: Length of Stay, Mortality, and Readmission at 30 and 365 Days." Journal of Clinical Epidemiology **52**(3): 171-179.

Liew, D. and Kennedy, M. P. (2003). "Emergency department length of stay independently predicts excess inpatient length of stay." Medical Journal of Australia **179**(10): 524-526.

Lobach, M. D. P. M. S. D. F. and Hammond, P. W. E. (1997). "Computerized Decision Support Based on a Clinical Practice Guideline Improves Compliance with Care Standards." The American Journal of Medicine **102**(1): 89-98.

Losee, R. M. (2006). "Browsing mixed structured and unstructured data." Information Processing & Management **42**(2): 440-452.

Lucas, R., Farley, H., Twanmoh, J., Urumov, A., Olsen, N., Evans, B. and Kabiri, H. (2009). "Emergency department patient flow: the influence of hospital census variables on emergency department length of stay." Academic Emergency Medicine: Official Journal Of The Society For Academic Emergency Medicine **16**(7): 597-602.

Luke, D. A. (2005). "Getting the big picture in community science: methods that capture context." American Journal of Community Psychology **35**(3-4): 185-200.

MacLellan, D. G., Cregan, P. C., McCaughan, B. C., O'Connell, T. J. and McGrath, K. M. (2008). "Applying clinical process redesign methods to planned arrivals in New South Wales hospitals." Medical Journal of Australia **188**(6): S23–S26.

Mah, R. (2009). "Emergency department overcrowding as a threat to patient dignity." CJEM: Canadian Journal Of Emergency Medical Care **11**(4): 365-374.

Malhotra, S., Jordan, D., Shortliffe, E. and Patel, V. L. (2007). "Workflow modeling in critical care: Piecing together your own puzzle." Journal of Biomedical Informatics **40**(2): 81-92.

Mans, R., Aalst, W. P., Vanwersch, R. B. and Moleman, A. (2013). Process Mining in Healthcare: Data Challenges When Answering Frequently Posed Questions. Process Support and Knowledge Representation in Health Care. Lenz, R., Miksch, S., Peleg, M., Reichert, M., Riaño, D. and Teije, A., Springer Berlin Heidelberg. **7738:** 140-153.

Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S. and van der Aalst, W. (2008). "Process mining techniques: an application to stroke care." Studies in Health Technology and Informatics **136**: 573-578.

Mazzocato, P., Savage, C., Brommels, M., Aronsson, H. and Thor, J. (2010). "Lean thinking in healthcare: a realist review of the literature." Quality & Safety in Health Care **19**(5): 376-382.

McClelland, M. S., Jones, K., Siegel, B. and Pines, J. M. (2012). "A Field Test of Time-Based Emergency Department Quality Measures." Annals of Emergency Medicine **59**(1): 1-10.

McHaney, R. and Cronan, T. P. (2000). "Toward an empirical understanding of computer simulation implementation success." Information & Management **37**(3): 135-151.

Meneu, T., Traver, V., Guillen, S., Valdivieso, B., Benedi, J. and Fernandez-Llatas, C. (2013). "Heart Cycle: Facilitating the deployment of advanced care processes." Conference proceedings: Engineering in Medicine and Biology Society **2013**: 6996-6999.

Miller, A. R. and Tucker, C. (2014). "Health information exchange, system size and information silos." Journal of Health Economics **33**(0): 28-42.

Mitra, B., Cameron, P. A., Archer, P., Bailey, M., Pielage, P., Mele, G., Smit, D. V. and Newnham, H. (2012). "The association between time to disposition plan in the emergency department and in-hospital mortality of general medical patients." Internal Medicine Journal **42**(4): 444-450.

Mooi, E. and Sarstedt, M. (2011). Cluster Analysis. A Concise Guide to Market Research, Springer-Verlag**:** 237 - 284.

National Academy of Engineering. (2012). "NAE Grand Challenges for Engineering." Retrieved 01/03/2014, 2014, from http://www.engineeringchallenges.org/cms/8996/8938.aspx.

National Health Performance Authority (2012). Hospital Performance: Time patients spent in emergency departments in 2011−12. Australia, C. o. Sydney, Australia.

National Primary and Care Trust Development Programme Website. (2003). "Introduction to lean thinking " Retrieved 21 December, 2010, from http://www.natpact.info/demand_management/wizards/big_wizard/index.php?page=/demand_management/wizards/big_wizard/Step_3/Introduction_to_lean_thinking.php.

O'Brien, D., Williams, A., Blondell, K. and Jelinek, G. A. (2006). "Impact of streaming "fast track" emergency department patients." Australian Health Review **30**(4): 525-532.

O'Connell, T. J., Bassham, J. E., Bishop, R. O., Clarke, C. W., Hullick, C. J., King, D. L., Peek, C. L., Verma, R., Ben-Tovim, D. I. and McGrath, K. M. (2008). "Clinical process redesign for unplanned arrivals in hospitals." Medical Journal of Australia **188**(6): S18-22.

