Business process analysis and process mining, particularly within the health care domain, remain under-utilised. Applied research that employs such techniques to routinely collected, health care data enables stakeholders to empirically investigate care as it is delivered by different health providers. However, cross-organisational mining and the comparative analysis of processes present a set of unique challenges in terms of ensuring population and activity comparability, visualising the mined models and interpreting the results. Without addressing these issues, health providers will find it difficult to use process mining insights, and the potential benefits of evidence-based process improvement within health will remain unrealised.

In this paper, we present a brief introduction on the nature of health care processes; a review of the process mining in health literature; and a case study conducted to explore and learn how health care data, and cross-organisational comparisons with process mining techniques may be approached. The case study applies process mining techniques to administrative and clinical data for patients who present with chest pain symptoms at one of four public hospitals in South Australia. We demonstrate an approach that provides detailed insights into clinical (quality of patient health) and fiscal (hospital budget) pressures in health care practice. We conclude by discussing the key lessons learned from our experience in conducting business process analysis and process mining based on the data from four different hospitals.

1. INTRODUCTION

Health systems both in Australia and abroad face significant challenges in how best to manage the effects of an ageing population with complex and costly health services needs [Commission 2005]. The pressure to contain costs and the expectations of continuous improvements in service quality have increased the need to understand how health care is provided and to determine whether cost-effective improvements to care practices can be made. Clinical guidelines and patient pathways are already developed with consideration of economic analysis that help ensure cost-effectiveness [Antioch et al. 2002; Eckard et al. 2011]. However, these guidelines and pathways are implemented within heterogeneous clinical contexts (e.g., at different hospitals), resulting in continued variations in cost, quality of care, and performance efficiencies [Runciman et al. 2012].

While some variations are necessary to account for different patient characteristics and preferences, unnecessary and unexplained over- or under-servicing of care contributes to inefficiencies through unnecessary spending or readmissions and mortality, respectively [Kennedy et al. 2010; Impellizzeri et al. 2009]. Through recent developments in the availability of clinical and administrative data, analyses have begun to look at the outcome of various patient treatments, with the aim of informing the economic impact of investment and practice alteration decisions; however, economic analyses of cost and health outcomes themselves are not enough to instigate change [Karnon et al. 2011]. Empirical insights into specific drivers of these outcomes and the modifiable elements within the process of care are also required. Findings from analyses into detailed patient pathways and clinical decision-making process would be useful to identify areas of practice that require redesign efforts and enable the monitoring of performance improvement actions.

Process mining [van der Aalst 2011], a research discipline that combines data mining and process analysis techniques, has the potential to use both supervised and unsupervised computer learning of big-data, held within hospital administrative and clinical data-warehouses, to derive descriptive models and statistics of health care processes and the pathways that patients travel through hospitals.

As a case study of current research and a necessary focus of future work, we propose the application of process mining as *an evidence-based business process analysis method* to investigate the variations in clinical practice and delivery of care across different hospital settings. Preliminary findings are presented from a trial conducted within the context of existing performance improvement initiatives in South Australia. Administrative and clinical data for over 13,000 patient journeys, extracted from two enterprise hospital information systems, were linked together and analysed to better understand the differences in clinical practices associated with chest pain management within comparable populations at four public hospitals.

The remainder of the paper is organised as follows. Section 2 presents the context and motivation behind the study and discusses related work. Section 3 describes the main research questions, our approach in addressing these questions using various process mining techniques, and a summary of preliminary findings from the case study. Section 4 discusses specific challenges and lessons learned from the case study. Section 5 concludes the paper.

2. BACKGROUND

Firstly, it is important to consider the background of health care processes, the concept of process mining with its various techniques and the current state of process analytics in health care.

2.1. Health care processes

A patient journey within a hospital setting consists of many different activities undertaken by different hospital staff (often in collaboration/consultation with one another) with the common goal of obtaining the best possible outcome for the patient in a timely manner. Some of these activities are *administrative* in nature, such as the registration of patient presentation, the admission and movement of patients to a ward, and the subsequent discharge; while others are *clinical* such as the triaging and risk stratification of patients, the ordering and delivery of tests and scans, disease diagnosis and therapy interventions [Lenz and Reichert 2007]. To help conceptualise these processes and how they interact as 'health care', the ability to view the different pathways taken by patients (with certain diagnoses and certain required treatments) through a hospital is very useful. Clinical and patient pathways are familiar to health services researchers and are typically used to communicate protocolised maps of how patients should be managed following evidence-based-medicine (EBM) guidelines and site-specific practice norms [De Bleser et al. 2006; Lenz and Reichert 2007].

However, it is not a simple task to capture these processes within 'as-is' descriptive models. Health care processes are recognised as "non-trivial" as the steps involved are often not linear and do not necessarily exist within a planned structure of sequences to the same extent as steps involved in other domains (e.g., manufacturing). Systems of clinical practice are not designed to be fully automated; instead, they rely on the professional expertise of medical specialists in the shaping of a care path. Thus, the occurrence of a task is not dependent merely on the completion of a previous task as it would be in e.g., the production line. Many other factors, such as a patient's overall health condition, his/her reaction to therapy, the dynamic professional environment with rapid changes to procedural options, the multi-disciplinary interaction of highly-specialised knowledge areas, and the human choice element are all inherent in decisions [Poulymenopoulou et al. 2003]. In addition, the majority of health care processes are *time-sensitive*, whereby timeliness of care affects patient health outcomes and the length of waiting times between activities can be a significant driver of cost [Scott 2003]. This is especially true for processes that provide acute care for patients.

2.2. Process Mining

Process mining is a research discipline which focuses on providing evidence-based process analysis techniques and tools for effective process management. Process mining techniques make use of the data in event logs to carry out detailed analysis on the behaviour of operational processes [van der Aalst 2011]. Process Mining studies have been carried out in over 100 organisations across a number of domains including banking and insurance, government agencies, education, transportation, and health care [van der Aalst 2011]. Many valuable insights have been gained regarding the importance of data quality, the stakeholder input and feedback as well as the relative importance of certain process mining perspectives or techniques over others, depending on the nature of the processes being analysed and the particular domain.

The three main categories of process mining techniques are process discovery, conformance, and enhancement [van der Aalst 2011]. Process discovery aims to adequately capture different behavioural aspects of non-trivial operational processes by taking an event log and producing process models without any additional information. Discovered process models can be used as a starting point for process improvement. Process conformance focuses on replaying the events recorded in a log on a process model to detect inconsistencies between the log and the model. The replay results can provide valuable insights for auditing and compliance purposes. Process enhancement focuses on extensions or improvement of existing process models using information contained in the log.

There are four different analysis perspectives through process mining techniques: the control-flow perspective, the organisational (resource) perspective, the case perspective, and the time perspective [van der Aalst 2011]. The control-flow perspective focuses on the ordering of activities. The organisational perspective is concerned with analysing resource information within an event log to better understand the roles that resources (both human and non-human) play in process enactment. The case perspective focuses on taking into account attributes related to a particular case for the classification of event logs and discovered process models. The time perspective focuses on the frequency and timing of events within an event log to derive useful insights, such as process bottlenecks. These four perspectives are orthogonal to the three categories of process mining techniques. In the next section, we specifically focus on existing work on the application of process mining in the area of health care.

