Project Details:	HCF Research Foundation Project Grant, 2012 - 2014
Report Title:	**Comparing hospital performance to reduce variation in clinical practice: it's not just the data, it's how you use it**
Authors:	Andrew Partington[1], Jodi Gray[1], Matthew Horsfall[2], Derek Chew[3,5], Suriadi Suriadi[4], Chun Ouyang[4], Moe Thandar Wynn[4], David Ben-Tovim[3], Paul Hakendorf[3,5], Jonathan Karnon[1].
Author Affiliations:	[1] University of Adelaide [2] South Australian Health and Medical Research Institute [3] Flinders University [4] Queensland University of Technology (QUT) [5] The South Australian Department of Health (SA Health)
Corresponding Author:	Professor Jonathan Karnon, School of Population Health, University of Adelaide, Level 7, 178 North Terrace, Adelaide, SA 5005 Tel: 08 83133562 Email: jonathan.karnon@adelaide.edu.au

Executive Summary

Background

In all developed countries, there is a funding bias in favour of new technologies, ahead of strategies to improve existing services. In Australia, the conduct of quality improvement initiatives are generally funded at the level of the hospital, i.e. with funds diverted directly from the provision of services. This inevitably means that quality improvement is assigned a low priority.

Decisions around the appropriate allocation of funds to new technologies and to quality improvement interventions are constrained by the lack of common processes to assess the costs and benefits of spending in both broad areas. With a common analytic framework, we could start to assess the relative value of funding new technologies versus quality improvement.

Estimates of the potential value of quality improvement provide a basis for prioritising areas for action, where the capacity for benefit is greatest. The subsequent step is to assess and address the causes of unwarranted variation, which may include structural, systems, and individual-level factors, across multiple interfaces.

Analyses of variation in costs, outcomes, and processes of care provide a basis for improving hospital performance, but the subsequent implementation of a quality improvement initiative requires potentially significant time investments from clinical and managerial stakeholders, and scarce health care funds to support quality improvement initiatives.

There has been growing interest in the public reporting of performance measures, and pay-for-performance as levers to incentivise quality improvement. A recent review reported strong and consistent evidence that public reporting of performance data incentivised quality improvement activities at a hospital level. Effects on clinical outcomes were mixed. The evidence on P4P is less promising. A recent review of reviews reported limited and inconsistent effects of P4P, especially in hospitals.

Aims and Methods

The overall aim was to demonstrate the value of linked, routinely collected clinical and administrative hospital data to prioritise clinical processes for improvement activities, and to inform and support such activities

To meet this aim, the following three sub-studies were completed:

- Study 1: A literature review, and preliminary application of process mining methods to health care data.
- Study 2: An applied comparative analysis of costs, outcomes, and processes of care to prioritise and inform quality improvement in hospitals.
- Study 3: A systematic literature review of the effectiveness of alternative approaches to the feedback of comparative hospital performance data to improve processes of care and patient outcomes.

Results

The findings from a literature review of the application of process mining techniques to health care, and a primary application of process mining techniques to health care data included the following:

- Mining based analyses of workflow and performance are useful tools for representing processes at individual hospitals, and for initiating conversations regarding comparative practices,
- The cognitive load of comparing complex processes across multiple sites, whilst controlling for potential differences in casemix, was large,
- Cluster analyses (the mining based identification of process clusters at each hospital) have the potential to highlight process differences within common casemix groupings, but these analyses were hampered by the need to represent continuous timing variables (e.g. time in the ED).

The main empirical study demonstrated that routinely collected health care data can inform relevant and robust evidence of important variation in the costs and outcomes of health care provided at alternative hospitals. Supporting analyses of key process indicators provide meaningful insights on important differences in the process of care for clinically relevant sub-groups of the aggregate eligible population.

At an aggregate level, statistically significant, casemix-adjusted differences were observed in mean inpatient costs (up to $672 per admitted patient), and 30 day and 12 month cardiovascular or mortality event rates (odds ratios up to 2.42 and 1.64, respectively) across providers. The analysis of costs and patient outcomes did not identify a single benchmark hospital, but rather identified an apparent outlier (non-benchmark) hospital that was incurring higher costs and achieving poorer patient outcomes than other hospitals.

The non-benchmark hospital had the lowest inpatient admission and active management rates, which was driven by much lower rates for patients who might initially be categorised as being at low risk of a poor health outcome (i.e. with a negative diagnostic test and no pre-existing circulatory condition on presentation). These results suggest that the clinical pathway for patients presenting with chest pain, with a negative troponin test, might be reviewed at Hospital 2.

The non-benchmark hospital also reported longer inpatient lengths of stay for higher risk patients (those with a positive diagnostic test result), which may be driving increased in costs at this hospital.

The third study was a systematic literature review of the feedback of comparative performance data to hospitals, with a specific focus on services for acute coronary syndromes. The review provided moderate evidence of effect, and suggest that the reporting of comparative performance data is not necessarily sufficient to motivate significant and sustained improvements in patient outcomes.

From the review, we identified three components of the evaluated approaches to the feedback of comparative performance data that appear to be important:

- quality of the comparative data;
- form and focus of feedback;
- ongoing support for quality improvement activities.

Discussion

The findings from the review support the potential value of comparative performance data, whilst emphasising the need for such data to be used in the context of broader, well designed quality improvement systems that target key stakeholders and provide support to implement and maintain quality improvement activities.

Timely access to routinely collected data is a key prerequisite for the use of comparative performance data to inform quality improvement in health care. Delays in accessing the data precluded their planned application to inform quality improvement activities in the study hospitals. Given the importance of timeliness of the analysis to inform investment in quality improvement, and the time lag in accessing administrative data, important ongoing analyses will assess the validity of analysing hospital performance using clinical data alone.

A report from the King's Fund in the UK recommended a program of work to identify causes of variation at specific local levels, and to prioritise those variations and causes that have the most important impact on equity, effectiveness, efficiency and patient health outcomes [Appleby et al, 2011]. The reported methods of analysis should be applied to a range of clinical areas, using relevant data (i.e. those available in a timely manner). Such analyses could support the ongoing comparative analysis of performance, as part of a transparent process of selecting priority clinical areas and providers for quality improvement intervention.

The King's Fund report also suggested that the reporting of unwarranted variations may not incentivise sufficient action. Based on the reviewed literature around the use of comparative performance data as part of a quality improvement program, such feedback interventions require careful design and evaluation.

Table of Contents

Chapter 1	Background and Study Aims	…….....1
Chapter 2	Process mining of health care data: a review and preliminary application	...…….....5
Chapter 3	Comparative analysis of costs, health outcomes, and processes of care: a case study in patients presenting at hospital with chest pain	…….18
Chapter 4	A systematic literature review of the effectiveness of alternative approaches to the feedback of comparative hospital performance data	…….……28
Chapter 5	Conclusions	…….……42
References		…….46
Appendix 1	Literature Search strategy	…….……52
Appendix 2	Quality Criteria Checklist	…….53

Chapter 1 Background and Study Aims

In general, health services, and hence patient outcomes, can either be improved by investing in new, more effective technologies (e.g. new pharmaceuticals or medical devices), or by making better use of existing resources and associated technologies (e.g. quality improvement).

There is ever increasing evidence of variation in the use of existing resources and technologies (e.g. non-adherence to clinical guidelines) [Runciman et al, 2012], as well as variation in the organisation and delivery of health care across alternative providers [Kennedy et al, 2010]. Some of this variation may be warranted on the basis of relevant differences in the clinical, geographical, and socioeconomic characteristics of patient populations. Areas of unwarranted variation in clinical practice should be identified, and actions implemented to improve service delivery in these areas.

This scenario implies a two-stage process of identifying unwarranted variation, followed by the implementation of a quality improvement process. But the health care system is large and complex, and resources are limited. How should we identify the most important areas of unwarranted variation, and how much should we spend on quality improvement when we identify such variation?

Compared to spending on new technologies, budgets allocated to quality improvement remain small. In Australia, even with new legislation to reduce government subsidies for medicines, spending by the Pharmaceutical Benefits Scheme (which does not cover all public spending on pharmaceuticals) rose by $500 million in 2010/11. The annual budget for the Australian Commission on Safety and Quality in Health Care (ACSQHC) is less than $15 million. In the UK, NHS Improving Quality (IQ) commenced in April this year, with the aim of implementing effective improvement programmes and building improvement capacity. The NHS IQ budget is not clear, but it is unlikely to match new spending on pharmaceuticals in the NHS, which increases by around £400 million per year.

Decisions around the appropriate allocation of funds to new technologies and to quality improvement interventions are hampered by the lack of common processes to assess the costs and benefits of spending in both broad areas. If we could apply a common analytic framework, we could start to assess the relative value of funding new technologies versus quality improvement.

1.1. Valuing new technologies

The concept of value-based pricing is applied to decisions regarding the funding of new technologies [Claxton et al, 2008], such as pharmaceuticals and medical devices. Bodies such as the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, and the National Institute for Health and Clinical Excellence (NICE) in England, commonly use decision analytic models to predict differences in costs and quality adjusted life years (QALYs) over the remaining lifetime of patients receiving competing interventions.

The cost-effectiveness of new technologies are generally presented in the form of an incremental cost-effectiveness ratio (e.g. $30,000 per QALY gained), which is interpreted as

the additional resources to be spent on the new technology in order to gain one additional QALY. To inform funding decisions, PBAC uses an implicit threshold of around $40,000 to $50,000 per QALY gained, which is intended to represent the opportunity cost of spending on new technologies. This means that the best alternative option to funding a new technology is assumed to have an incremental cost-effectiveness ratio of around £40,000 per QALY gained. So, if a new technology gains additional QALYs at less than $40,000, it is considered to provide better value than the best alternative funding option and should be funded.

1.2. Valuing quality improvement

To inform the value of quality improvement, and hence decisions regarding investment to improve existing services, the effects of improvement initiatives should be estimated with respect to service costs and patient outcomes. In line with the value-based pricing of new technologies, decisions to intervene to improve hospital performance should reflect the expected costs of identifying and implementing improved healthcare provision, and the value of improved patient outcomes.

The linkage of de-identified individual-level hospital and mortality data provides longitudinal data on the whole populations of patients treated at alternative hospitals. In Australia, bottom up costing to inform casemix funding provides individual-level cost data for most inpatient episodes. Previous studies have reported analyses of differences in costs and outcomes, using individual-level administrative hospital data to identify benchmark providers (e.g. those achieving the best incremental cost-effectiveness ratio across the comparator hospitals) [Pham et al, 2012; Karnon et al, 2013]. Such analyses also inform the expected value of benchmark performance, i.e. the gains in costs and outcomes (efficiency) associated with improving performance to that of the benchmark provider(s).

