

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

LUA PERIMAL-LEWIS¹, DENISE DE VRIES¹, CAMPBELL H THOMPSON²

¹ School of Computer Science, Engineering and Mathematics
Flinders University of South Australia

² Medicine
The University of Adelaide, South Australia

lua.perimal-lewis@flinders.edu.au, denise.devries@flinders.edu.au, campbell.thompson@adelaide.edu.au

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.

Keywords: patient journey, process mining, simulation model, event logs, hospital key performance indicators, emergency department (ED), general medicine (GM), inliers, outliers.

1 Introduction

The demand on Australian Public Hospitals continues to intensify as the life expectancy of the overall Australian population increases especially the rise in residents who are 65 years old or older. The trend in South Australia 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 from 2.0 years among males and 1.5 years among females to 1.4 years among males and 1.2 years among females respectively. According to the Australian Bureau of Statistics (2013) there will be greater proportion of people aged 65 years or more and this population is projected to increase to between 23% and 25% in 2056. The Australian Institute of Health and Welfare (2011-2012) reported that there was an increase of 4.3% on average each year between 2007-08 and 2011-12 in Emergency Department (ED) presentations accounting to over 6.5 million presentations. The department also reported that the overall proportion of 'seen on time' were 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 the ED patients were admitted as hospital inpatients to continue care after the completion of ED treatment.

The data presented in the paragraph above shows the imminent increase in the demand for services at Australian Public Hospitals. Proportion of patients 'seen on time' is one of the ED Key Performance Indicators (KPIs) used Australia wide as a measurement of ED efficiency. The increased demand for services together with the pressure of progressing patients within a set timeframe in an attempt to satisfy certain KPI requirements often contributes to patients being streamed to wards that are not equipped with all the required facilities to treat the patient's conditions. Patients who stay outside of their homewards are referred as outliers.

EDs around Australia are often brought into the limelight by the media as being affected by access block. A patient is considered to have experienced access block if the patient has waited in the ED for 8 hours or more waiting for an inpatient bed and General Medicine (GM) patients are more likely to experience access block (Perimal-Lewis et al., 2013). O'Connell et al. (2008) assert that ED congestions are intensified by regular

Copyright © 2014, Australian Computer Society, Inc. This paper appeared at the Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2014), Auckland, New Zealand. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 153. J. Warren and K. Gray, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.

failure to manage processes involved in progressing patients through the hospital and highlight that the lack of shared understanding among staff, patients and carers of the probable patient path as one of the contributing factors amongst others. The authors call for better inpatient management for a better flow which will in turn ease issues faced within ED.

Hospitals are continually trying to improve their processes in order to cope with the increase in the demand for service and to increase the efficiency of delivery of care. Hospitals need to go beyond conventional aggregate information produced which is collected as part of performance reporting for the in-depth knowledge needed for process improvement activities. In previous work in this area Perimal-Lewis et al. (2012) stated that Clinical Process Re-engineering (CPR) is considered similar to Business Process Re-engineering (BPR) where both activities focus on continuous improvement to core business or clinical processes. The authors also regarded patient journeys as the core business process for hospitals and differentiated it as being patient-centred and service-oriented rather than business-oriented. Ben-Tovim et al. (2008) stated that clinical process redesign aims to harmonise the poorly coordinated patient journeys as patients move across multiple departments taking a holistic approach by looking at a wider area during the redesign process.

The performance of public hospitals is compared and judged publically by certain KPIs. Conformance to KPIs may soon become essential because of competitive government funding. KPI reporting is generally presented as an aggregate figure, such as *'20% of patients are seen on time'*. Often a poor KPI reported on core business areas would initiate a closer look at the underlying processes to investigate and redesign the processes in order to improve performance.

Processes in hospitals are complex. Process mining aims to gain insight into a process from carrying out detailed analysis using historical event data pertaining to that process. Generally, these event logs are easily sourced from Process Aware Information Systems (PAIS). 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) who is also the pioneer of this field. As such applying process mining in health will contribute to in depth analytical knowledge contributing to Health Intelligence (HI). Unlike many mainstream BI and data mining tools which are data-centric, process mining is process-centric aimed to gain insight into the processes the data refer to and the focus is not on fancy-looking dashboards rather a deeper analysis of the data (Van der Aalst, 2011).