2.3. Application of Process Mining Techniques in Health Care

There has been an increase in the application of process mining to the health care domain. This is not surprising given the unique ability of process mining to derive meaningful insights from the complex temporal relationships between activities and resources involved in processes. For example, Mans et al. [Mans et al. 2012] identified twelve studies related to the application of process mining in a variety of health care processes, such as the gynaecological oncology process in a Dutch hospital [Mans et al. 2008b], the emergency process in a public hospital in Portugal [Rebuge and Ferreira 2012], the process of an inpatient's journey from admission to discharge in an Australian public hospital [Perimal-Lewis et al. 2012], and the process of activities related to breast cancer treatment in a hospital in Belgium [Poelmans et al. 2010].

Nevertheless, the different foci and goals amongst these studies make it quite difficult to gauge the extent to which process mining has been applied in the health care domain, and more importantly, to identify potentially-interesting application areas that are yet to be explored. To these ends, a systematic literature review was conducted in late 2012. Using keyword-based literature search over three scholarly databases (Web of Science, Scopus, and Google Scholar) in addition to backward and forward search techniques [vom Brocke et al. 2009], 28 related papers (published as late as November 2012) were identified.

The extent to which process mining is applied in each of the identified papers was measured according to four dimensions: (1) *data preparation* - were there any explanations about data preparation activities in the papers? (2) *process mining techniques* - which types of analysis (discovery, conformance, and/or enhancement) were used in the studies? (3) *process mining perspectives* - which perspectives (control flow, organisational, time, and/or case) were being analysed in the studies? and (4) *comparative analysis* - did the studies focus on processes within a single hospital, or across multiple hospitals? Table I summarizes the evaluation of the 28 related papers.

Table I: Literature Review Evaluation Summary

Pre-processing	Mining Techniques			Perspectives			Comp. Anal.
	Disc.	Conf.	Enhc.	Control	Orgs.	Case	
15 (54%)	23 (82%)	6 (22%)	1 (3.5%)	25 (89%)	3 (11%)	7 (25%)	1 (3.5%)

Firstly, data pre-processing is an important step as health data often comes from heterogeneous sources and is thus fragmented. An explanation of how each study manipulated the data into a form that is suitable for process mining analysis is thus valuable knowledge. About half of the papers evaluated [Mans et al. 2008b; Bose and van der Aalst 2012; Staal 2010; Binder et al. 2012; Rebuge and Ferreira 2012; Perez-Castillo et al. 2011; Gupta 2007; Janssen 2011; Han et al. 2011; Manninen 2010; Elghazel et al. 2007; Mans et al. 2012; Poelmans et al. 2010; Ferreira and Alves 2011; Perimal-Lewis et al. 2012] focused on data pre-processing activities. A useful recent study is the one conducted by Mans et al. [Mans et al. 2012] whereby different types of data encountered in four Dutch hospitals' information systems were described, and options for using the data to address frequently posed questions by clinicians were explained. Nevertheless, given that over half of the studies reported on data preparation activities, the depth of study in this dimension could be improved.

With respect to process mining techniques, a clear majority of papers (82%) covered process discovery techniques [Mans et al. 2008b; Mans et al. 2008a; Quaglini 2010; Lang et al. 2008; Bose and van der Aalst 2012; Staal 2010; Binder et al. 2012; Poelmans et al. 2010; Gunther and Van der Aalst 2007; Perez-Castillo et al. 2011; Rebuge and Ferreira 2012; Song et al. 2009; Ferreira and Alves 2011; Gupta 2007; Janssen 2011; Fernandez-Llatas et al. 2010; Han et al. 2011; Manninen 2010; Huang et al. 2012; McGregor et al. 2011; Mans et al. 2012; Perimal-Lewis et al. 2012; Blum et al. 2008]. There were only six studies [Mans et al. 2008b; Dunkl et al. 2011; Binder et al. 2012; Quaglini 2010; Peleg et al. 2007; Kuo and Chen 2012] that reported on the use of conformance analysis, and only one paper [Mans et al. 2008b] that reported on process enhancement. Therefore, the use of ‘conformance’ and ’enhancement’ process mining techniques seems to be currently under-utilised in the health care field.

In terms of the process mining perspectives, most studies (about 89%) focused on control-flow analysis [Mans et al. 2008b; Mans et al. 2008a; Quaglini 2010; Lang et al. 2008; Dunkl et al. 2011; Bose and van der Aalst 2012; Staal 2010; Binder et al. 2012; Peleg et al. 2007; Poelmans et al. 2010; Gunther and Van der Aalst 2007; Perez-Castillo et al. 2011; Rebuge and Ferreira 2012; Song et al. 2009; Gupta 2007; Janssen 2011; Fernandez-Llatas et al. 2010; Han et al. 2011; Huang et al. 2012; Kuo and Chen 2012; Manninen 2010; McGregor et al. 2011; Mans et al. 2012; Perimal-Lewis et al. 2012; Blum et al. 2008]. Only seven studies (25%) looked into the time perspective [Mans et al. 2008b; Mans et al. 2008a; Staal 2010; Quaglini 2010; Peleg et al. 2007; Rebuge and Ferreira 2012; Perimal-Lewis et al. 2012], and only three studies (11%) [Mans et al. 2008b; Rebuge and Ferreira 2012; Ferreira and Alves 2011] reported on the organizational perspective. We can thus see that the mining of the organisational, time, and case perspectives seems to be over-looked in the health care setting.

Finally, since the focus of this paper is to conduct comparative analysis for identifying the (dis)similarity of practices across different hospitals, each paper was also evaluated to see if comparative analyses were attempted. Amongst all 28 papers identified, there was only *one* paper [Mans et al. 2008a] that attempted to conduct such analysis. In [Mans et al. 2008a], data from 368 patients diagnosed with ‘first-ever ischemic stroke’ from four Italian hospitals were analyzed in order to discover the procedures involved in the treatment of the patients. With only *one* comparative analysis study identified, the use of cross-organisational process mining for comparative analysis purposes is clearly under-exploited.

3. COMPARATIVE ANALYSIS CASE STUDY

Through reporting on the results of a *comparative analysis* case study across four hospitals in South Australia, and by focusing on the *control-flow and the time* perspectives, we attempt to address some of the research gaps identified in Section 2. Furthermore, by providing a detailed report on data *pre-processing* activities, we also seek to add knowledge to the way in which data from Australian public hospitals can be utilised for process mining purposes. With the aim of demonstrating comparative health care analyses through process mining techniques, it was envisaged that this would be easiest in a service area that contains a high volume of activity and one which also demands significant financial resources within a hospital. With this in mind, we focused on a population of patients with symptoms suggestive of acute coronary syndrome (ACS), who presented at the Emergency Department (ED) of one of four South Australian hospitals.

3.1. Approach

The study was exploratory in nature, with the goal to identify differences in practice between hospitals. The approach, as illustrated in Figure 1, was conducted in stages that aligned with those of the *L* life-cycle model* for process mining [van der Aalst et al. 2012]. First, activity data was extracted and reformatted according to the project and process mining method requirements (Stage 1). Second, the preprocessing and initial exploration and checks of the event log data occurred (Stage 2). A number of process mining techniques were then used to discover the control-flow and the performance of the processes at each hospital (Stage 3) for comparative analysis. The results were then iteratively interpreted and enhanced through clinical stakeholder engagement. Note that Stage 4 of the *L* life-cycle model* (explicit operational support) was beyond the scope of this particular study.