Previous comparative analyses of costs and patient outcomes across hospitals have relied on administrative hospital data, which is collated by hospitals for reporting to a centralized body (e.g. the National Hospital Cost Data Collection in Australia). The advantages of such data for informing hospital performance include that they are routinely collected in a standardized manner by all hospitals. Disadvantages include the potentially limited nature of the data collected, and the lag between data collection and availability of the data for analysis.

Clinical data comprises data recorded during the process of managing patients, and may include details of care provided, such as the timing and results of diagnostic tests and investigations, and interventions. The advantages of clinical data include the additional detail on the clinical pathway, and the potential for quicker access to the data. The main disadvantage is the non-routine collection and reporting of clinical data, which means that primary data extraction has to be undertaken. However, with increasing use of electronic data systems, the extraction of large amounts of clinical data is more feasible than previously, when such data were manual extracted from paper-based clinical records.

An aim of the current study was to improve the representation of casemix at alternative hospitals through the use of clinical and administrative data to compare costs and outcomes. Casemix adjustment is the basis for adjusting observational data to account for differences in the underlying risk or prognosis of presenting patients.

1.3. Supporting Quality Improvement

Estimates of the potential value of quality improvement provide a basis for prioritising areas for action, where the capacity for benefit is greatest. The subsequent step is to assess and address the causes of unwarranted variation, which may include structural, systems, and individual-level factors [Lilford et al, 2004], across multiple interfaces (e.g. emergency, acute, sub-acute, etc.).

The expansion of the dataset used to compare costs and outcomes across hospitals, to include clinical as well as administrative data, also facilitated more in depth analysis of the processes or pathways of care at the comparator hospitals. The clinical data provided information on a wide range of aspects of the clinical pathway experienced by presenting patients, including:

- Mode and time of ED presentation,
- Time and destination of ED discharge (e.g. to home or inpatient ward),
- Timing and results of tests and procedures.

Another aim of the study was to analyse the collated dataset to identify specific areas of variation in the processes of care at the comparator hospitals. Interpreted in conjunction with the comparative analyses of costs and outcomes, particular areas of the clinical pathway may then be targeted for quality improvement at non-benchmark hospitals.

To compare processes of care across hospitals, process mining techniques were applied to the data. Process mining applies data mining techniques to process information, in this case, searching for patterns in processes across patients presenting at each hospital. The methodology originated in the field of business process engineering, but has recently begun to be applied in the health care setting.

1.4. Funding and Incentivising Quality Improvement

Analyses of variation in costs, outcomes, and processes of care provide a basis for improving hospital performance, but the subsequent implementation of a quality improvement initiative comprises multiple stages: (1) primary investigation of the causes of variation from benchmark practice, (2) identification of potential barriers and facilitators to quality improvement, (3) decisions regarding appropriate actions, (4) implementation of the defined improvements, and (5) post-implementation evaluation. This requires potentially significant time investments from clinical and managerial stakeholders, and scarce health care funds to support quality improvement initiatives.

This study also aimed to assess the use and feedback of the comparative costs, outcomes, and processes of care data, to inform actions to improve quality at non-benchmark hospitals.

There has been growing interest in the public reporting of performance measures, and pay-for-performance (P4P) as levers to incentivise quality improvement. A review published in 2010 reported strong and consistent evidence that public reporting of performance data incentivised quality improvement activities at a hospital level [Chen, 2010]. Effects on clinical outcomes were mixed. Chen noted that more recent studies have better study design, data, and analytic frameworks, and are more likely to show a positive effect of public

reporting on clinical outcomes. The evidence on P4P is less promising. Despite some success stories [Sutton et al, 2012], a recent review of reviews reported limited and inconsistent effects of P4P, especially in hospitals [Eijkenaar et al, 2013].

On the basis of the above studies, a systematic literature review was undertaken to assess the effectiveness of alternative approaches to the feedback of comparative hospital performance data. The current review is not restricted to public reporting, but the intervention must involve the feedback of comparative performance data to hospitals. No restrictions are placed on other aspects of the evaluated interventions, for example, the feedback of comparative data may be the sole component of an intervention, or feedback might be combined with additional components that are intended to support quality improvement initiatives. Thus, the review is intended to capture the full range of applied approaches to using comparative performance data to incentivise improved hospital performance.

1.5. Project aims

The stated aims of the project were to demonstrate the potential value of linked, de-identified clinical and administrative hospital data to:

- Inform priority areas for quality improvement, through analyses of the incremental costs and effects of care provided by alternative hospitals for specific conditions,
- Identify significant differences in the processes of care provided at alternative hospitals, to provide a rationale for any estimated differences in costs and outcomes, as well as providing targets for service improvement,
- Feedback comparative data on costs, patient outcomes, and processes of care to inform quality improvement initiatives.

In the light of delays in accessing the linked, clinical and administrative hospital data, additional objectives were specified:

- Application of process mining techniques on existing datasets to investigate and develop appropriate methods for the analysis of health care data,
- A systematic literature review to identify the most effective methods for using comparative hospital performance data to improvement the quality of hospital care.

The following sections of the report describe the conduct and findings of the three sub-studies that were completed over the course of the funded project:

- Study 1: An explorative application of process mining methods to health care data.
- Study 2: An applied comparative analysis of costs, outcomes, and processes of care to prioritise and inform quality improvement in hospitals.
- Study 3: A systematic literature review of the effectiveness of alternative approaches to the feedback of comparative hospital performance data to improve processes of care and patient outcomes.

The final section of the report provides a general discussion of the findings, priority areas for further research, and conclusions.

Chapter 2 Process mining of health care data: a review and preliminary application

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 of operational processes [van der Aalst 2011]. To better inform the planned comparative analysis of processes of care across health care providers, two sub-studies were undertaken:

The study reported by Partington et al. [2014] comprised a literature review and detailed case study, outlining the landscape for comparative analyses of healthcare, and the insights that mining of process data may provide to support performance improvement initiatives. Focused on both emergency department (ED) and inpatient workflow and timing, the authors detailed their methods for pre-analysis data manipulation and preparation and reported on comparative findings associated with admission rates, throughput timing, procedure use and length of stay (LoS). This paper contributed to both the understanding of working with clinical data for process analysis.

Looking at ED workflows, the second study focussed specifically on comparing hospital activity using *conformance analyses* methods that 'replay' data on previously mined workflow models [Suriadi et al, 2014]. The analysis attempted to identify comparative sub-groups of patients through semi-supervised clustering, based on patient and process attributes.

Drawing on these two studies, this section provides a detailed synopsis of the analysis of process of care data with respect to the following issues:

2.1. Conceptualisation of 'health care processes';
2.2. Process mining: underlying concepts;
2.3. Literature Review of the application of process mining within health care;
2.4. Data management and manipulation requirements for analyses of process;
2.5. Discovery of workflow and timing statistics from data;
2.6. Visualisation of workflows and timing statistics for clinical engagement;
2.7. Practical insights and information feedback; and
2.8. Methodological lessons learned for further analyses within this project.

2.1 Conceptualisation of '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 'non-trivial' as the steps involved are often non-linear and do not necessarily exist within a planned structure of sequences to the same extent as the 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, but on many other factors, such as a patient's overall health condition, his/her reaction to therapy, the dynamic professional environment with rapidly changing procedural options, the multi-disciplinary interaction of highly-specialised knowledge areas, patient/doctor preferences, and availability of relevant resources (e.g., a bed, a clinician) are all inherent in the decisions to execute tasks [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, as those that we are focused on in this project.

2.2 Process mining: underlying concepts

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 (resource) perspective, the case perspective, and the time perspective [van der Aalst 2011]. The control-flow perspective focuses on the ordering of activities. The perspective is concerned with analyzing 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 the 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 Literature Review of the application of process mining within 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]. To further explore potentially interesting application areas, 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.

Table 2.1 summarizes the evaluation of the 28 related papers. 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 2.1 – 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 was found to be an important step as health data is often collated from heterogeneous sources and is often fragmented. An explanation of how each study manipulated the data into a form that was suitable for process mining analysis was thus valuable knowledge. Approximately half of the reviewed papers made reference to data pre-processing activities [Mans et al. 2008b; Bose and van der Aalst 2012; Staal 2010; Binder et al. 2012; Rebuge & 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 & Alves 2011; Perimal-Lewis et al. 2012]. Specifically, a recent study conducted by Mans et al. [Mans et al. 2012] details different types of data encountered within four Dutch hospitals' information systems, and illustrates options for using the data to address frequently posed questions by clinicians. Nevertheless, given that only 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:

- The majority of papers (82%) covered process discovery techniques [Mans et al. 2008b; Mans et al. 2008a; Quaglini 2010; Lang et al. 2008; Bose & van der Aalst 2012; Staal 2010; Binder et al. 2012; Poelmans et al. 2010; Gunther & van der Aalst 2007; Perez-Castillo et al. 2011; Rebuge & Ferreira 2012; Song et al. 2009; Ferreira & 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].
- Six studies reported on the use of conformance analysis [Mans et al. 2008b; Dunkl et al. 2011; Binder et al. 2012; Quaglini 2010; Peleg et al. 2007; Kuo and Chen 2012],
- One paper reported on process enhancement [Mans et al. 2008b].

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 & van der Aalst. 2012; Staal 2010; Binder et al. 2012; Pelegetal 2007; Poelmansetal 2010; Gunther & van der Aalst. 2007; Perez-Castillo et al. 2011; Rebuge & 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]

- 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]
- Three studies (11%) reported on the organizational perspective [Mans et al. 2008b; Rebuge and Ferreira 2012; Ferreira and Alves 2011].

The mining of the organisational, time, and case perspectives seems to have been over-looked in the health care setting.