This paper outlines how unstructured event data were processed to derive the event logs needed as an input for process mining in the absence of PAIS. Using the processed event data, process mining was then applied for an evidence-based process model discovery of patient journeys from start to end at Flinders Medical Centre (FMC). The discovered process model is the base for a simulation model. The discovered process models could also be used for performance analysis and verification.

The knowledge discovered using process mining techniques could be used as a basis for process optimisation.

The remaining section of the paper is structured as follows:

Section 2 gives a brief background to the conventional ways of modelling used in healthcare. Section 3 gives information on the study setting and the origin of data for the case study used in this paper. Section 4 describes the methodology used to derive the event log for process mining in the absence of PAIS. Section 5 describes only pertinent results related to the argument of this paper where the final output is the petri net model for simulation project. Section 6 is the discussion and finally the paper is concluded in Section 7 with the conclusion and future work.

2 Background

Traditionally health care modelling has been done using various mathematical modelling techniques focused on forecasting and predicting in order to improve health care performances (Perimal-Lewis et al., 2012). New approaches which use a combination of techniques to complement the strengths and weaknesses of the one technique are quickly emerging. One such research was carried out by Ceglowski et al. (2007) who proposed combining Data Mining techniques and discrete event simulation for identifying bottlenecks in the patient flow between ED and a hospital ward by providing insight into the complex relationship between patient urgency, treatment and disposal and the occurrence of queues for treatment.

Simulation is widely used in health care as a basis to understand processes and for prediction. There are many simulation projects which are discontinued after implementation as these models fail to improve the underlying processes. Process models are the core component of any simulation projects; 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. The conventional ways of creating process models by hand, ignoring event data are error prone which can lead to wrong conclusions. Using process mining solves this issue as models are extracted from events that have already taken place giving a close reflection of reality (Van der Aalst, 2011).

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

The proposition made in this paper is to use process mining techniques for the discovery of process model/s from historical event data which form the foundation for a simulation model. This model would give a closer reflection of reality. This paper introduces a novel way of

deriving event logs for process mining used to discover a process model that is validated by the domain experts as an authentic representation of the patient journey from start-to-end. The discovered process model could then be used as an input model for simulation projects.

3 Study Setting & Data

The analysis was undertaken on inpatient records for patients admitted and discharged by the General Medicine (GM) service at Flinders Medical Centre (FMC) and one specialty unit which is the Cardiology Unit. FMC is a public teaching hospital in South Australia and it attends to approximately 62,000 admitted inpatients per annum. The two largest medical inpatient specialties are GM and Cardiology. The typical patient admitted to each of these units is very different in terms of their age, complexities of disease and diagnosis. The GM service looks after a wide variety of diagnoses. The GM service controls about 100 inpatient beds out of about 500 beds in FMC as a whole. Cardiology is a specialty unit looking after a limited number of specific diagnoses that treats the highest number of patients compared to other specialty units. The analysis was carried out on inpatient records of the GM service; that is, on those patients whose inpatient care had been allocated to a GM team and on inpatient records of the Cardiology service; that is, on those patients whose inpatient care had been allocated to a Cardiology team. The wards that were 'home-wards' for this service were clearly defined. A 'home-ward' is a ward that is equipped with the appropriate medical team and specialised equipment to treat the patient's primary disease. Patients who were not allocated a 'home-ward' of the medical units responsible for their care were defined as being an outlier and stayed in an outlier ward. A similar concept is applied to any other specialty units.

The Patient Journey Database 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 with a unique journey number 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 occupied 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 of 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. The individual patients are not identifiable at any point.

Ethics approval for the use of data was granted by the Southern Adelaide Health Service / Flinders University Human Research Ethics Committee.

4 Methodology

Process mining activities can be categorised into three different perspectives, which are the process perspective, the organisational perspective and the case perspective. (Weijters et al., 2006). The process mining category presented in this paper 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 key to producing a good simulation model is to first understand the model with all possible variations which cater for all scenarios to produce a model that will be as close to reality as possible. A model accounting for all variations would help the domain experts perceive the entire process as it has taken place. Once the entire model is produced, it can then be deduced to represent the major behaviour of the system.