O'Neill, S., Jones, T., Bennett, D. and Lewis, M. (2011). "Nursing works: the application of lean thinking to nursing processes." Journal of Nursing Administration **41**(12): 546-552.

O’Connell, T. J., Ben-Tovim, D. I., McCaughan, B. C., Szwarcbord, M. G. and McGrath, K. M. (2008). "Health services under siege: the case for clinical process redesign." MJA **188**(6): S9–S13.

Office of the Auditor General. (2013, 27 June 2013). "Emergency Department Information System – Department of Health." Retrieved 16 February, 2014, from https://audit.wa.gov.au/reports-and-publications/reports/information-systems-application-controls-audits/emergency-department-information-system-department-of-health/.

Oxford University Dictionary (2014). Oxford Dictionaries. Oxford Dictionaries, Oxford University Press.

Paul, J. A. and Lin, L. (2012). "Models for improving patient throughput and waiting at hospital emergency departments." Journal of Emergency Medicine **43**(6): 1119-1126.

Perimal-Lewis, L., Ben-Tovim, D. I., Li, J. Y., Hakendorf, P. H. and Thompson, C. H. (2014a). "Relationship Emergency Department lengths of stay: characteristics favouring a delay to the admission decision as distinct from a delay while awaiting an inpatient bed." Internal Medicine Journal. (In Press).

Perimal-Lewis, L., Li, J. Y., Hakendorf, P. H., Ben-Tovim, D. I., Qin, S. and Thompson, C. H. (2012a). "The relationship between in-hospital location and outcomes of care in patients of a large general medical service." Internal Medicine Journal.

Perimal-Lewis, L., Qin, S., Thompson, C. and Hakendorf, P. (2012b). Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach. 5th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012).

Australasian Computer Science Week 2012 (ACSW2012). , RMIT, Melbourne, Australia, Australian Computer Society.

Perimal-Lewis, L., Thompson, C., Qin, S. and Hakendorf, P. (2013). Analysing homogenous patient journeys to assess quality of care for patients admitted outside of their "home-ward". 6th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2013), Adelaide, South Australia, Australian Computer Society Inc.

Perimal-Lewis, L., Vries, D. D. and Thompson, C. H. (2014b). Health intelligence: Discovering the process model using process mining by constructing Start-to-End patient journeys. 7th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand, ACS.

Process Mining Group. (2009). "Process Mining (ProM)." Retrieved 21 December, 2010, from http://prom.win.tue.nl/research/wiki/research/start.

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., Saunders, L. D., Beck, C. A., Feasby, T. E. and Ghali, W. A. (2005). "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data." Medical Care **43**(11): 1130-1139.

Rebuge, Á. and Ferreira, D. R. (2012). "Business process analysis in healthcare environments: A methodology based on process mining." Information Systems **37**(2): 99-116.

Richardson, D. B. (2006). "Increase in patient mortality at 10 days associated with emergency department overcrowding." Medical Journal of Australia **184**(5): 213-216.

Richardson, D. B. and Mountain, D. (2009). "Myths versus facts in emergency department overcrowding and hospital access block." Medical Journal of Australia **190**(7): 369–374.

Rozinat, A., Wynn, M. T., Aalst, W. M. P. v. d., Hofstede, A. H. M. t. and Fidge, C. J. (2008). Workflow Simulation for Operational Decision Support Using Design, Historic and State Information. Proceedings of the 6th International Conference on Business Process Management (BPM 2008), Milan, Italy, Springer Verlag.

Sackett, D. L., S. E. Straus, W. S. Richardson, W. Rosenberg and R. B. Haynes (2000). Evidence based medicine : how to practice and teach EBM. Edinburgh, Edinburgh : Churchill Livingstone.

Sahin, F. and Robinson, E. P. (2002). "Flow Coordination and Information Sharing in Supply Chains: Review, Implications, and Directions for Future Research." Decision Sciences **33**(4): 505-536.

Sally C. Brailsford (2007). Advances and challenges in healthcare simulation modeling: tutorial. Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come, Washington D.C., IEEE Press.

Santamaria, J. D., Tobin, A. E., Anstey, M. H., Smith, R. J. and Reid, D. A. (2014). "Do outlier inpatients experience more emergency calls in hospital? An observational cohort study." Medical Journal of Australia **200**(1): 45-48.

Schafermeyer, R. W. and Asplin, B. R. (2003). "Hospital and emergency department crowding in the United States." Emergency Medicine **15**(1): 22-27.

Shetty, A., Gunja, N., Byth, K. and Vukasovic, M. (2012). "Senior Streaming Assessment Further Evaluation after Triage zone: a novel model of care encompassing various emergency department throughput measures." Emergency Medicine Australasia **24**(4): 374-382.