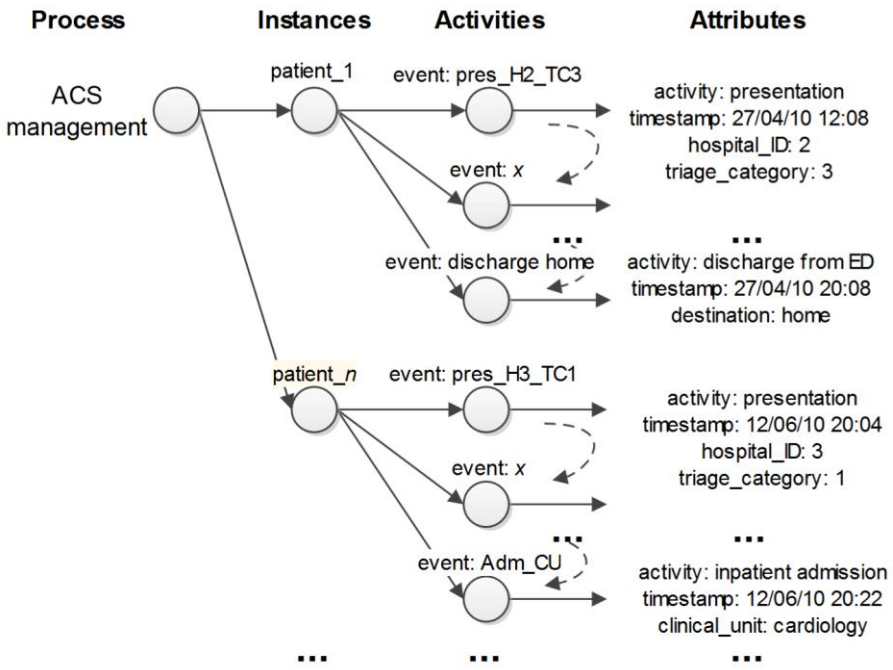
For the purposes of informing methods and approaches relevant to our study, each paper was also evaluated to see if comparative analyses were attempted. Amongst the 28 identified papers, only one paper presented a comparative analysis of health care processes [Mans et al. 2008a]. This study analysed data from 368 patients diagnosed with 'first-ever ischemic stroke' from four Italian hospitals to compare procedures involved in the treatment of stroke patients. With only one comparative analysis study identified, the use of cross-organisational process mining for comparative analysis purposes has been under-exploited.

2.4 Data management and manipulation requirements for analyses of process

Variables of interest to the analysis were identified through a review of the Australian clinical guidelines for the management of ACS [Aroney et al. 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 practices exist across the hospitals. Annonymisation of patient records was applied at the extraction level in order to preserve privacy, and in accordance with ethics committee clearance.

After extraction, the anonymised data was reformatted from a *case-log* data format, into transactional *event logs* (see Figure 2.1). 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. 2.1: Structure of data within event logs, based on [van der Aalst 2011]

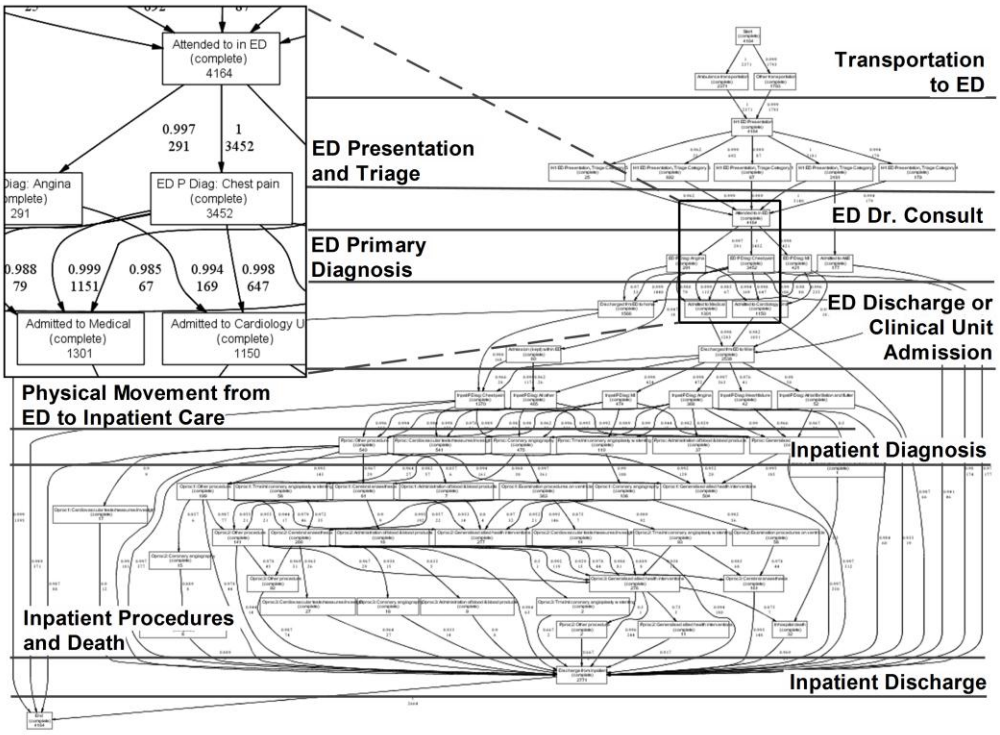


2.5 Discovery of workflow and timing statistics from data

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 this study suggested 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 2.2, 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 2.2, 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 2.2 shows that there exist many alternate pathways that a patient may follow.

The performance analysis with the petri-net plug-in within ProM [Hornix 2007] was also applied, in order to replay the event log data through the events and transitions obtained from the heuristically mined workflow models (as in Fig. 2.2). 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.

Fig. 2.2: Illustrative workflow net with nine process layers



2.6 Visualisation of workflows and timing statistics for clinical engagement

We found that the visualisation of processes for comparing activities across hospitals was crucial for engaging clinical stakeholders, and was one of the key benefits of the process mining analyses. However, while the heuristically mined workflow maps were helpful for understanding the concepts of relative flow, these models were often unintelligible to anyone but the analyst and did not represent key timing elements within a single static diagram. Further, we required a way to effectively display both timing and control flow perspectives across multiple process models for cross-comparison.

2.6.1 BPMN

Business Process Modelling Notation (BPMN) diagrams together with timing statistics were employed, as a first step, to help report the main workflow variants, the quantitative routing statistics, and performance statistics. In order to do this, we abstracted a ‘standard’ set of pathways, reflective of common activities across the hospitals that were evident within the mined workflow models. The resulting 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 2.3 and 2.4 depict the pathways traversed by patients who received a similar (ICD-10-AM, R07 Chest Pain) diagnosis at one of four hospitals, and illustrates the range of activities, the alternate pathways, the frequency of events, and the timing [mean HH:mm (standard deviation HH:mm)] associated with LoS.

Fig. 2.3: BPMN models illustrating the Flow & Timing for Hospitals 1 & 2

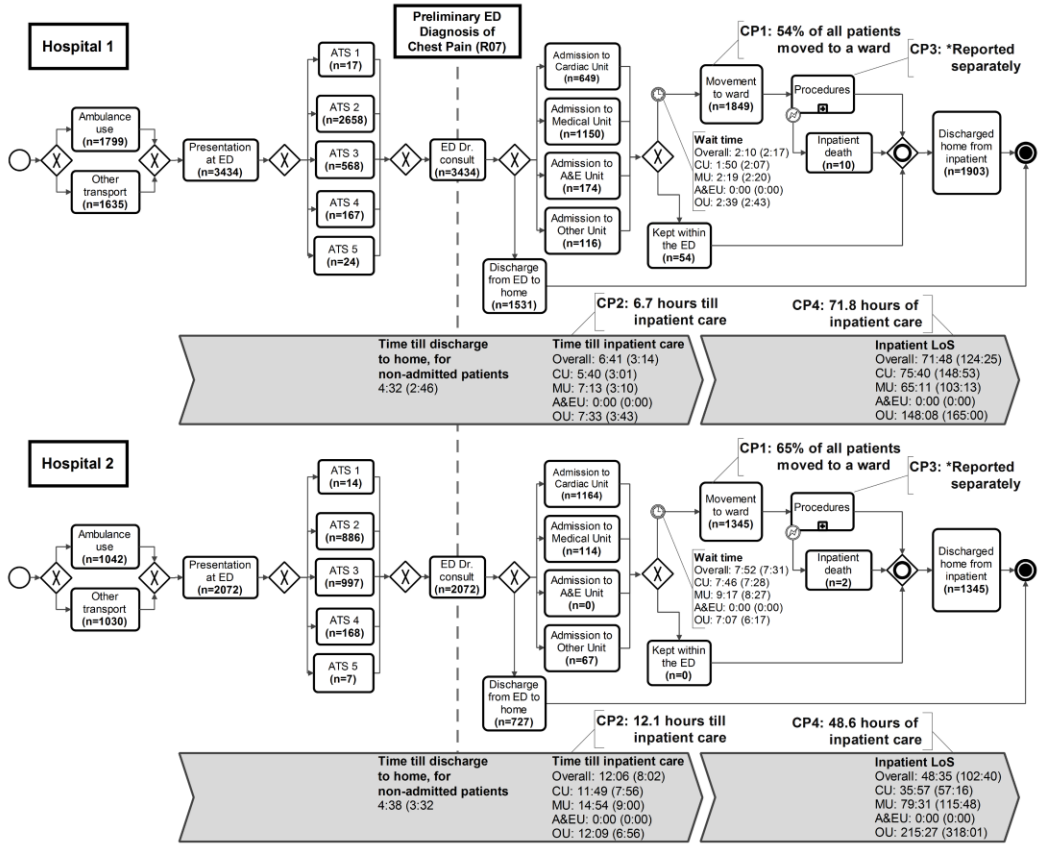
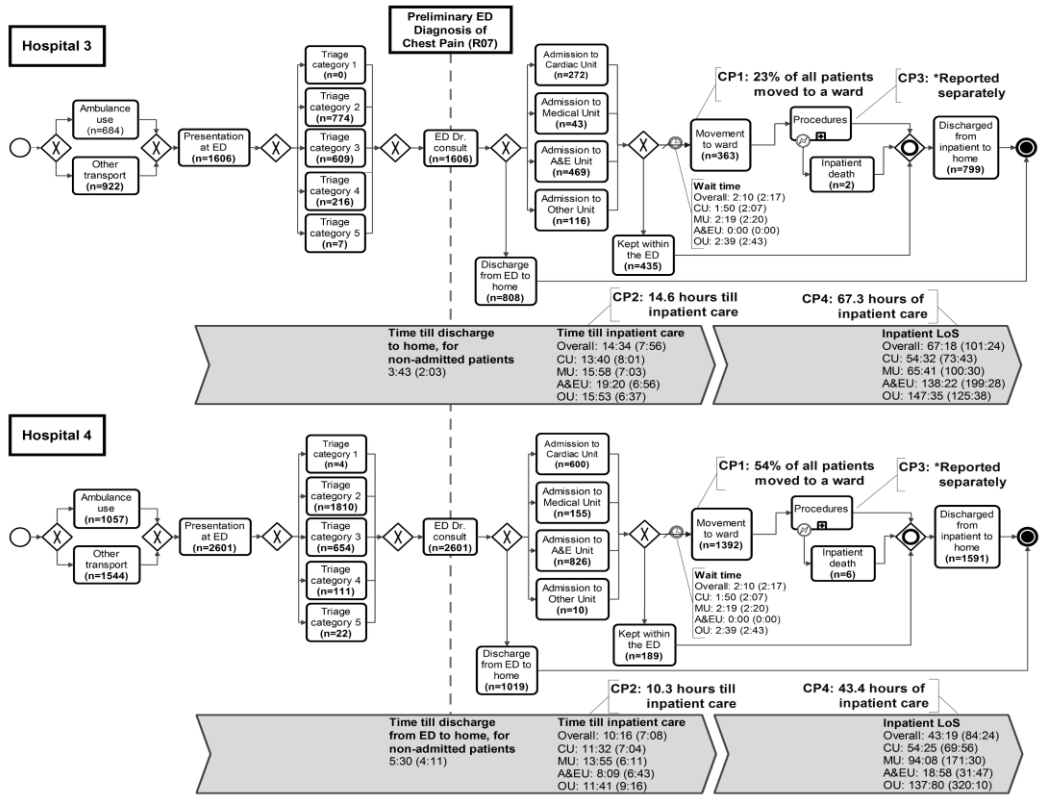