For a complex process such as the health care environment formulating a process model close to reality is far from trivial, therefore the use of historical data to derive a process model for simulation is advocated. This section describes the methodology used to develop a petri net based simulation model for the GM inpatients and for the Cardiology specialty unit's inpatients. The main software used for Process Mining is ProM (The Process Mining Group, 2010). ProM is open-source specialised process mining software. The constructed event log was preprocessed into the MXML format required as input to ProM. This conversion was done using the Disco software package (Fluxicon, 2012). Other ancillary software used was MS Excel and MS Access which was also necessary for the pre-processing of the event log. In section 4.1, the hospital admission process as explained by the domain experts is described to set the context of the area and process being investigated. The knowledge of the area and the underlying process is the foundation for constructing the event log which is described in section 4.2. Then in section 4.4 after the required event log is derived, the application of one of the process mining algorithms within ProM to derive the process model for simulation is described.

4.1 FMC's admission process

FMC offers both inpatient and outpatient services. Outpatients are seen during business hours at the outpatient clinics and sometimes these patients might be admitted as an inpatient. The inpatients could be categorised into two streams: those who enter the hospital as an emergency admission and those whose admissions are pre-planned for the 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 (FMC-WTS), Assessment time (FMC-RT) and Boarding (FMC-Boarding). The time patients spend at FMC-WTS and overall ED waiting time (= FMC WTS + FMC RT + FMC-Boarding) is measured and reported by the hospital as these times are part of the hospital's KPI. All the three phases of time take place within the ED.

The flow chart in Figure 4.1, as illustrated by the domain experts, 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.

An understanding of the processes and how patients flow through the ED and the KPIs surrounding these processes was a starting point for the process mining activities. This knowledge is essential in order to identify the data set and fields required to construct the inpatient journeys discussed in the next section. The flow-chart in Figure 4.1 as depicted by domain experts is used to verify the discovered process model/s and could also be used for conformance checking.

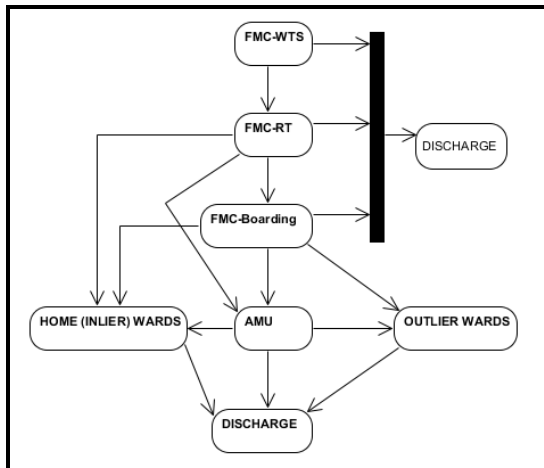


Figure 4.1: GM patient journey flowchart

4.2 Process Mining – Feature Extraction

In order to gain a holistic view of the patient journey process, it was necessary to derive an event log containing the required fields needed to discover the operational process model of patient journeys. The process of deriving the operational event log is far from a trivial exercise. The sensitive nature of health data and the ethics laws surrounding the use of health data meant that access to the information systems were not practical therefore the historical raw data supplied were sourced from various databases by the hospital. 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 patient journey event log used for process mining. Figure 4.2 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 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 1 | " " | 31 | 28 Jan 04 | 0 | 1439 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 1 | " " | 31 | 29 Jan 04 | 0 | 1439 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 5 | " " | 84 | 30 Jan 04 | 0 | 1439 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 5 | " " | 84 | 31 Jan 04 | 0 | 1439 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 5 | " " | 84 | 01 Feb 04 | 0 | 1439 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |
| 5 | " " | 84 | 02 Feb 04 | 0 | 825 | "60" | "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 | "60" | "CARD" | "Medicine, Cardiac and Critical care" |