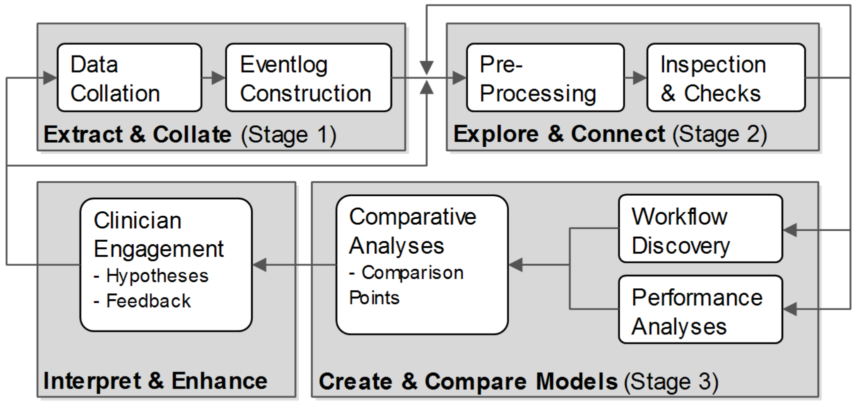

Fig. 1: Process mining approach used during the case study

Initial methodological questions were explored to determine whether we could collate and link the right event data, whether the data could be reformatted appropriately, and whether the data accurately portrayed clinical pathways. As we set out to investigate the elements of the process known to impact service performance and efficiency, four *Comparison Points* were identified based on known drivers of costs and/or patient health outcomes. These points were used to direct the focus of the comparative analyses, around which the context of preceding and subsequent events could then be explored.

Comparison Point 1. (CP-1)
The proportion of patients admitted to an inpatient care setting
— of those admitted, to which clinical unit(s) were they admitted to?

Comparison Point 2. (CP-2)
The throughput timing between ED presentation and movement to an inpatient setting (Admission)
— are there associated differences in initial risk (triage) categorisation?
— does throughput differ depending on the clinical unit to which patients are admitted to?

Comparison Point 3. (CP-3)
The frequency of procedures (diagnostic/treatment) provided
— does procedure use differ depending on the clinical unit to which patients are admitted?

Comparison Point 4. (CP-4)
The total length of stay for patients
— does the length of stay differ depending on the clinical unit to which patients are admitted to?

3.2. Data Collation

Variables of interest to the analysis were identified through a review of the Australian clinical guidelines for the management of ACS [Group 2006], and based on existing literature regarding the cardiac care process and measures of quality and performance [Scott 2003; Scott et al. 2004]. The resulting data framework was finalised following subsequent discussions with a consultant cardiologist. While each hospital site is responsible for the collection and input of their patient data, centrally collated ED and inpatient data repositories exist within the health department, from which the initial collection of data was extracted and linked. Both ED and inpatient activities were captured at a patient-level of granularity, across which standardised nomenclature, clinical ontologies and collection practice exist across the hospitals. Annonymisation of patient records was applied at the extraction level in order to preserve privacy.

3.3. Eventlog Construction

After extraction, the anonymised data was reformatted from a *case-log* data format, into transactional *event logs* (see Figure 2). Certain administrative and clinical attributes, such as *hospital ID* and *triage category*, were incorporated into the name of activities and used to characterise a specific event. This was important for enhancing the process models and the visualisation of patient trajectories through the hospital. In some instances, proxy timestamps were needed for data elements for which there were no recoded timestamps. Specifically, such data elements included the mode of transport to the ED (i.e., ambulance or other), the issuance of a working diagnosis (i.e. Chest Pain), and for the implementation of some therapeutic procedures. As a result, some assumptions were made regarding the temporal order of these events, but they occurred only when there is already an implied and clear temporal order (e.g., transportation to ED must necessarily happen before the ED presentation). Of course, such timestamps were not used for performance analysis purposes. This approach enables us to visually represent real steps in the process, on which we had activity data, but for which timestamps are not available.

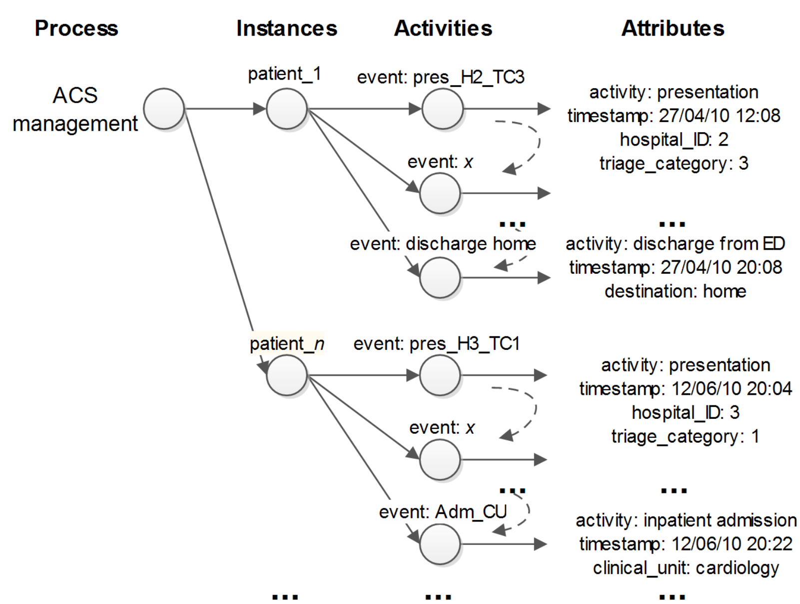

Fig. 2: Structure of data within event logs, based on [van der Aalst 2011, pp. 100]

3.4. Preprocessing and Log Inspection

Event logs were filtered using the software, Nitro [Laboratories 2012] to minimise the presence of incomplete instances and/or variables inconsistently collected across the sites, which may have complicated the models and affected the accuracy of the mapping. Three types of filters were used: attribute, start, and end point. The event logs were also split based on presentation and diagnosis attributes, for development of attribute specific (diagnosis, presenting hospital) models.

The initial dataset contained 27,773 instances of patients who had engaged the services at one of the four hospitals with either a suspected cardiac presentation, or cardiac related discharge between 1 July 2009 and 30 June 2010. To focus analyses on a specific, comparable population, only those instances with an ED preliminary diagnoses of Chest Pain (ICD10: R07) were included. This resulted in 9,713 instances of patient presentations: 3,434; 2,072; 1,606; and 2,601 at Hospitals 1 through 4, respectively.

3.5. Hospital Specific Workflow Discovery and Performance Analysis

Initially, hospital specific analyses were conducted in order to describe and understand local activities and facilitate site-by-site feedback. Using the heuristic mining plug-in [Weijters et al. 2006] within the Process Mining tool, ProM, causative work-flow nets were derived for each hospital. Figure 3 presents an illustrative example, describing the flow of patients attending one of the four hospitals.