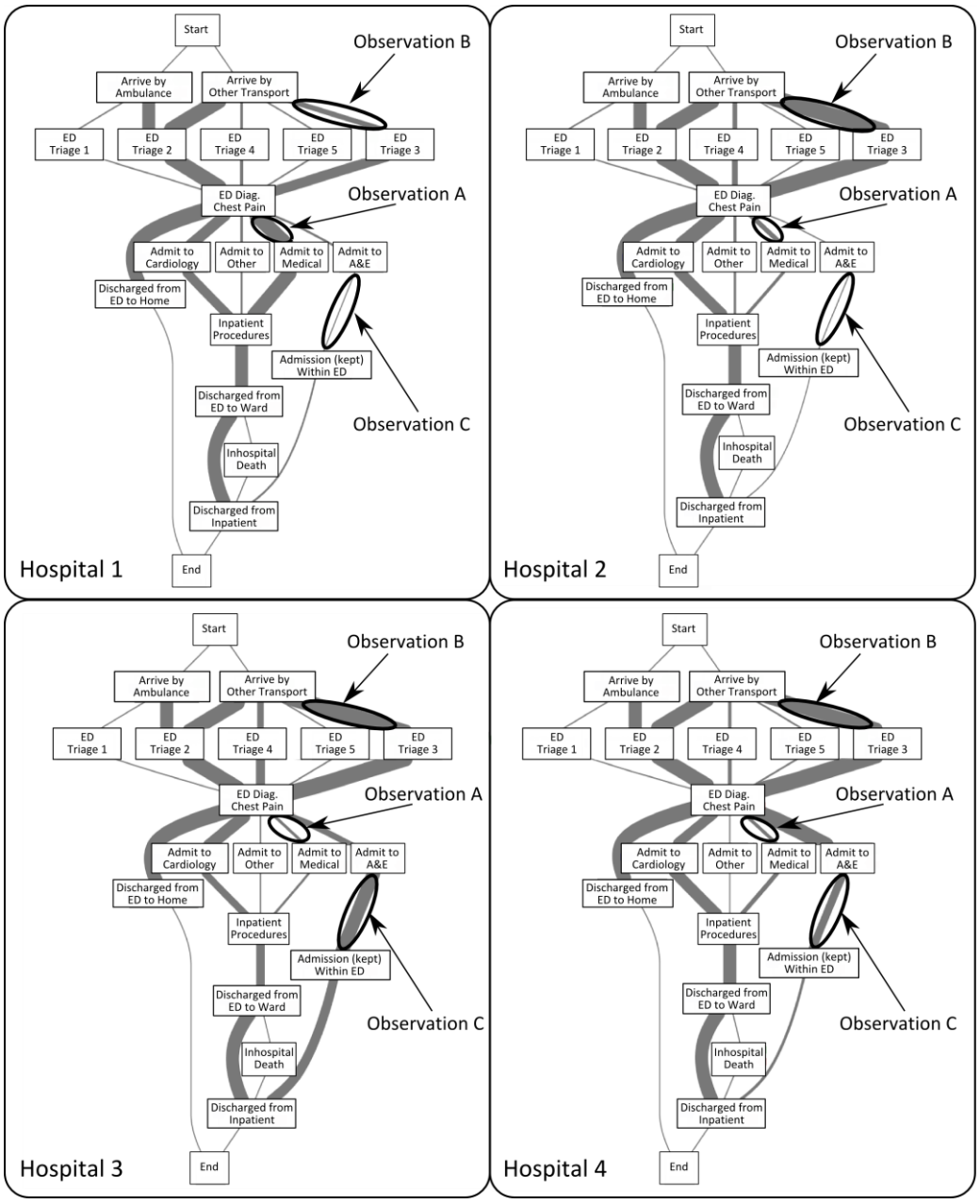
Fig. 2.4: BPMN models illustrating the Flow & Timing for Hospitals 3 & 4



2.6.2 Fuzzy Mined Models

In addition to the results reported as BPMN models, we also trialed other approaches to visualize and analyse the comparative results. In the first instance, Partington et al. created one common process model that captured the patient pathways within all four hospitals by applying the Fuzzy Miner Plug-in [Gunther 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 2.5) from which we could identify and communicate the differences between hospitals.

Fig. 2.5: Fuzzy models depicting the key variations (Observations A-C) in process



2.6.3 Conformance Analysis

Suriadi et al. attempted to extend the idea of the Fuzzy comparative models and sought to measure the 'conformance' or 'fit' of the data from each hospital's event-log, to each the mined process models. As shown in Table 2.2, Suriadi et al. were able to demonstrate that the models of Hospitals 2 and Hospitals 3, and their associated logs, were comparable.

Table 2.2: Fitness values – replay of data logs on hospital models

	Process Model (Petri Net)			
Hospital log	H1	H2	H3	H4
Hospital 1 (H1)	0.918	0.756	0.745	0.749
Hospital 2 (H2)	0.651	0.861	0.836	0.748
Hospital 3 (H3)	0.586	0.784	0.847	0.726
Hospital 4 (H4)	0.611	0.725	0.77	0.871

2.7 Practical insights and information feedback

The analysis of processes of care was focused around elements of the process that were identified by clinical experts as key factors. These *Comparison Points* provided an initial context for the analysis and interpretation of preceding and subsequent events:

Comparison Point 1. (CP1)
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. (CP2)
The throughput timing between ED presentation and movement to an in- patient setting (Admission)
— are there associated differences in initial risk (triage) 14ategorization? —does throughput differ depending on the clinical unit to which patients are admitted to?

Comparison Point 3. (CP3)
The frequency of procedures (diagnostic/treatment) provided
— does the use of procedures differ depending on the clinical unit to which patients are admitted?

Comparison Point 4. (CP4)
The total length of stay for patients
— does the length of stay differ depending on the clinical unit to which patients are admitted to?

Significant differences were observed in comparison point (CP) 1, where the proportion of patients admitted and transferred through to a ward ranged from 23-65%, with similar patients (all given preliminary chest pain diagnoses) being admitted to alternative types of clinical unit. As highlighted as 'Observation A' in Figure 2.5, Hospital 1 admitted a higher proportion of patients to the Medical Unit in comparison to other hospitals. Unexpectedly, 'Observation C' in Figure 2.5 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 *CP2*, reported in Figures 2.3 and 2.4, 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.

The observed differences in ED throughput time might be due, in part, to variation in the urgency categorization (triage) at different hospital sites. For example, as shown in Figure 2.5 as 'Observation B', the proportion of patients being assigned to Triage Category 3 is *lower* in Hospital 1 than in the 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 is quicker (and hence seemingly preferred), this hospital was seen to discharge a majority (65%) of patients to non-cardiac (i.e. General Medical) clinical units. One possible interpretation of this is that Hospital 1 was directing patients to any unit, simply when the first resources (i.e., a bed or a clinician) became available, irrespective of specialty. This is somewhat corroborated within the data, as there was little difference observed in the throughput and waiting times between the different clinical units at Hospital 1.

When looking at procedures for chest pain patients (not shown in the figures), the two largest hospitals, Hospitals 1 and 2, were found to provide an almost identical rate of angiography. Interestingly however, the two hospitals with the smaller patient volumes made the least use of angiography. In looking more closely at *CP3*, 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. Perhaps Hospital 1 was selecting patients with greater need for such procedures, for admission to a cardiac unit? Alternatively, the cardiac unit at Hospital 1 may simply have a lower threshold for undertaking these procedures.

Finally, Hospitals 1 and 3 had the longest inpatient LoS (mean of 70 hours), whilst 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 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.

2.7.1 Additional Insights

The work by Suriadi et al. provided additional insights into how we might approach alternative, 'natural groupings' of patients that may better represent observed variations in process. Clustering analyses were used to investigate similarities in patient (e.g., age, gender etc.) and process characteristics (e.g., the number of times a certain activity is undertaken, the number of events in a case) and obtain clusters of cases with respect to the processes within clinically relevant groupings. These analyses identified variation between groupings based on ED diagnoses and ED discharge destination, and (cluster analysis) groupings based on similarities with respect to the full process of care within the ED. These analyses suggest non-uniformity with respect to pathways of care for patients with similar presenting and exit characteristics.

2.8 Methodological lessons

2.8.1 Piecing together the 'jigsaw puzzle' of data

The successful representation of pathways for a population defined by their initial categorisation (i.e. ED diagnosis) provided insights into the entire 'at risk' population. However, the problem of piecing data from multiple sources is recognised within the Process Mining literature 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.

However, linking health care data can be 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 always straightforward to merge data via unique identifiers.

2.8.2 Timestamps

In order to discover process knowledge, the temporal ordering of events within a data log is important and hence the accuracy of timestamps associated with key events in the process is an important issue. Methods for checking the accuracy of timestamps included the use of the "Timestamp Issue Detector" plugin of software platform ProM 6, which detects duplicate timestamps, timestamps with different accuracy, and timestamps that may be outliers within a trace. We also visually inspected the data for other timestamp issues. There were many duplicate data points due to the linking of alternative data sources with common variables, many of which were subject to slight variations, possibly due to rounding, collection and data storage.

2.8.3 Comparability of what?

In summary, the workflow and performance analysis were found to be useful tools for generating an understanding of the steps involved at each hospital site, but these tools were less useful for comparing their relative practices. While the insights generated were useful to initiate conversations about practices, the idea that individual clinicians and hospital sites could still explain variation in process on the basis of 'assumed' variations in casemix

(presenting characteristics of patients at the different hospitals) could not be overcome using existing process mining methodologies.

To address this issue, the application of the suite of process mining techniques to sub-groups of patients with similar presenting characteristics at each hospital was considered. However, the interpretative capacity required to assess full process models for multiple hospitals and multiple patient sub-groups was considered too great. This led to a shift in methodological focus, away from qualitative, descriptive process models, to statistical models of process metrics, controlling for relevant patient and hospital level covariates (as described in Chapter 3).