Figure 4.2: Patient journey tab-separated flat file

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 4.3.

```

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,KL,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,VM,9/08/1993 16:02,9/08/1993 16:49,D/C,3657921,3657921,3657921,????
11/08/1993 8:50,9872357,4,VUL,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 4.3: 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 was established. Each field had to be converted to its appropriate data type. The major challenge for this process is the conversion of the date/time fields. In the patient journey dataset both ‘time1’ and ‘time2’ fields are 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. 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 Patient Unit Record Number ‘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_id’s. 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. The ED time is broken into three phases as described in Section 4.1: elapsed time between triage and when the patient is seen by a doctor (FMC-WTS), elapsed time from being seen by the doctor until a decision is made to admit the patient which is the overall assessment time (FMC-RT), and the time spent in the ED waiting for an inpatient bed after the decision to admit is made which is the boarding time (FMC-Boarding). All three components: FMC-WTS, FMC-RT and FMC-Boarding have been derived from the raw ED dataset. The snippet of the derived event log is shown in Figure 4.4. For each journey, the three components of ED time are derived from the ED dataset and the ward movement data are derived from the patient journey dataset. This extracted event log now contains information on the inpatient journey process from start to end.

| 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 |

Figure 4.4: The derived event log

Each tuple contains information about an event. The next section will discuss this derived event log and how it relates to the process mining methodology and concepts as discussed by Van der Aalst (2011).

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.

4.3 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 the 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. Within the ED there are three distinct events which relate to distinct ED processes as described in the previous section: FMC-WTS, FMC-RT, and FMC-Boarding. These ED processes also relate to the KPIs reported by the ED hence the availability of event data to produce the aggregate statistical information. Once the ED processes are completed the 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 in an outlier ward. Each event or activity (e.g. ward movement) is recorded with a timestamp (the ‘DateIn’ and ‘DateOut’) field. These fields are some of the many possible attributes that could be derived for an event. The timestamp of the event reflects the order of patient movement.

4.4 Process mining – Heuristics Miner - algorithm

It is apparent from the GM patient journey flowchart (Figure 4.1) that the patient journey flow from the ED to the various wards outside of the ED 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 as 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 5. The finer details of the workings on this algorithm are addressed in “*Process Mining with the Heuristics Miner-algorithm*” (Weijters et al., 2006).

The control flow process model is constructed by analysing for causal dependency. In this context the event log is analysed to see if 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 $A \Rightarrow_w B$). 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 $A \Rightarrow_w B = 5/6 = 0.833$. 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 $A \Rightarrow_w B = 50/51 = 0.980$. 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 $A \Rightarrow_w B = 49/52 = 0.94$ indicating that there is also strong dependent relationship.