Simon, T. G., Beland, M. D., Machan, J. T., DiPetrillo, T. and Dupuy, D. E. (2012). "Charlson Comorbidity Index predicts patient outcome, in cases of inoperable non-small cell lung cancer treated with radiofrequency ablation." European Journal of Radiology(0).

Simpao, A. F., Ahumada, L. M., Galvez, J. A. and Rehman, M. A. (2014). "A review of analytics and clinical informatics in health care." J Med Syst **38**(4): 45.

Singer, A. J., Thode, H. C., Jr., Viccellio, P. and Pines, J. M. (2011). "The association between length of emergency department boarding and mortality." Academic Emergency Medicine **18**(12): 1324-1329.

Solberg, L. I., Asplin, B. R., Weinick, R. M. and Magid, D. J. (2003). "Emergency department crowding: consensus development of potential measures." Annals of Emergency Medicine **42**(6): 824-834.

Song, M. and van der Aalst, W. M. P. (2008). "Towards comprehensive support for organizational mining." Decision Support Systems **46**(1): 300-317.

Sprivulis, P. C., Jacobs, I. G., Frazer, A. R. L. and Jelinek, G. A. (2006). "The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments." Medical Journal of Australia **184**(5): 208-212.

SPSS. (2001). "The SPSS TwoStep Cluster Component." SPSS - White Paper - Technical Report Retrieved 27 October, 2012, from http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf.

Tan, P.-N., Steinbach, M. and Kumar, V. (2005). Cluster Analysis: Basic Concepts and Algorithms. Introduction to Data Mining, Addison-Wesley**:** 487-568.

The Australian Council on Healthcare Standards (2013). Australasian Clinical Indicator Report: 2005–2012. Newcastle, H. S. R. G. U. o. Australia, The Australian Council on Healthcare Standards.

The Process Mining Group. (2010). "ProM." from http://www.processmining.org/.

Thomas, J. W., Guire, K. E. and Horvat, G. G. (1997). "Is patient length of stay related to quality of care?" Journal of Healthcare Management **42**(4): 489-507.

Trebble, T. M., Hansi, N., Hydes, T., Smith, M. A. and Baker, M. (2010). "Process mapping the patient journey through health care: an introduction." British Medical Journal **341**(c4078).

van der Aalst, W. M. P. (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer.

van der Aalst, W. M. P. and Günther, C. W. (2007). Finding Structure in Unstructured Processes: The Case for Process Mining. 7th International Conference on Application of Concurrency to System Design, 2007. ACSD 2007.

van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M. and Verbeek, H. M. W. (2007). "Business process mining: An industrial application." Information Systems **32**(5): 713-732.

van der Aalst, W. M. P., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G. and Weijters, A. J. M. M. (2003). "Workflow mining: A survey of issues and approaches." Data & Knowledge Engineering **47**(2): 237-267.

Van Walraven, C., Seth, R., Austin, P. C. and Laupacis, A. (2002). "Effect of Discharge Summary Availability During Post-discharge Visits on Hospital Readmission." Journal of General Internal Medicine **17**(3): 186-192.

Vankipuram, M., Kahol, K., Cohen, T. and Patel, V. L. (2010). "Toward automated workflow analysis and visualization in clinical environments." Journal of Biomedical Informatics **In Press, Corrected Proof**.

Viccellio, A., Santora, C., Singer, A. J., Thode, H. C., Jr. and Henry, M. C. (2009). "The association between transfer of emergency department boarders to inpatient hallways and mortality: a 4-year experience." Annals of Emergency Medicine **54**(4): 487-491.

Victor, C. R., Healy, J., Thomas, A. and Seargeant, J. (2000). "Older patients and delayed discharge from hospital." Health & Social Care in The Community **8**(6): 443-452.

Warne, S., Endacott, R., Ryan, H., Chamberlain, W., Hendry, J., Boulanger, C. and Donlin, N. (2010). "Non-therapeutic omission of medications in acutely ill patients." Nursing in Critical Care **15**(3): 112-117.

Weijters, A., van der Aalst, W. M. and De Medeiros, A. A. (2006). "Process mining with the heuristics miner-algorithm."

Whitney, K. M. and Daniels, C. B. (2013). "The Root Cause of Failure in Complex IT Projects: Complexity Itself." Procedia Computer Science **20**(0): 325-330.

Wynia, M. K. (1999). "Performance measures for ethics quality." Effective Clinical Practice **2**(6): 294-298.

Xu, G., Whitaker, E. and Hubbard, I. (2011). "Inpatient elderly care: Reducing re-admission rates." British Journal of Health Care Management **17**(12): 599-602.

Yeh, J.-Y. and Lin, W.-S. (2007). "Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department." Expert Systems with Applications **32**(4): 1073-1083.

Zubizaretta, G. (2013). "HOW TO SAFELY RACE to an OPTIMIZED FINANCIAL CLOSE." Financial Executive **29**(7): 40-43.