Chapter 3 Comparative analysis of costs, health outcomes, and processes of care: a case study in patients presenting at hospital with chest pain

This chapter describes the analysis of routinely collected administrative and clinical hospital data to inform the potential value and conduct of quality improvement activities. The data are analysed in order to identify important areas of variation in the health care costs, health outcomes, and processes of care associated with the management of patients presenting with chest pain at the emergency departments (EDs) of four alternative hospitals in South Australia.

The aim of these analyses was to answer the following questions:

1. Is there sufficient variation, across the four study hospitals, in the health care costs and health outcomes associated with the management of patients presenting with chest pain, to warrant action to improve services at hospitals that are incurring greater costs or achieving poorer patient outcomes?
2. Can empirical analyses of the available data describing the processes of care at the four study hospitals further inform the decision to act to improve services, and the focus of any improvement activities (i.e. particular components of the care pathway, or patient sub-groups, for which variation in processes of care is greatest).

The following section describes the data and data sources, and the methods of analysis used to compare costs, outcomes, and processes of care. The results of the analyses are presented in detail, followed by a discussion of the findings, with specific reference to the above questions.

3.1 Methods

The Data: The four study hospitals in South Australia maintain a suite of local data warehouses containing patient-level information describing key procedures, pathology test results, movement between hospital departments and wards, etc. as well as automated linkages to population-based mortality data. These local systems have comparable nomenclature and collection practices, and are collated by the State health department in the form of a single, State-wide reporting repository.

Separate administrative data, submitted to the State health department for every inpatient separation at all public and private South Australian hospitals, were available from 2003 to June 2011. These data provided key variables such as age, gender, and postcode of normal residence (to inform Socio-Economic Indexes for Areas (SEIFA) scores), as well as co-morbidities (coded on the basis of principal and additional diagnoses in the 12 months preceding the index ED presentation [Duckett et al, 2008]) and hospital admissions for related conditions within the previous 12 months.

Probabilistic data linkage methods using name, gender, and date of birth were used to group public hospital separations by patient. Private hospitals separations were assigned to these groups on the basis of matching Medicare numbers.

The required administrative data to July 2011 had been collated within SA Health by project commencement in January 2012. The project timelines allowed six months for the linkage of the administrative and clinical data (thus allowing six months for data analysis, and one year for the feedback and quality improvement component of the project). Unfortunately, due to budgetary and other internal pressures within the South Australian Department of Health, there was a one year delay in gaining access to the final dataset.

Data analysis: The cost of the index hospital episode was estimated for every eligible patient, representing both the ED and inpatient component (for admitted patients). Detailed patient-level costs for all inpatient separations were built up across 17 cost categories, one of which represented costs incurred in the ED. ED cost data were analysed using a generalised linear regression model to predict ED costs as a function of age, gender, ED diagnosis, and time in ED.

Outcomes were specified as a related readmission (for unstable angina, MI, or stroke) or mortality within 30 days, or within 12 months. Available process variables were mapped to the Australian clinical guidelines for the management of acute coronary syndrome [ACSGWG, 2006], and discussed with clinical experts to identify those variables of most relevance to a comparative analysis of key process indicators for patients presenting with chest pain. The selected process indicators included: the proportion of presenting patients admitted to hospital, the time to admission (i.e. length of stay (LoS) in the ED), the proportion of patients undergoing an invasive diagnostic procedure who went on to receive an invasive management procedure, and the inpatient LoS for admitted patients.

To identify variation between the hospitals, separate multiple regression models were fitted to the data for each of the cost, outcome, and process variables. Binary hospital attendance variables were used to test hospital effects, and hospital interaction terms were tested to identify patient sub-groups that might be driving variation observed at the aggregate hospital level. The models tested the following patient-level variables to control for confounding: age, gender, troponin test result (positive/negative), SEIFA score, and a wide range of binary co-morbidity and recent hospital admission variables. Interactions between key patient-level covariates were also tested.

3.2 Results

The analysis included 7,950 eligible patients, ranging from 1,527 at Hospital 2 to 2,368 at Hospital 3. Table 3.1 describes the key characteristics of the patients presenting at the comparator hospitals. There were statistically significant differences in some key baseline characteristics, including age, socioeconomic status, objective risk markers (troponin test results), and existing circulatory conditions and diabetes.

Table 3.1 Patient characteristics

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Difference (p value)
No. patients	1997	1527	2368	2058	
Age (SD)	60.2 (17.5)	62.5 (16.8)	59.1 (18.2)	57.9 (16.2)	<0.001
Male	0.54	0.53	0.53	0.54	0.768
SEIFA decile (SD)	5.4 (2.6)	4.0 (2.4)	6.0 (2.6)	2.6 (2.0)	<0.001
Positive troponin test result	0.17	0.13	0.12	0.11	<0.001
Existing circulatory disorder	0.37	0.33	0.37	0.33	0.002
Cancer	0.03	0.03	0.03	0.02	0.189
COPD	0.03	0.03	0.03	0.03	0.706
Renal disease	0.07	0.08	0.09	0.07	0.398
Diabetes	0.07	0.06	0.09	0.10	<0.001
Dementia/Alzheimer's disease	0.01	0.02	0.01	0.01	0.716
After hours presentation	0.61	0.62	0.59	0.58	0.082
Weekend presentation	0.24	0.25	0.23	0.24	0.636

SEIFA – SocioEconomic Indicator For Areas: 1=lowest decile

In adjusting for difference in baseline characteristics, all of the fitted regression models for the cost, outcomes, and process dependent variables passed the a priori specified tests for goodness-of-fit and model specification. The following sections describe the outputs of the cost, outcome, and process models, respectively.

3.2.1 Index presentation costs

Table 3.2 presents the analysed differences in inpatient costs associated with the index chest pain presentations for the eligible cohort. The cost per presenting patient is reported, based on the proportions of patients who were admitted at each hospital in the reported patient groups. Across all patients, hospital 4 reported the lowest standardised costs per presenting patient, though hospital 1 reported very similar costs. Hospitals 2 and 3 had statistically significantly increased costs, with a range of between $300 and $1,000 per presenting patient. Over the almost 4,000 patients presenting to these hospitals, the mean excess costs sum to over $2million every year.

Increased costs of care are observed at all hospitals in the sub-group of patients with prior experience of a circulatory condition on presentation to the ED, which is primarily due to the increase probability of these patients being admitted to inpatient care (Table 4). However, at hospital 2, the cost differential to hospital 4 appears to be driven by costs in this patient group. A similar pattern is observed at hospital 1, though the cost differences do not reach statistical significance relative to hospital 4.

Table 3.2 Index inpatient costs, per presenting patient

Outcome/ Sub-group	No. of patients	Hospital 4	Hospital 1 - Hospital 4	Hospital 2 - Hospital 4	Hospital 3 - Hospital 4
All patients	7950	$2,868 ($2729 to $3006)	$42 (-$282 to $395)	$630 ($307 to $989)	$510 ($299 to $713)
Existing circulatory condition, Out-of-hours presentation	1706	$4,539 ($4264 to $4840)	$229 (-$431 to $989)	$1,542 ($786 to $2365)	$510 ($299 to $713)
No existing circulatory condition, Out-of-hours presentation	1082	$1,579 ($1449 to $1722)	-$451 (-$661 to -$251)	$136 (-$73 to $361)	$510 ($299 to $713)
Existing circulatory condition, In-hours presentation	3050	$4,640 ($4366 to $4928)	$643 (-$58 to $1367)	$1,265 ($513 to $2111)	$510 ($299 to $713)
No existing circulatory condition, In-hours presentation	2112	$1,610 ($1492 to $1748)	-$6 (-$410 to $423)	-$91 (-$318 to $143)	$510 ($299 to $713)

3.2.2 Outcomes

Outcomes were analysed at 30 days and 12 months, with respect to hospital admissions for cardiovascular events, or mortality. Table 3.3 describes the outcome probability for the hospital with the best outcomes (hospital 1), and the risk of an event at the other hospitals, relative to Hospital 1. At 30 days, each of hospitals 2, 3, and 4 reported significantly higher event rates across the aggregate patient group. The patient age / hospital interaction term was statistically significant, indicating that relative and absolute differences in outcomes were greater in younger patients.

Differences across the hospitals were reduced with respect to outcomes at 12 months. Only hospital 2 maintains a statistically significant difference in outcomes across the aggregate patient group and all of the sub-groups, defined with respect to age and gender. The only other significant difference in 12-month outcomes is in the old male patients sub-group, where patients at hospital 3 have a 50% increased risk of an event at 12 months.

Figure 3.1 combines the costs and 12 month outcomes model outputs in the form of a cost-effectiveness acceptability curve across the aggregate patient cohort, which reports the probability that each hospital is the benchmark performer according to the value attached to improved outcomes. The figure shows that if the avoidance of a cardiovascular inpatient admission or death at 12 months is valued at an equivalent monetary value of $10,000, there is a 96% probability that hospital 4 is the benchmark performer. As the value of achieving better outcomes increases, there is less certainty regarding the benchmark performer: at values above $63,000, hospital 1 becomes the mostly likely benchmark performer; at all values, it is clear that hospital 2 is under performing.