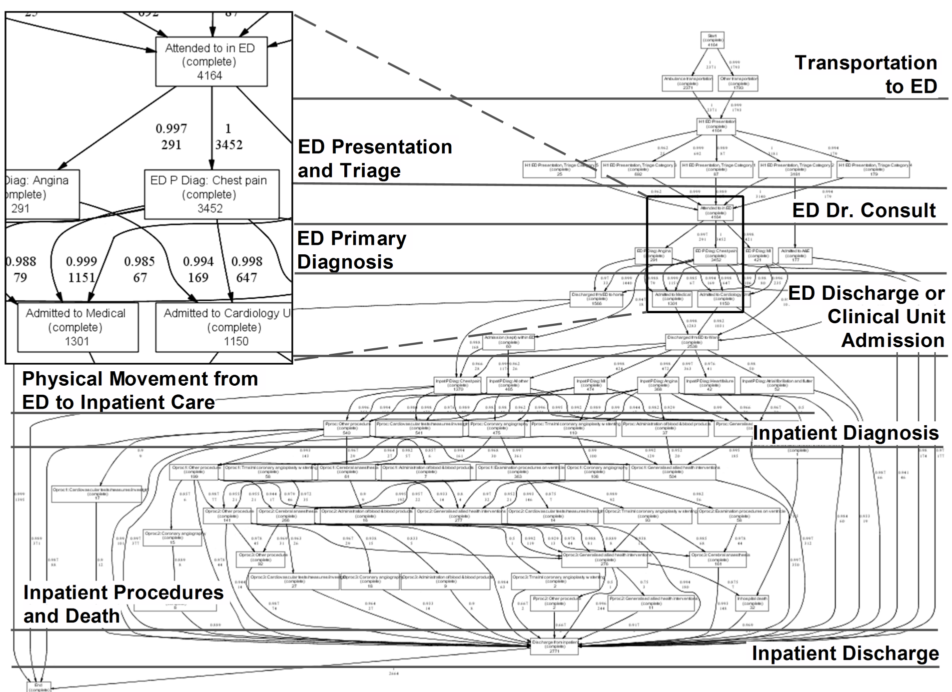

Fig. 3: Illustrative workflow net with nine process layers

Although most hospital processes can be quite complex and discovered process models can be unreadable and *spaghetti-like* [van der Aalst 2011], our experience with clinical practice suggests that an overall acute patient flow, on a high aggregate level of abstraction, is rather structured and *lasagna-like* [van der Aalst 2011]. In fact, as shown in Figure 3, the flow can be mainly organised into a number of stages: entry, assessment, stratification, action and exit. As such, nine potential stages or steps within the clinical process are represented and described within Figure 3, as indicated by the horizontal, descending lines overlaid on the model. Each rectangular box in the model represents an event in the care process. An overview of the model in Figure 3 shows that there exist many alternate pathways that a patient may follow.

The performance analysis using the performance analysis with the Petri-net plug-in within ProM [Hornix 2007] was also applied by replaying the event log data through the events and transitions obtained from the heuristically mined workflow models. Mean, standard deviation and mean of the interquartile range were captured to compute timing metrics of interest to the study, such as *waiting times*, *throughput* and *Length of Stay (LoS)*. Events with a proxy timestamp, were *excluded* from the performance analyses and had no bearing on the timing metrics of interest to the study. Because of the structured *lasagna-like* process, the heuristic workflow models translated into Petri nets and used in performance analyses had continuous semantic fitness scores of >0.95 on a 0-1 scale, meaning that the behaviours captured in the models were representative of the activities recorded in the event log.

3.6. Comparative Analysis across Hospitals

Firstly, a 'standard' set of pathways reflective of common activities across the hospitals was abstracted from the mined, workflow models and re-expressed into Business Process Modelling Notation (BPMN) models. These BPMN models were then populated with the quantitative results from the hospital specific workflow and performance analyses, thereby enabling the direct consideration of both workflow and timing elements within the same visual model.

Figure 4 and Figure 5 depict the number of patients who receive a similar (chest pain) diagnosis at one of four hospitals (n=x). The figures also illustrate the different pathways taken by these patients and the timing [mean HH:mm (standard deviation HH:mm)] associated with throughput and LoS. The BPMN models are also annotated with information to address the questions raised earlier with the four comparison points (i.e., CP1-4).

As another approach to visualise and analyse the differences in the patient pathways between hospitals, we created one common process model that captures the patient pathways within all four hospitals by applying the Fuzzy Miner Plug-in [Günther 2009] on one large combined log from all hospitals. The log from each hospital was then replayed separately on the common model using the Fuzzy Animation capability. During the animation, as paths connecting any two activities were traversed, the line connecting the activities became thicker. As a result, well-traversed (or dominant) pathways became visibly thicker than infrequently-traversed paths. Thus, we were able to obtain comparable 'maps' of patient pathways (as shown in Figure 6) from which we could identify and communicate the differences between hospitals.

3.7. Comparative Findings

From both the abstraction of workflow and performance analyses and comparative Fuzzy Model, a number of significant differences in practice were observed. A synopsis of the main comparative findings, aligned with our four *Comparison Points*, are presented below.

We were able to see significant differences in *CP-1*, whereby the proportion of patients admitted and transferred through to a ward ranged from 23-65%, and with these very similar patients (all given preliminary chest pain diagnoses) being admitted to various clinical units. As highlighted as 'Observation A' in Figure 6, Hospital 1 admitted a higher proportion of patients to the Medical Unit in comparison to other

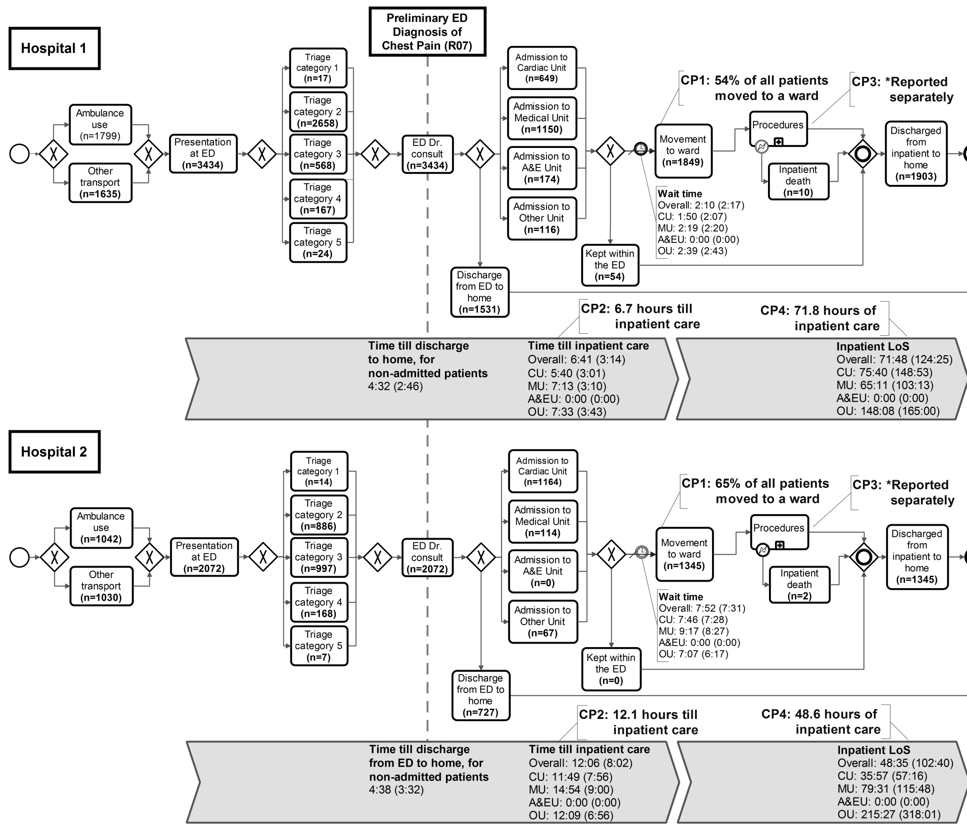

Fig. 4: BPMN models illustrating the Flow & Timing for Hospitals 1 & 2

hospitals. Unexpectedly, 'Observation C' in Figure 6 highlights that Hospitals 3 and 4 made use of an inpatient Accident and Emergency (A&E) Unit, to which they admitted patients, and 'kept them within ED' rather than moving them to a ward. Such practice is not readily observed at Hospitals 1 and 2, despite these two hospitals also possessing similar inpatient A&E facilities.