5 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 helped the domain experts with identifying the relevant data set needed to be extracted from various systems used by the hospital. Using this data a specific event log was constructed as described in section 4.2.

Once the event log which is the most fundamental element in a process mining activity is available, this event log could be used for knowledge discovery applying various process mining algorithms available within ProM. The result presented here is specific and limited to the discovery of a simulation process model and corresponding analysis. The same event log could be used with various other process mining algorithms and analysis features within ProM which is not presented in this paper. Section 5.1 below presents the descriptive statistics in relation to the ED and the processes under investigation.

5.1 Descriptive Statistics

Descriptive statistical analysis of the overall FMC's patient data shows predictable patterns. Various statistical analyses are already being carried out to improve the efficiency and quality of patient care in general. One such previous work relating to 'Quality of Care' received by inlier and outlier patients has been addressed (Perimal-Lewis, 2013). Figure 5.1 and 5.2 show the trend in average waiting time and average patient count at the ED during triage.

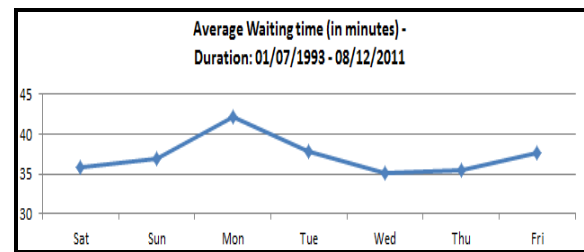


Figure 5.1: Trend in average waiting time (FMC-WTS)

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

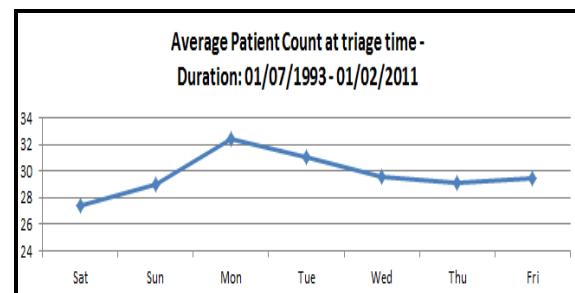


Figure 5.2: Average patient count at triage time

Based on the data captured for the various ED KPI reporting, 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 mature statistical analysis to understand this process that extends beyond the 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 specialty unit. Similar models could be built for other units as required.

5.2 Control flow perspective – heuristic models

As the models discovered are the base for simulation model/s, it was essential to choose a timeframe within the data set that reflected the processes within FMC where there was stability. Deriving this timeframe was done with close consultation with the domain experts. The timeframe used for these models was between 01/01/2007 and 31/12/2009.

Figure 5.3 shows the heuristic model for the cardiology specialty 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 patients that received 100% of their care from the Cardiology team. This means that these patients would have received their entire care from the same team of doctors. This is a rare but good occurrence as fewer unit (*team of doctors providing care*) changes are better for

the patients as the patients would receive uninterrupted care. The model is verified by the domain experts to be a correct reflection of patient journey for 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 section 4.4. As reflected in Figure 5.3 the dependency relationship is not too strong as firstly only a small subset of patients are modelled. Secondly, the possibility of such a transition where a patient's care from ED to discharge stays with the same team of doctors is often rare.

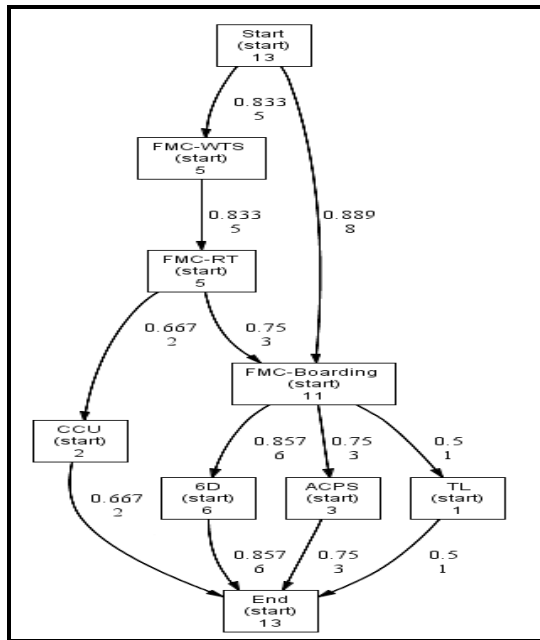


Figure 5.3: Cardiology patient journey

The next model shown in Figure 5.4 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 5.4 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 domain experts as the nature of 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 in 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 a better model. Further analysis of this model using '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 5.5 and for better readability a small section is of the model is shown in Figure 5.6

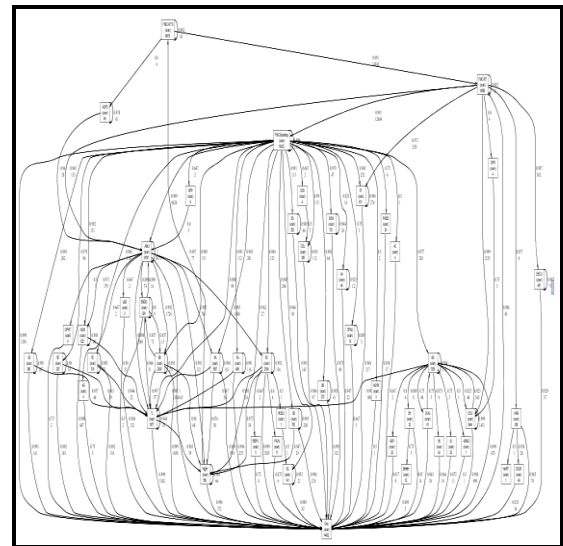


Figure 5.4: Complexity of first patient journey process model for GM patients