Table 3 Thirty day and twelve month outcomes

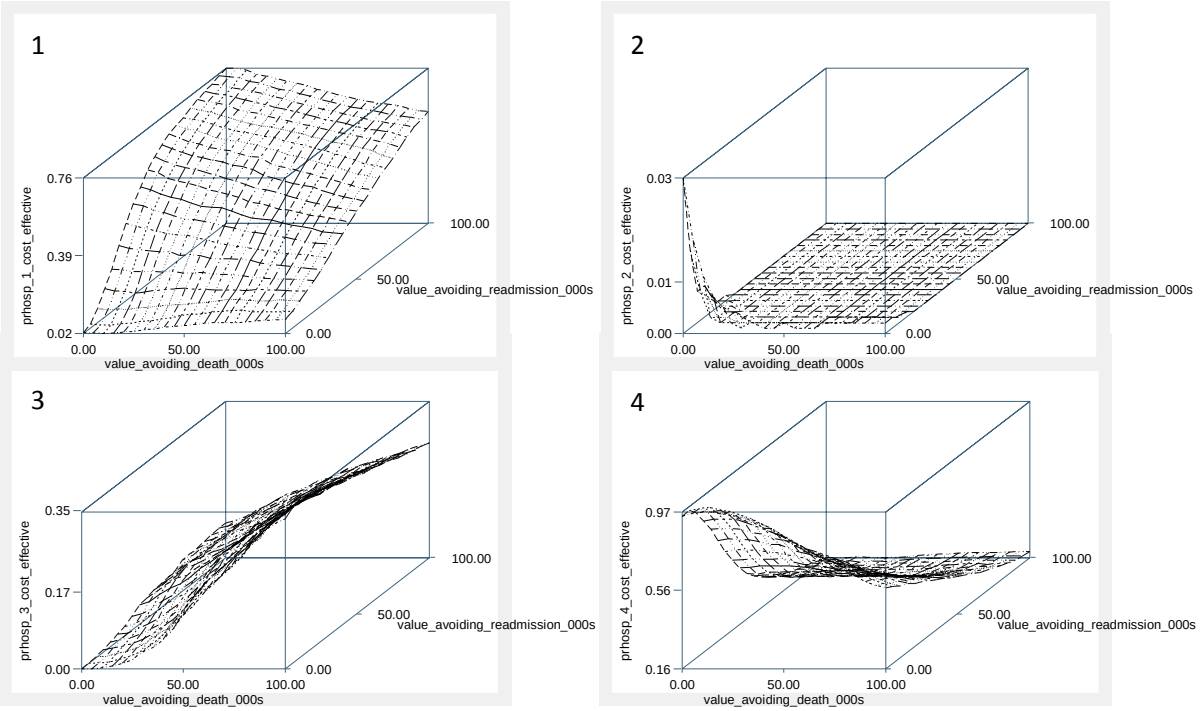
	No. of patients	Hospital 1	Hospital 2	Hospital 3	Hospital 4
		Pr(event)	Pr(event) [RR (95%CI)]	Pr(event) [RR (95%CI)]	Pr(event) [RR (95%CI)]
30 day readmission or death					
All patients	7950	0.009 (0.005 - 0.014)	0.021 [2.39 (1.36 - 4.11)]	0.016 [1.83 (1.10 - 3.07)]	0.017 [1.96 (1.11 - 3.32)]
Young patients*	3982	0.005 (0.002 - 0.009)	0.015 [3.25 (1.52 - 6.80)]	0.011 [2.48 (1.18 - 5.20)]	0.012 [2.66 (1.23 - 5.49)]
Old patients*	3968	0.021 (0.015 - 0.027)	0.031 [1.55 (1.05 - 2.22)]	0.024 [1.19 (0.82 - 1.69)]	0.026 [1.28 (0.83 - 1.85)]
12m readmission or death					
All patients	7950	0.045 (0.037 - 0.055)	0.072 [1.59 (1.27 - 1.93)]	0.048 [1.06 (0.80 - 1.35)]	0.052 [1.16 (0.91 - 1.47)]
Young male patients	2281	0.043 (0.033 - 0.055)	0.069 [1.59 (1.27 - 1.94)]	0.047 [1.09 (0.75 - 1.51)]	0.05 [1.16 (0.91 - 1.47)]
Old male patients	1979	0.102 (0.082 - 0.124)	0.155 [1.54 (1.25 - 1.85)]	0.152 [1.50 (1.18 - 1.93)]	0.116 [1.14 (0.91 - 1.43)]
Young female patients	1701	0.025 (0.017 - 0.033)	0.04 [1.61 (1.28 - 1.97)]	0.019 [0.76 (0.47 - 1.20)]	0.028 [1.16 (0.90 - 1.48)]
Old female patients	1989	0.086 (0.068 - 0.105)	0.132 [1.55 (1.26 - 1.87)]	0.092 [1.09 (0.81 - 1.42)]	0.098 [1.15 (0.91 - 1.44)]
12m mortality					
All patients	7950	0.02 (0.01 to 0.03)	0.024 [1.42 (0.93 to 2.05)]	0.014 [0.82 (0.47 to 1.23)]	0.021 [1.06 (0.67 to 1.45)]
12m readmission					
All patients	7950	0.03 (0.02 to 0.04)	0.048 [1.72 (1.27 to 2.27)]	0.034 [1.22 (0.84 to 1.66)]	0.039 [1.3 (0.89 to 1.71)]

Pr(event) – probability of death or hospital admission for unstable angina, myocardial infarction, or stroke; RR – relative risk (base hospital – hospital 1); statistically significant results highlighted in bold

* young patients defined as patients aged under the mean age, old patients defined as those aged over the mean age

Figure 3.1 combines the analyses of costs and outcomes in the form of cost-effectiveness acceptability planes, which present the probability that each hospital is the most cost-effective (or the benchmark) hospital, for alternative equivalent monetary values associated with the avoidance of death at 12 months, and cardiovascular hospital admission within 12 months (as represented on the horizontal axes). As an example, if we assign an equivalent monetary value of $100,000 to avoiding a death at 12 months, and a value of $50,000 to avoiding a hospital admission, hospitals A to D have probabilities of being the most cost-effective hospital of 30%, 0%, 32%, and 38%, respectively. It is clear that the choice of the benchmark hospital varies significantly between hospitals A, C, and D according to the values associated with the avoidance of mortality and hospital admissions. However, it is apparent that there is a very small likelihood that Hospital B is the benchmark hospital, which is consistent with the statistically significantly increased costs and poorer outcomes at hospital 2 (as shown in Tables 3.2 and 3.3).

Figure 3.1 Cost-effectiveness acceptability planes: probabilities of cost-effectiveness*



The probability that each hospital is the most cost-effective hospital, for alternative equivalent monetary values associated with the avoidance of death at 12 months, and cardiovascular hospital admission within 12 months: A – Hospital 1; B – Hospital 2; C – Hospital 1; D – Hospital 4.

3.2.3 Process indicators

Four separate process indicators were analysed to identify differences across hospitals, controlling for patient presenting characteristics. Table 3.4 describes the probabilities of presenting patients being admitted as an inpatient. Hospital 1 reported the highest mean probability of admission, which was significantly higher than admission rates at the other hospitals. However, the largest differences were observed in the sub-groups of patients who did not have an existing circulatory condition at the time of presentation. In the largest sub-group – patients with a negative troponin test and no existing circulatory condition – patients at hospital 2 were 30% less likely to be admitted (an absolute difference of 18%).

Table 3.4 also describes the probability that patients who underwent an invasive diagnostic procedure went on to receive an invasive therapeutic procedure (a percutaneous coronary intervention – PCI). Hospital 1 has the lowest conversion rate, of 14% across all patients receiving an angiogram, increasing to 30% in patients with a positive troponin test who present at the ED on a weekend. Hospital 2 has the largest conversion rate, with a relative risk of conversion above 2 across all patient groups. Hospitals 3 and 4 also report higher PCI proportions, which appear to be mainly driven by higher conversion rates in patients with a negative troponin presenting during the week.

The third process variable describes the probability of undergoing a PCI, which shows that Hospitals 2 and 4 have lowest uptake. In both cases this finding is driven by significantly lower rates in low risk (troponin negative) patients who presented during the week.

Table 3.4 Processes: Pr(admitted), Pr(PCI, given angiography), Pr(PCI)

Pr(Admitted)	No. pts	Hospital 1	Hospital 2	Hospital 3	Hospital 4
		Pr(admitted) (95%CI)	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]
All patients	7950	0.77 (0.75 - 0.79)	0.67 [0.87 (0.83 - 0.91)]	0.71 [0.92 (0.89 - 0.96)]	0.71 [0.92 (0.88 - 0.96)]
Troponin +ive, Existing circ.	842	0.97 (0.94 - 0.99)	0.96 [0.99 (0.98 - 1.01)]	0.94 [0.97 (0.94 - 0.99)]	0.94 [0.97 (0.94 - 0.99)]
Troponin +ive, No Existing circ.	216	0.87 (0.77 - 0.94)	0.77 [0.88 (0.79 - 0.95)]	0.78 [0.89 (0.78 - 0.97)]	0.78 [0.89 (0.78 - 0.97)]
Troponin -ive, Existing circ.	1946	0.89 (0.86 - 0.91)	0.87 [0.98 (0.94 - 1.03)]	0.86 [0.97 (0.95 - 0.99)]	0.86 [0.97 (0.94 - 0.99)]
Troponin -ive, No Existing circ.	4946	0.62 (0.58 - 0.65)	0.44 [0.71 (0.65 - 0.77)]	0.55 [0.90 (0.84 - 0.96)]	0.55 [0.90 (0.83 - 0.96)]
Pr(PCI | Angiography)		Pr(PCI|angio) (95%CI)	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]
All patients	7950	0.14 (0.10 - 0.19)	0.34 [2.40 (1.71 - 3.36)]	0.25 [1.79 (1.31 - 2.44)]	0.22 [1.57 (1.14 - 2.15)]
Troponin +ive, Weekend	279	0.30 (0.18 - 0.42)	0.62 [2.18 (1.36 - 3.44)]	0.30 [1.07 (0.62 - 1.87)]	0.27 [0.94 (0.54 - 1.53)]
Troponin +ive, Weekday	779	0.22 (0.15 - 0.30)	0.45 [2.03 (1.44 - 2.79)]	0.27 [1.22 (0.83 - 1.74)]	0.38 [1.72 (1.26 - 2.28)]
Troponin -ive, Weekend	1632	0.17 (0.10 - 0.26)	0.44 [2.76 (1.48 - 4.74)]	0.28 [1.72 (0.96 - 2.93)]	0.15 [0.93 (0.49 - 1.64)]
Troponin -ive, Weekday	5260	0.12 (0.08 - 0.17)	0.28 [2.36 (1.56 - 3.41)]	0.24 [2.02 (1.35 - 2.92)]	0.23 [1.90 (1.32 - 2.64)]
Pr(PCI)		Pr(PCI) (95%CI)	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]	Pr() [RR (95%CI)]
All patients	7950	0.9 (0.08 – 0.11)	0.07 [0.71 (0.57 - 0.88)]	0.1 [1.02 (0.82 - 1.26)]	0.06 [0.63 (0.50 - 0.78)]
Troponin +ive, Weekend	279	0.87 (0.76 – 0.94)	0.87 [1.00 (0.93 - 1.08)]	0.95 [1.10 (1.03 - 1.22)]	0.91 [1.05 (0.99 - 1.13)]
Troponin +ive, Weekday	779	0.90 (0.82 – 0.95)	0.85 [0.94 (0.85 - 0.99)]	0.94 [1.05 (1.01 - 1.11)]	0.89 [0.99 (0.94 - 1.03)]
Troponin -ive, Weekend	1632	0.04 (0.03 - 0.05)	0.04 [1.04 (0.70 - 1.50)]	0.07 [1.72 (1.09 - 2.50)]	0.03 [0.86 (0.57 - 1.23)]
Troponin -ive, Weekday	5260	0.05 (0.04 - 0.07)	0.03 [0.62 (0.47 - 0.80)]	0.05 [1.03 (0.74 - 1.41)]	0.03 [0.51 (0.39 - 0.66)]

Table 3.5 reports differences in two length of stay (LoS) variables. For patients admitted as an inpatient, the differences in the mean time to admission at hospital 3 was between 3.9 and 5.5 hours shorter than at the other hospitals. The findings did not vary greatly by patient sub-group, though the times to admission, and the absolute differences between hospitals were slightly longer in patients with negative troponin test results.