A second point of difference at *CP-2*, reported in Figures 4 and 5, was that patients admitted to Hospital 1 enjoyed a much faster throughput time between presentation and admission. At 6.7 hours, Hospital 1 was able to move these patients to a ward for inpatient care, 3.5 hours faster than the next fastest Hospital 4 and up to 8 hours faster than the third fastest hospital.

When we look at the associated pathways, we see that this may be caused, in part, by variances in the urgency categorisation (triage) at different hospital sites. For example, as shown in Figure 6 as 'Observation B', the proportion of patients being categorised with Triage Category 3 is *lower* in Hospital 1 than in three other hospitals where patients are more commonly processed through Triage Category 2. This may have been influenced by the clinical unit location to where patients were being admitted and the capacity of these different units. So while the throughput speed for inpatient care at Hospital 1 seemed to be ideal, this hospital was seen to discharge a majority (65%) of patients to non-cardiac (i.e. Medical) clinical units. One possible interpretation of this is that Hospital 1 was directing patients to any non-cardiac (i.e. medical) unit, simply

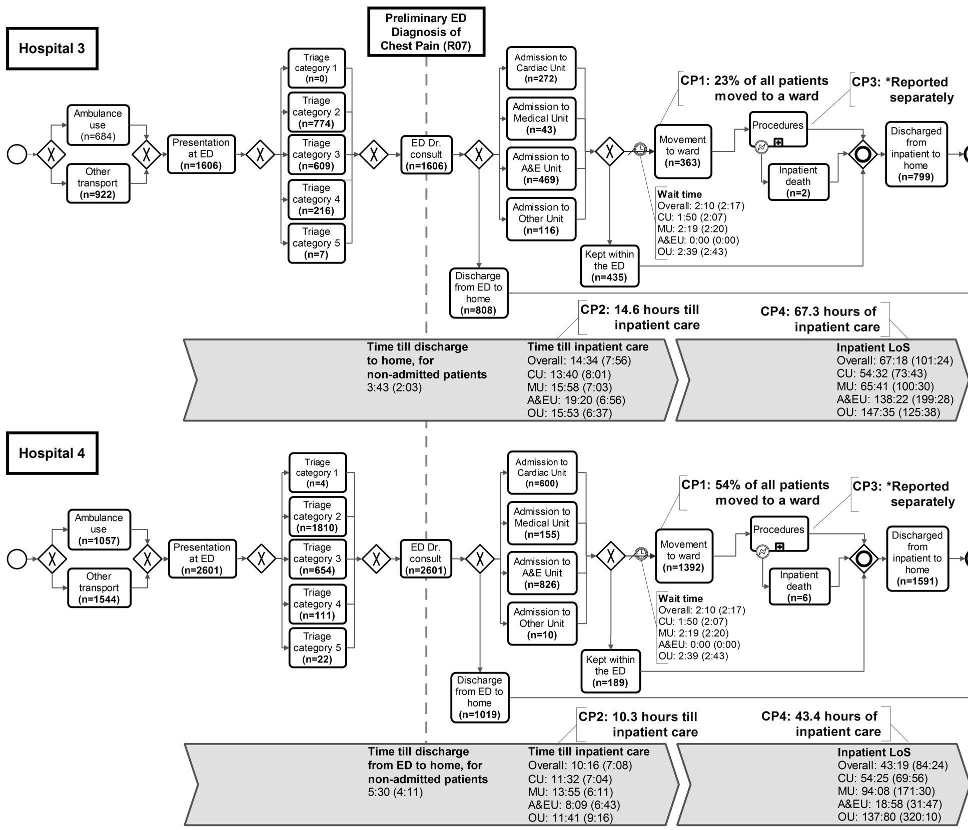

Fig. 5: BPMN models illustrating the Flow & Timing for Hospitals 3 & 4

when the first bed became available. This, however, is hard to corroborate within the data, particularly as there was little difference in the throughput and waiting times between the different clinical units at Hospital 1.

When looking at the use of various procedures within these chest pain patients (not shown in the figures), the two largest hospitals, Hospitals 1 and 2, were found to provide an almost identical volume of angiography. Interestingly however, the two smaller hospitals were found to make the least use of angiography. In looking deeper into *CP-3*, the rate of angiography use for patients admitted within Cardiac Units at Hospitals 2, 3 and 4 ranged from 12-18%, while it was 33% at Hospital 1. Given that Hospital 1 is seemingly selective in admitting cases to its Cardiac Unit, perhaps those patients have a greater need for such a procedure and the data can be interpreted as such.

Finally, Hospitals 1 and 3 had the longest inpatient LoS (mean of 70 hours), whilst length of stay (LoS) at Hospitals 2 and 4 was 20 to 25 hours shorter. LoS at H4 was driven down by a large proportion of patients admitted to an A&E clinical unit, whose mean LoS was a mere 19 hours. At Hospitals 1 and 3, LoS of patients admitted to either the Cardiac and Medical Units was very similar, whilst patients admitted to a Medical Unit at Hospitals 2 and 4 had much longer LoS than patients admitted to a Cardiac Unit at these hospitals.