The model in Figure 5.5 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, journeys with a full 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 model accounted for 75% of the GM inpatients and was more interpretable. The number of patterns was now 113 patterns rather than 2000 patterns or more 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.

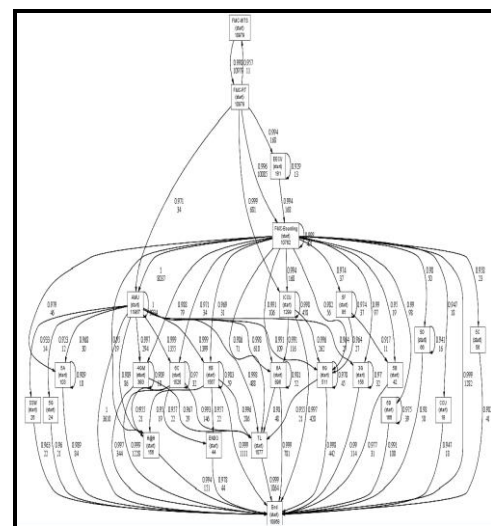


Figure 5.5: Complexity of the second patient journey process model for GM patients

The second heuristic model in Figure 5.5 and the snippet of the model in Figure 5.6 are also close representations in conformance with the GM patient journey flowchart shown in Figure 4.1. This knowledge helps enforce the validity of the discovered model. As well as validating the process model as depicted by the domain experts, the discovered model also revealed other ward movement patterns which were not accounted for by the domain experts. The model also helped identify potential deviations in the process which could also 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.

Furthermore 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, enforcing more confidence in a particular pathway as being a common feature of the GM patient journey. Other movements were verified correct by the domain experts to depict the 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 paper 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 which were verified to be a correct reflection of wards where the care of a patient might be transferred to another team because either the patients were being wrongly diagnosed and hence "sorted" into that unit or the unit offered a higher acuity of care and a significant deterioration in patients' condition often required a change in the team of doctors looking after those patients.

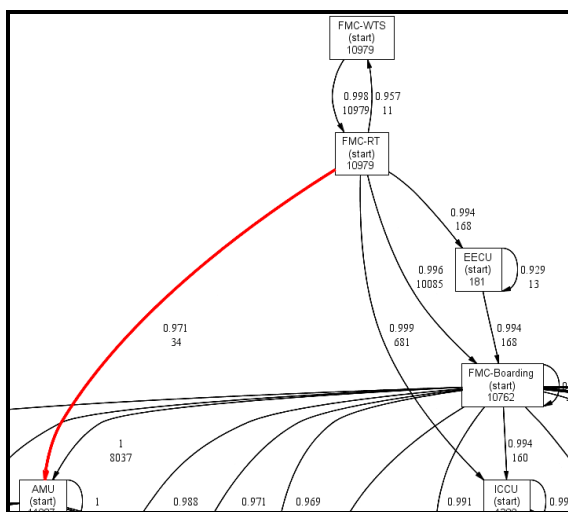


Figure 5.6: Snippet of the second patient journey process model for GM patients

Finally, once the verification of the models is to the satisfaction of all concerned, these models were converted into a Petri Net model. Figure 5.7 is a snippet of a Petri Net model which was derived from the second

patient journey model for GM patients shown in Figure 5.5. The Petri Net model can now be exported into a simulation tool such as the Coloured Petri Net (CPN) tool for simulation.

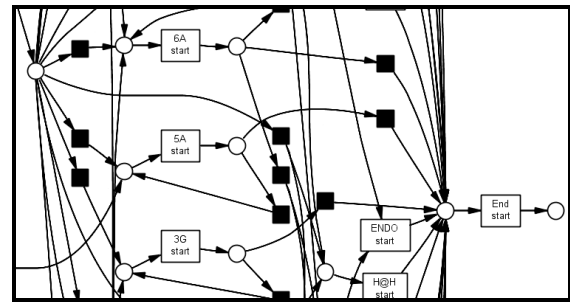


Figure 5.7: Snippet of Petri Net for GM patients

Coloured Petri Nets (CPNs) are a discrete-event modelling language for modelling systems where concurrency, communication and synchronisation plays a major role (Jensen et al., 2007). The CPN process model that will be used for simulation is a sound model derived from historical event logs and validated as a close reflection of reality by the domain experts. As a result, the patient journey process models discovered gives high confidence into the output of simulation exercise as these models are based of 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 when constructing these models by ‘hand’.

6 Discussion

It is important to be mindful of the scope and boundary for the data needed as otherwise the big data files 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 health care could lead to discontinuation of such projects. Therefore the collaboration with the domain experts should start at the very inception of the project and continue at every stage of the project.

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 wider spectrum of behaviour of the patient journey process. Insight such as this from the domain experts is significant for successful application of process mining in healthcare settings. Choosing a diverse patient group and then focussing on single specialised unit with less diverse patient separately is an important strategy for the big hospital data.

A distinction is 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 following a structured process. For example patients under the FMC-WTS phase are treated according to the Australian Triage Scale (ATS) which is a measure of severity and in the order of arrival.