Greater differences were observed between patient sub-groups with respect to inpatient LoS. At all hospitals, patients with a positive troponin test had significantly longer LoS, but mean LoS for positive troponin patients at hospitals 2 and 3 was between 13.8 and 18 hours longer than at hospital 4. Hospital 1 had a greater LoS differential for troponin negative patients.

Table 5 Emergency Department, and Inpatient Length of Stay

Length of Stay in the ED (if admitted as inpatient)	No. pts	Hospital 3, LoS in hrs	Hospital 1 - Hospital 3	Hospital 2 - Hospital 3	Hospital 4 - Hospital 3
All patients	5409	6.1 (5.94 - 6.30)	5.5 (5.10 - 5.97)	4.4 (3.94 - 4.96)	3.9 (3.42 - 4.37)
Troponin +ive, Out-of-hours presentation	578	5.7 (4.92 - 6.42)	3.4 (2.48 - 4.29)	4.6 (3.24 - 5.95)	2.1 (1.33 - 2.99)
Troponin +ive, In-hours presentation	333	4.9 (4.18 - 5.69)	4.3 (3.29 - 5.34)	4.6 (3.24 - 5.95)	2.1 (1.33 - 2.99)
Troponin -ive, Out-of-hours presentation	2705	6.6 (6.33 - 6.80)	5.5 (4.97 - 6.08)	4.4 (3.87 - 4.97)	4.3 (3.74 - 4.74)
Troponin -ive, In-hours presentation	1793	5.8 (5.53 - 6.12)	6.5 (5.67 - 7.26)	4.4 (3.87 - 4.97)	4.3 (3.74 - 4.74)
Length of Stay as inpatient		Hospital 4, LoS in hrs	Hospital 1 - Hospital 4	Hospital 2 - Hospital 4	Hospital 3 - Hospital 4
All patients	5373	37.2 (35.09 - 39.31)	7.4 (4.46 - 10.32)	4.5 (0.79 - 8.38)	10.5 (7.58 - 13.42)
Troponin +ive, Out-of-hours presentation	561	56.4 (48.10 - 65.35)	3.3 (-7.28 - 14.39)	18.0 (2.52 - 34.19)	16.3 (5.84 - 27.24)
Troponin +ive, In-hours presentation	325	51.1 (41.44 - 61.02)	7.0 (-3.44 - 17.45)	13.8 (-1.55 - 29.42)	15.2 (4.33 - 26.11)
Troponin -ive, Out-of-hours presentation	2702	33.8 (31.43 - 36.24)	6.5 (3.29 - 9.52)	3.8 (-0.13 - 7.87)	9.9 (6.50 - 13.03)
Troponin -ive, In-hours presentation	1785	35.3 (32.76 - 37.99)	10.2 (5.88 - 14.53)	-0.4 (-5.02 - 4.19)	8.7 (4.76 - 12.46)

3.3 Discussion

This study has demonstrated the feasibility and significance of comparative analyses of cost, outcomes, and processes of care for similar patient cohorts presenting at alternative hospitals, as well as identifying patient sub-groups for whom differences between hospitals are exaggerated. At an aggregate level, statistically significant, casemix-adjusted differences were observed in mean inpatient costs (up to $672 per admitted patient), and 30 day and 12 month cardiovascular or mortality event rates (odds ratios up to 2.42 and 1.64, respectively) across providers. The analysis of costs and patient outcomes did not identify a single benchmark hospital, but rather identified an apparent outlier (non-benchmark) hospital that was incurring higher costs and achieving poorer patient outcomes than other hospitals.

The interpretation of the analyses of processes of care might focus on Hospital 2, as the non-benchmark hospital, and which processes vary most significantly from the processes of care

observed at the other hospitals. Hospital 2 has the lowest inpatient admission rate, but this difference is driven by a much lower rate of admission for patients with a negative troponin test and no existing circulatory condition on presentation (Table 3.4).

Hospital 2 has the highest conversion rate of invasive diagnostic procedures into invasive management procedures, which indicates high specificity with respect to the selection of patients for an angiogram. However, it may also indicate low sensitivity, i.e. patients who should have received invasive management are not receiving the prerequisite invasive diagnostic procedure. The finding of lower use of PCI at Hospital 2, primarily in the patients with a negative troponin test is in line with the identified lower admission rates at Hospital 2.

These results suggest that the clinical pathway for patients presenting with chest pain, with a negative troponin test, might be reviewed at Hospital 2.

Hospital 2 is not an outlier with respect to aggregate time to admission or inpatient length of stay, but inpatient length of stay is much higher in higher risk patients (those with a positive troponin test result, as well as Hospital 3). This may be driving the increased in costs at Hospital 2 and 3, and so may be factor for further investigation at both of these hospitals.

3.3.1 Next steps

This study has built on previous studies with respect to the linkage of administrative and clinical data to improve casemix adjustment through the inclusion of objective pathology test results [Pham et al, 2012; Karnon et al, 2013]. In addition, the analysis of process of care variables facilitated the identification of specific patient sub-groups and aspects of the care pathway as potential targets for quality improvement. There are, however, areas in which the application could be improved.

A major limitation of the study, with respect to informing actions to improve services, was the delay experienced in obtaining access to the administrative hospital data. There was a learning curve with respect to the extraction of the clinical and population-based mortality data, but competing priorities led to significant delays in accessing and linking the administrative hospital data. This meant that on analysing the data in late 2013, the results related to patients presenting in 2009/10. The gap reduced the relevance of the analyses to contemporary service provision, and so has not informed actual actions to improve services. Delays in access to administrative data are likely to remain significant, but clinical and population-based mortality data can now be extracted and linked within a month of the data being collected. Ongoing analyses are assessing the added value of the administrative data, to determine whether analyses based on the clinical data alone are sufficiently robust to prioritise and inform quality improvement activities.

With respect to the scope of the data, casemix adjustment could be improved. In the chest pain study, other diagnostic indicators such as electrocardiogram results, and more detail on the pathway of patients to the hospital (e.g. via ambulance records) would be useful. Likewise, a wider range of process of care indicators might provide more directed guidance with respect to areas for improvement, for example, data describing the type and timing of medication use, bedside visits, and the use of allied health would provide a fuller picture of

the care pathways at the comparator hospitals. Increasing use of electronic patient records should facilitate more detailed casemix adjustment and process analyses in the near future.

Another area in which data access could be improved is with respect to staffing levels (i.e. number, type, and grade of clinicians in the ED, and in the wards to which patients are admitted), and competing demand for resources (i.e. ED presentation rates, and inpatient operating capacity). Such data would indicate whether differences in costs, outcomes, and processes may be partly explained by varying resource constraints. However, the presented analyses are not intended to provide definitive guidance, but rather to act as a screening tool for the identification of service areas with significant potential for quality improvement. Possible next steps include:

- Additional collection of primary data to further inform the comparative analysis of cost, outcomes, and processes,
- Unilateral quality improvement processes at non-benchmark hospitals (e.g. lean redesign methods [Ben-Tovim et al, 2007]),
- Cross-hospital quality improvement processes, providing a basis for quantitative and qualitative comparative data to inform improvement at multiple sites.

3.5 Conclusions

The analysis of individual-level hospital and mortality data to represent comparative hospital performance provides a link between decisions to fund new technologies and to fund service improvement activities. As in the study reported in this paper, such analyses may highlight greater potential for increasing population health at low incremental cost from improving use of existing technologies. The translation of this research methodology to practice will require improvements in the timeliness of access to linked data, and demonstration of the applicability of the analysis to a larger set of hospitals. To motivate quality improvement, the feedback of such comparative performance data needs to be combined with practical support with respect to change management.

The presented analyses identify clinically and statistically significant differences in casemix adjusted costs and outcomes across alternative providers, for patients presenting at an ED with chest pain. These results indicate the potential value of engaging stakeholders to identify and implement service improvements at one or more of the non-benchmark hospitals. The analyses of variation in processes provide complementary evidence to support the validity of the reported differences in costs and outcomes. The process data may also usefully inform subsequent stakeholder engagement through the identification of patient groups and process components for which variation across providers is greatest, and at which improvement efforts might be focussed.

More generally, such comparative analyses of costs, outcomes, and process across healthcare providers can inform the potential value of improving existing services (e.g. relative to allocating scarce resource to new technologies), and prioritise and guide efforts to service improvement.

Chapter 4 A systematic literature review of the effectiveness of alternative approaches to the feedback of comparative hospital performance data

A systematic review of the literature was undertaken to identify effective approaches to the feedback of comparative hospital performance data, and supplementary activities (e.g. support and monitoring of quality improvement). The review focuses on the clinical area of acute coronary syndromes (ACS), and any form of feedback of comparative hospital performance data. As a systematic review, rather than a rapid review, detailed analyses of study quality, and the evaluated feedback processes and supplementary activities are reported.

4.1 Literature review methods

Data sources and searches

To identify potentially relevant publications a systematic search of PubMed was conducted from 1 January 2000 to 28 April 2014. The search terms are listed in Appendix 1. The search separated reviews (systematic or otherwise) from primary source articles. Language restrictions were not applied.

Study selection

The following criteria were used to select studies for inclusion in the review:

Population: Patients receiving emergency or inpatients care for Acute Coronary Syndrome or related diagnoses (e.g. acute myocardial infarction, STEMI, NSTEMI, or unstable angina).

Intervention: The provision of feedback on activity or performance as a basis for improving either adherence to evidence-based care or patient outcomes. The feedback provided to each hospital must include a comparator (e.g. with peer hospitals, or state or national averages). The intervention must be provided across more than one hospital (multiple sites).

Comparator: A comparator is required, but was not specified. Possible comparators included: a comparison made between pre and post intervention periods, or a concurrent comparison across multiple sites.

Outcomes: Studies must assess the impact of the intervention on the change in either process outcomes (adherence to evidence-based care guidelines), clinical outcomes (patient health), or financial outcomes.

Study type: Unspecified.

Following the PubMed search, the titles and abstracts of 3146 publications were screened to determine eligibility for inclusion in the review. The full texts of 108 potentially eligible studies were retrieved and screened further. Twenty-five studies were considered to be eligible for inclusion.

Data extraction and analysis

Data was extracted from the 25 included studies using a standardised data extraction table. Study design was classified using the National Health and Medical Council's Evidence Hierarchy (www.nhmrc.gov.au/_files_nhmrc/file/guidelines/stage_2_consultation_levels_and_grades.pdf). During extraction it became obvious that two types of studies had been identified – those with an adequate control group (n=11), and those with no control group where only a comparison of the pre and post intervention periods had been made (n=14).