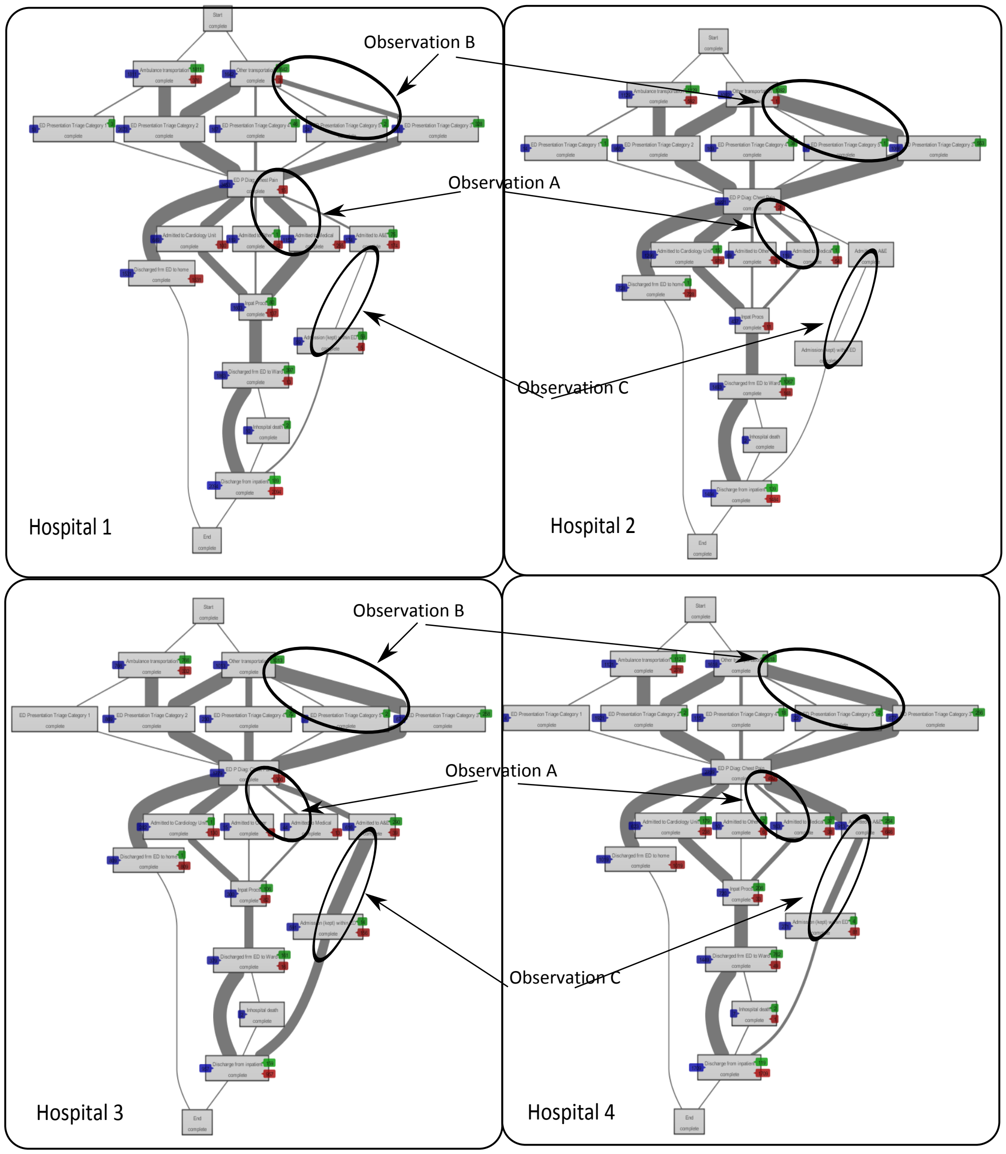

Fig. 6: Fuzzy Model Animations for All Four Hospitals

4. DISCUSSION

While it may seem easy to judge observed differences from the analysis alone, it is important to recognise that such results provide only a *“2D slice of a 3D reality”* [van der Aalst 2011, pp. 123]. As such, the results from the case study were communicated back to the stakeholders, as part of an iterative engagement to ask questions, such as: *‘is the data actually telling us what we think it is telling us?*

In this section, we draw on the case study experience, previous knowledge of health care, and comments from a stakeholder engagement to discuss the successes of cross-organisational process analysis in using the data to provide empirical insights together with the challenges faced. In linking the discussion of the case study to existing challenges highlighted in the Process Mining Manifesto [van der Aalst et al. 2012], we make some recommendations for others seeking to utilise similar process analyses in health care. Further, we illuminate the next steps that are currently underway here in Australia, and highlight future work that is needed to progress process-oriented analysis in health care.

4.1. Challenges, Successes and Lessons Learned

Piecing together the "jigsaw puzzle" of data. The problem of piecing data from multiple sources is recognised within the manifesto as one of the key challenges of cross-organisational analyses [van der Aalst et al. 2012, pp. 12]. Within this case study, such linkage of data pieces across legacy information systems was essential to enabling a 'whole of process' analysis and the 'end-to-end' visualisation across both the ED and inpatient settings.

As captured in the quote below, the clinical staff usually lament the fact that they only receive a limited view of patient pathways, and as a result it is hard to gain an accurate representation of lost opportunity to provide better care:

> *"[we / clinicians] generally only see a selected sample of those who come across their desk.. [we] only see a selected sample ... [we] don't see the denominator, only the numerator" – Clinical Stakeholder Group*

In this case, the successful representation of pathways for a population defined by their initial categorisation (i.e., ED diagnosis), allowed us to gain insights into the entire 'at risk' population (the denominator) and thus, we could identify any missed opportunities to provide better care to those discharged home from ED or admitted to non-cardiac units in a hospital.

In general though, piecing health care data together can be fairly difficult. In many instances, relational databases do not exist across, or even within, hospitals. Even if they do exist, numerous unique identification numbers are often used to represent patients and are inconsistently maintained. Therefore, it is not straight forward to merge data via unique identifiers. Instead, the correlation of cases must use, as what we have done in this case study, deterministic and probabilistic methods that consider loosely identifying variables (i.e., date of birth, postcode, gender, etcetera) and calculate the likelihood that a patient case in one system relates to the same patient case across other databases [Roos et al. 1986; Roos and Wajda 1991]. Our experience with the application of these methods prove to be quite positive.

Ensuring Comparable Populations. The facilitation of organisational learning though comparison, is also highlighted in the manifesto as a challenge for cross-organisational mining [van der Aalst et al. 2012, pp. 12]. This challenge is particularly relevant within the health care domain, where despite the existence of standardised ontologies built into the information systems (i.e., SNOMED-CT, UMLS, ICD10), some doubts still exist regarding consistency in the use of the ontologies. When the case study results were presented to the stakeholder group, there emerged some doubt as to whether there was consistency in the practice or ICD10 coding of a diagnosis. While it was generally agreed among the stakeholders that there was likely to be little difference, it did bring into question the true comparability of ACS patient populations across hospitals.

For future analysis, we plan to incorporate the results of diagnostic tests as a discretised categorical attributes for each patient wherever possible. Through the consensus among clinicians on the sensitivity and specificity of these tests, such medical attributes could be used to characterise and define clinically comparable patient populations in a more precise manner.

Ensuring Comparable Activity. The greatest success we had was in comparing the activities across the four hospitals. For the purpose of this case study, we defined relevant and comparable behaviour within the four *Comparison Points*. We found that it was important to provide some focus as to which areas that we should analyse in order to avoid being overwhelmed by too many 'potentially-relevant' comparison points. As mentioned in Section 3, the *Comparison Points* were elicited from discussions with the stakeholders as to which activities were known to add to, or detract from, the value (health outcomes and costs) of health care services. As a result, we were able to elicit the following insights:

> *"Fascinating... we've known [informally] that you've got to get past the cardiology registrars [at Hospital 1].. whereas [at Hospital 2] they come to [cardiology] first and then they are sorted out" – Clinical Stakeholder Group*

> *"[Hospital 1] are clearly allocating to triage 2 in the more protocolised way... the protocol says only a tiny number of chest pain patients should ever go home directly from the emergency department because they've all got to have [a specific set of tests] etcetera" – Clinical Stakeholder Group*

However, similar to the critique on the use of key performance indicators (KPI) in health, a myopic comparison of only the *Comparison Points*, or certain steps in the processes are open to gaming, manipulation and misinterpretation [Scobie et al. 2006]. Thus, an end-to-end view of clinical pathways is important to ensure that points of comparison are objectively and quantitatively contextualised by preceding and subsequent routing of patients in the pathways. In our case study, we attempt to minimise this problem by using discovered process models in the discussions with the stakeholders to visually 'prompt' them to consider the 'broader' view of an end-to-end patient pathway in identifying key *Comparison Points*.

Communicating Results. We also found that the visualisation of processes for comparing activities across hospitals was crucial for engaging clinical stakeholders. This is one of the key benefits the process mining analysis.