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

7 Conclusion & Future Work

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 show 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 domain experts to identify key process improvement areas. For healthcare, these areas are normally areas where the performances are measured by KPIs. This will lead to the identification of appropriate data needed 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 is the most crucial and time consuming activity. 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 for process mining activities to be successful. This paper presented one such method for deriving event log in the absence of data warehouse or PAIS. In such instances, using the available data used for KPI reporting could be a starting point. Since similar KPIs are used in all public hospitals in Australia, this method is generalizable and similar approach could be used for other hospitals that would like to embrace process mining to dive deeper into their process as a starting point for 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 to investigate and reveal insights to access block and bottlenecks. Therefore constructing a process model for patient journeys from start-to-end as a base for simulation model derived from historical event log and validated by domain experts is advocated as a sound starting ground for future simulation projects.

The final start-to-end patient journey process models discovered as a base for simulation discussed in this paper could be further extended by including petri nets from other process mining perspectives such as the organisational perspective and the case perspective. The entire integrated model would give a holistic view of the process and therefore produce an all-encompassing input for simulation projects.

8 References

- Australian Bureau of Statistics (2013) *Population Projections, Australia, 2006 to 2101* <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3222.0Main+Features12006%20to%202101?OpenDocument>
- Australian Institute of Health and Welfare (2011-2012) (Ed, Welfare, A. I. o. H. a.).
- Banham, D., Woollacott, T. and Lynch, J. (2011) *Population Health Metrics*, **9**, 13.
- Ben-Tovim, D. I., Dougherty, M. L., O'Connell, T. J. and McGrath, K. M. (2008) *The Medical Journal of Australia*, **188**, S14-7.
- Ceglowski, R., Churilov, L. and Wasserthiel, J. (2007) *The Journal of the Operational Research Society*, **58**, 246-246-254.
- Fluxicon (2012) *Disco* <http://fluxicon.com/disco/>
- Jensen, K., Kristensen, L. and Wells, L. (2007) *International Journal on Software Tools for Technology Transfer*, **9**, 213-254. <http://dx.doi.org/10.1007/s10009-007-0038-x>
- Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S. and Van Der Aalst, W. (2008), Vol. 136, pp. 573-578.
- 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) *The Medical Journal of Australia*, **188**, S18-22.
- Perimal-Lewis, L., Li, J. Y., Hakendorf, P. H., Ben-Tovim, D. I., Qin, S. and Thompson, C. H. (2013) *Internal Medicine Journal*, **43**, 712-716. <http://dx.doi.org/10.1111/imj.12066>
- Perimal-Lewis, L., Qin, S., Thompson, C. and Hakendorf, P. (2012) *Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012) Melbourne, Australia*, **129**, 59-66
- Perimal-Lewis, L., Thompson, C., Qin, S. and Hakendorf, P. (2013) In *6th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2013)*(Ed, Koronios, K. G. a. A.) Australian Computer Society Inc., Adelaide, South Australia.
- Rebuge, Á. and Ferreira, D. R. (2011) *Information Systems*, **In Press**, **Corrected Proof**. <http://www.sciencedirect.com/science/article/pii/S0306437911000044>
- The Process Mining Group (2010) *ProM* <http://www.processmining.org/>
- Van der Aalst, W. M. P. (2011) *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer.
- Weijters, A., van der Aalst, W. M. and De Medeiros, A. A. (2006) Eindhoven University of Technology.