However, while the heuristically mined workflow maps and fuzzy models were helpful for understanding the concepts of relative flow, these models did not represent the timing element within a single static diagram. Further, there seemed to be no tool available which could effectively display both timing and control flow perspectives across multiple process models for cross-comparison. For this reason, the BPMN diagrams together with timing statistics were used to report the main workflow variants, the quantitative routing statistics, and performance statistics.

As such, it seems that further work is required to help improve the visualisation and communication of multiple flow and performance perspectives across multiple process models, so that a more comprehensive comparison of an end-to-end process model can be facilitated.

4.2. Next Steps

An immediate next step that we need to undertake in this case study, is the validation of the data: through clarifying definitional terms and ontologies, auditing the collection of practice norms across hospitals, and measuring the extent to which cross-hospital data is maintained. In the meantime, we are investigating the use of discretised pathology and imaging test results for characterising the risk/severity of each case, which can be incorporated as an attribute within the event logs by which comparable ACS populations can be filtered into separate logs.

The major next step, however, is that the comparative process analyses displayed in this case study must now be *combined with other comparative analyses* that focus on the outcomes of the observed differences in patient pathways. In particular, we need to link our study with those that directly evaluate the cost-effectiveness of alternative forms of service provision, by employing statistical methods to capture long-run cost and health outcomes for patients, e.g. [Karnon et al. 2013; Pham et al. 2012]. By linking these two analysis perspectives (i.e. our case study which focus on the process perspective and those studies that focus on the long-term cost-effectiveness of services) through common identifiers, we would then be positioned to investigate the economic impact and efficiency of practice changes and investments. This can address the question such as the one posted by the stakeholders:

> *"... what happens if [Hospital 1] pushes everyone to cardiology [like Hospital 2]? Well you might have to employ 6 more cardiologists and let go a number of general physicians... Or is it that [Hospital 2] are overly admitting patients to cardiology?" – Clinical Stakeholder Group*

5. CONCLUSION

Systems of clinical practice are not designed to be fully automated and rely on the expertise of clinical staff to apply care as appropriate to individual patient needs. However, initiatives are needed to ameliorate uncertainty and disagreement among clinical, executive and political stakeholders, with respect to the most appropriate tasks, decision criteria, work procedures, and performance measures. Ideally, such initiatives will be driven by empirical records of activity and objective analyses of processes and outcomes.

In this paper, we have presented preliminary findings from a case study of comparative process mining, which utilises routinely-collected data to describe differences in the process of care, as delivered at four Australian hospitals.

The existing case study results highlight key differences in the way the hospitals applied ACS management, and confirm the existing but implicit assumptions of the clinical stakeholder group, of the variations in clinical practice. While this itself is a fruitful outcome, this case study also demonstrates the potential for cross-organisational process mining to be used to inform and monitor decisions on changes in practice, and of investment decisions.

Future work is needed to improve the visualisation of comparative analyses, and to link observed processes with health and fiscal outcomes. In achieving this, insights from comparative process mining will help identify inefficiencies in process variations, and will help to inform health care quality and funding reform programs in Australia.

REFERENCES

K. Antioch, G. Jennings, M. Botti, R. Chapman, and V. Wulfsohn. 2002. Integrating cost-effectiveness into clinical practice guidelines in Australia for acute myocardial infarction. *European Journal of Health Economics* 3, 1 (2002), 26–39.

M. Binder, W. Dorda, G. Duftschmid, R. Dunkl, K.A. Froschl, W. Gall, W. Grossmann, K. Harmankaya, M. Hronsky, S. Rinderle-Ma, C. Rinner, and S. Weber. 2012. On Analyzing Process Compliance in Skin Cancer Treatment: An Experience Report from the Evidence-Based Medical Compliance Cluster (EBMC2). In *LNCS 7328*, J. Ralyte (Ed.). Springer, 398–413.

Tobias Blum, Nicolas Padoy, Hubertus Feuner, and Nassir Navab. 2008. Workflow mining for visualisation and analysis of surgeries. *International Journal of Computer Assisted Radiology and Surgery* 3, 5 (2008), 379–386. http://link.springer.com/article/10.1007/s11548-008-0239-0?LI=true#page-1

R. Bose and W. van der Aalst. 2012. Analysis of patient treatment procedures. In *Lecture Notes in Business Information Processing*, Vol. 99. 165–166.

Productivity Commission. 2005. *Impacts of Advances in Medical Technology in Australia*. Research report. Productivity Commission, Melbourne. http://www.pc.gov.au/projects/study/medical-technology/docs/finalreport

L. De Bleser, R. Depreitere, K. De Waele, K. Vanhaecht, J. Vlayen, and W. Sermeus. 2006. Defining pathways. *J. of Nursing Management* 14, 7 (October 2006), 553–63.

R. Dunkl, K. Fröschl, W. Grossman, and S. Rinderle-Ma. 2011. Assessing medical treatment compliance based on formal process modeling. In *Lecture Notes in Computer Science*, Vol. 7058. 533–546.

N. Eckard, M. Janzon, and LA. Levin. 2011. Compilation of cost-effectiveness evidence for different heart conditions and treatment strategies. *Scandinavian Cardiovascular Journal* 45, 2 (April 2011), 72–6. http://www.ncbi.nlm.nih.gov/pubmed/21329415

H. Elghazel, V. Deslandres, K. Kallel, and A. Dussauchoy. 2007. Clinical pathway analysis using graph-based approach and Markov models. In *2007 2nd International Conference on Digital Information Management (ICDIM 2007)*. 279–284.

C. Fernandez-Llatas, T. Meneu, J. Benedi, and V. Traver. 2010. Activity-Based Process Mining for Clinical Pathways Computer Aided Design. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.

D. Ferreira and C. Alves. 2011. Discovering user communities in large event logs. In *Lecture Notes in Business Information Processing*.

Acute Coronary Syndrome Guidelines Working Group. 2006. Guidelines for the management of acute coronary syndromes. *The Medical Journal of Australia* 184 (8 Suppl) (2006), S1–S32. https://www.mja.com.au/journal/2006/184/8/guidelines-management-acute-coronary-syndromes-2006

C. Gunther and W. Van der Aalst. 2007. Fuzzy mining- Adaptive process simplification based on multi-perspective metrics. In *Business Process Management Proceedings*. Springer.

Christian W. Günther. 2009. *Process mining in flexible environments*. Ph.D. Dissertation. Technische Universiteit Eindhoven.

S. Gupta. 2007. *Workflow and Process Mining in Healthcare*. Master's thesis. Eindhoven University of Technology.

B. Han, L. Jiang, and H. Cai. 2011. Abnormal process instances identification method in healthcare environment. In *Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011)*. 1387–1392.

P.T.G Hornix. 2007. *Performance Analysis of Business Processes through Process Mining*. Master's thesis. Eindhoven University of Technology.

Z. Huang, X. Lu, and H. Duan. 2012. On mining clinical pathway patterns from medical behaviors. (2012). In Press.

F. M. Impellizzeri, M. Bizzini, M. Leunig, N. A. Maffiuletti, and A. F. Mannion. 2009. Money matters: exploiting the data from outcomes research for quality improvement initiatives. *European Spine Journal* 18, Suppl 3 (2009), 348–59. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2899321/

R. Janssen. 2011. *Increasing accessibility and reproducibility of process mining in healt*. Master's thesis. Eindhoven University of Technology.

J. Karnon, D. Ben-Tovim, C. Pham, O. Caffrey, P. Hakendorf, M. Crotty, and P. Phillips. 2011. The efficient price: an opportunity for funding reform. *Australian Health Review* 35, 4 (2011), 501–506. http://www.ncbi.nlm.nih.gov/pubmed/22126956

Jonathan Karnon, Orla Caffrey, Clarabelle Pham, Richard Grieve, David Ben-Tovim, Paul Hakendorf, and Maria Crotty. 2013. Applying Risk Adjusted Cost-Effectiveness (RAC-E) Analysis to Hospitals: Estimating the costs and consequences of variation in clinical practice. *Health Economics* 22, 6 (2013), 631–642. DOI: http://dx.doi.org/10.1002/hec.2828

P. Kennedy, C. Leathley, and C. Hughes. 2010. Clinical practice variation. *Medical Journal of Australia* 193, 8 (2010), S97–S99.

M. Kuo and Y. Chen. 2012. A method to identify the difference between two process models. *Journal of Computers* 7 (2012), 998–1006.

Fluxicon Process Laboratories. 2012. Disco. (2012). http://fluxicon.com/disco/

M. Lang, T. Burkle, S. Laumann, and H. Prokosch. 2008. Process Mining for Clinical Workflows: Challenges and Current Limitations. In *Ehealth Beyond the Horizon- Get There*. 229–234.

R. Lenz and M. Reichert. 2007. IT support for healthcare processes- premises, challenges, perspectives. *Data and Knowlegde Engineering* 61(1) (2007), 39–58.

A. Manninen. 2010. *Applying the principles of process mining to finnish healthcare*. Master's thesis. Aalto University.

R. Mans, M. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, and W. van der Aalst. 2008a. Process Mining Techniques: an Application to Stroke Care. In *Ehealth Beyond the Horizon- Get it There*.

R. Mans, M. Schonenberg, M. Song, W. van der Aalst, and P. Bakker. 2008b. *Application of Process Mining in Healthcare- A Case Study in a Dutch Hospital*. Springer-Verlag, 425–438.

R. Mans, W. van der Aalst, R. Vanwersch, and A. Moleman. 2012. Process Mining in Healthcare: Data challenges when answering frequently posed questions. In *Proceedings of BPM 2012 Workshops*. http://www.processmining.org/_media/publications/processminingprohealth3r1.pdf

C. McGregor, C. Catley, and A. James. 2011. A Process Mining Driven Framework for Clinical Guideline Improvement in Critical Care. In *Learning from Medical Data Streams 13th Conference on Artificial Intelligence in Medicine (LEMEDS)*. 35–46.

M. Peleg, P. Soffer, and J. Ghatlas. 2007. Mining process execution and outcomes- Position paper. In *5th International Conference on Business Process Management*.

R. Perez-Castillo, B. Weber, J. Pinggera, S. Zugal, G. de Guzman, and M. Pialtini. 2011. Generating event logs from non-process-aware systems enabling business process mining. *Enterprise Information Systems* 5 (2011), 301–335.

L. Perimal-Lewis, S. Qin, C. Thompson, and P. Hakendorf. 2012. Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach. In *HIKM 2012 (CRPIT)*, Vol. 129. ACS, 59–66. http://crpit.com/confpapers/CRPITV129Perimal-Lewis.pdf

C. Pham, O. Caffrey, J. Karnon, D. Ben-Tovim, P. Hakendorf, and M. Crotty. 2012. Evaluating the effects of variation in clinical practice: a risk adjusted cost-effectiveness (RAC-E) analysis of acute stroke services. *BMC Health Services Research* 12 (2012), 266. http://www.biomedcentral.com/1472-6963/12/266

J. Poelmans, G. Dedene, G. Verheyden, H. Van der Mussele, S. Viaene, and E. Peters. 2010. Combining Business Process and Data Discovery Techniques for Analyzing and Improving Integrated Care Pathways. In *ICDM 2010*, Vol. 6171. Springer, 505–517.

M. Poulymenopoulou, F. Malamatniou, and G. Vassilacopoulos. 2003. Specifying workflow process requirements for an emergency medical service. *Journal of Medical Systems* 27(4) (2003), 325–335.

S. Quaglini. 2010. Information and communication technology for process management in healthcare: a contribution to change the culture of blame. *Journal of Software Maintenance and Evolution: Research and Practice* 22 (2010), 435–448.

A. Rebuge and D. Ferreira. 2012. Business Process Analysis in Healthcare Environments: a Methodology based on Process Mining. *Information Systems* 37 (2) (2012), 91–116.

LL. Jr. Roos and A. Wajda. 1991. Record linkage strategies. Part I: Essential information and evaluation approaches. *Methods of information in medicine* 32, 2 (1991), 117–23.

LL. Jr. Roos, A. Wajda, and JP. Nicol. 1986. The art and science of record linkage: methods that work with few identifiers. *Computers in Biology and Medicine* 16, 1 (1986), 45–57.

W.B. Runciman, T.D. Hunt, N.A. Hannaford, P.D. Hibbert, J.I. Westbrook, E.W. Coiera, R.O. Day, D.M. Hindmarsh, E.A. McGlynn, and J. Braithwaite. 2012. CareTrack: assessing the appropriateness of health care delivery in Australia. *MJA* 197, 2 (2012), 100–105.

S. Scobie, R. Thomson, J. McNeil, and P. Phillips. 2006. Measurement of the safety and quality of health care. *MJA* 184 (2006), S51–S55. background on appropriateness of measurements to inform CQI initiatives in Australia.

I.A. Scott. 2003. Determinants of Quality of In-Hospital Care for Patients with Acute Coronary Syndromes. *Disease Mgmt. and Health Outcomes* 11, 12 (2003), 801–816.

I. Scott, C. P. Denaro, A. C. Hickey, C. Bennett, A. M. Mudge, D. C. Sanders, J. Thiele, and J. L. Flores. 2004. Optimising care of acute coronary syndromes in three Australian hospitals. *Int. Jour. for Quality in Health Care* 16, 4 (2004), 275–284.

M. Song, C. Gunther, and W. Van der Aalst. 2009. Trace Clustering in Process Mining. In *BPM Workshops*. 109–120.

J. Staal. 2010. *Using process and data mining techniques to define and improve standardization in a healthcare workflow environment*. Master's thesis. Eindhoven University of Technology.

W. van der Aalst. 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag.

Wil van der Aalst, Arya Adriansyah, Ana Karla Alves de Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter van den Brand, Ronald Brandtjen, Joos Buijs, and others. 2012. Process mining manifesto. In *Business process management workshops*. Springer, 169–194.

W. van der Aalst et al. 2012. Process Mining Manifesto. In *BPM Workshops 2011*, Vol. 99. Springer, 169–194.

Jan vom Brocke, Alexander Simons, Bjoern Niehaves, Bjorn Niehaves, Kai Reimer, Ralf Plattfaut, and Anne Cleven. 2009. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In *Paper presented at the 17th European Conference on Information Systems (ECIS 2009)*. 2206–2217.

A. J. M. M. Weijters, W. van der Aalst, and A. K. Alves de Medeiros. 2006. Process Mining with the Heuristics Miner Algorithms. *BETA Working Paper Series* WP 166 (2006).