The quality of included studies with a control group (n=11) was assessed using a modified version of Downs and Black's *Checklist for Measuring Study Quality* [1998]. The checklist was adapted for use with a hospital level intervention, where potential quality issues may arise at both patient and hospital levels (see Appendix 2). The checklist examined study quality in reporting, external validity, internal validity (bias, confounding) and power. Given the diversity of outcomes reported it was not appropriate to statistically pool the findings of the included studies, therefore the results of the review are reported narratively only.

4.2 Literature review results

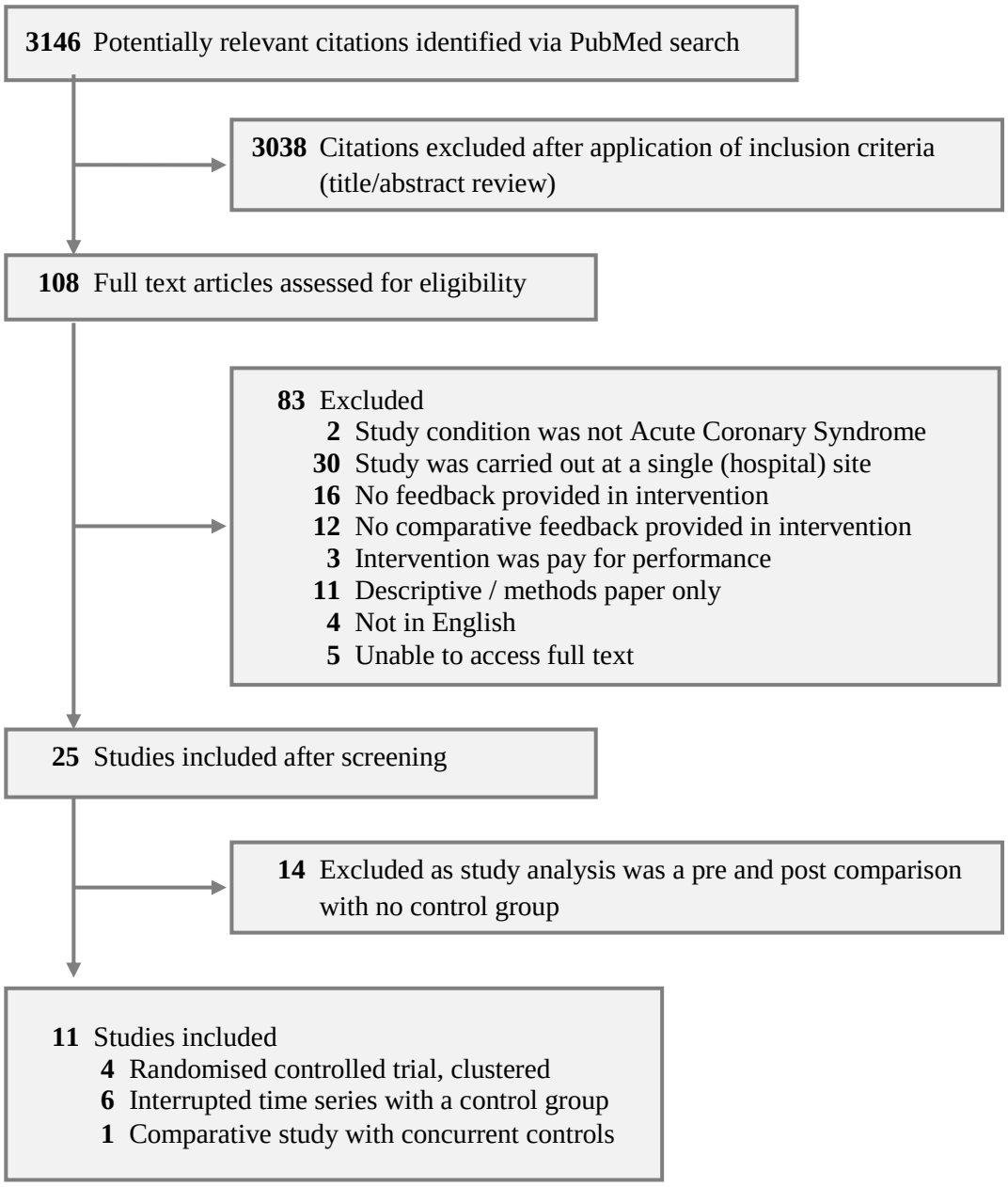
Figure 4.1. describes the numbers of studies identified and excluded at sequential stages of the review process. Twenty-five full study papers met the study inclusion criteria, but on review of the study designs, 14 of the 25 papers described study designs at NHMRC Level of Evidence III-3: Comparative studies without concurrent controls. Details of these 14 studies are available on request, but the current review focuses on studies with Level of Evidence III-2 and above, i.e. studies with concurrent controls.

Table 4.1 describes the quality scores of the 11 included studies, which shows the studies were within the moderate to high quality range.

Table 4.1 Quality criteria checklist scores

	Reporting	**External validity**	**Internal validity: bias**	**Internal validity: confounding**	**Power**	**Total score**
Section total	(/ 12)	(/ 4)	(/ 8)	(/ 8)	(/ 1)	(/ 33)
Renzi, 2012	8	4	4	3	0	19
Renzi, 2014	7	4	5	2	0	18
Carlhed, 2006	8	3	5	4	0	20
Carlhed, 2009	10	3	6	4	0	23
Carlhed, 2012	7	3	5	4	0	19
Sauaia, 2000	8	4	6	7	1	26
Beck, 2005	10	4	7	8	1	30
Berner, 2003	8	2	5	8	0	23
Heller, 2001	9	4	6	7	1	27
Hollenbeak, 2008	9	3	6	3	0	21
Moscucci, 2005	8	2	6	4	0	20

Figure 4.1 Flow diagram for included studies



Tables 4.2 to 4.4 provide details of the included studies with respect to the interventions evaluated, and the study designs.

Table 4.2 Included papers: Interventions and outcomes

Renzi, 2012	**Comparator**: Hospital performance data were released online, and communicated via patient and citizens associations and through press conferences (public reporting). Data were general, without inter-hospital comparisons or focus on specific indicators. **Intervention**: Hospital specific performance data with unblinded peer comparisons were fedback to hospitals via in-person meetings with general management, clinical directors and clinical staff. Staff were encouraged to suggest interpretations of the data. The aim was to foster a constructive approach to quality improvement. **Feedback data**: PCI within 48hrs **Study outcomes**: same as feedback
Renzi, 2014	**Comparator and Intervention**: as above **Feedback data**: *Procedures*: PCI within 48hrs; *Mortality:* at 30 days (overall, after PCI within 48hrs, in patients without PCI) **Study outcomes**: same as feedback
Carlhed, 2006	**Intervention**: a team of two cardiologists and two nurses were trained in QI methods over six months, with an additional twelve months of ongoing support. Training included how to use the national AMI registry (RIKS-HIA) to generate realtime reports enabling comparison of their hospital with current national performance and local performance over time. QI teams generated local action plans, which were tested, implemented and shared with other teams. Ongoing support for QI activities was provided by telephone, video conferencing and site visits. **Feedback data**: *Inhospital medications*: Heparin / LMWH; *Discharge medications*: ACE inhibitors, lipid-lowering therapies, clopidogrel; *Procedures*: coronary angiography **Study outcomes**: same as feedback
Carlhed, 2009	**Intervention**: as above **Feedback data**: As above **Study outcomes**: *Longer term outcomes* [a]: mortality (all cause), readmission (CV), mortality (all cause) / readmission (CV), bleeding complication (requiring hospitalisation)
Carlhed, 2012	**Intervention**: as above. Support for the intervention from the study management group was withdrawn after eighteen months. After a three month consolidation period, a second post-intervention evaluation was done to determine if changes in indicators were sustained. **Feedback data**: as above **Study outcomes**: *Sustainability indicators*: *Inhospital medications*: Heparin / LMWH; *Discharge medications*: ACE inhibitors, lipid-lowering therapies, clopidogrel; *Procedures*: coronary angiography; *Dissemination indicators*: stress testing, ECG, reperfusion for STEMI (given/time delay), length of stay

Sauaia, 2000	**Comparator**: Mailed feedback containing a summary of the research project and a comparison of hospital-specific, state and national performance data for the selected indicators. **Intervention**: Two hour on-site feedback presentation in additional to the written feedback materials. The presentations were delivered by locally respected cardiologists (or local experts in rural areas). They discussed the individual hospital data, followed by methods to translate the data into improvements in clinical practice. All hospitals were invited to submit a written QI plan to the research team. **Feedback data**: *Inhospital medications*: aspirin, avoidance of calcium channel blockers; *Discharge medications*: ACE inhibitors, aspirin, beta blockers; *Procedures*: reperfusion within 12hrs of arrival, thrombolytics within 1hr of arrival; *Advice*: smoking cessation **Study outcomes**: same as feedback
Beck, 2005	**Intervention**: Confidential, written report card presenting risk-adjusted performance data for twelve QI indicators for AMI. The report card included a cover letter, information sheet and the data in multiple forms (printed, overheads, electronic) to encourage further dissemination to hospital staff. A reminder to encourage dissemination was sent after two months. **Feedback data**: *Discharge medications* (30 days): beta blockers, ACE inhibitors, statins, and aspirin / clopidogrel; *Length of stay*; *Follow up*: physician visit within 4wks of discharge; *Waiting times*: angiography, PCI, CABG; *Mortality* (30 days); *Readmissions* (30 days): AMI, angina, CHF **Study outcomes**: same as feedback
Berner, 2003	**Health Care Quality Improvement Project (HCQIP) intervention**: QI coordinator (non-physician) attended a half day orientation. This included a review of UA guidelines, as well as materials and further discussion to support QI activities. Coordinators were given hospital-specific performance data for their hospital, with aggregated and de-identified hospital performance data for comparison. Data included benchmarks for quality of care for each indicator. **Opinion leader (OL) intervention**: A local physician opinion leader (OL) was identified at the hospital via survey of physicians. Both the hospital QI coordinator and the OL attended a half day orientation, as described above (HCQIP), and including greater detail on the research study design, evidence for the chosen indicators, and QI strategies that may be used to improve performance. Education materials for use with hospital staff were also provided. OLs were asked to develop and implement a QI plan in collaboration with their hospital's QI coordinator. **Feedback data**: *Inhospital medications*: Antiplatelets within 24hrs of admission, beta blockers, heparin (for moderate to high risk patients); *Discharge medication*: Antiplatelets; *Procedures*: ECG within 20mins of arrival **Study outcomes**: same as feedback

Heller, 2001	**Intervention**: Educational session delivered by a local opinion leader and research staff member. Sessions were delivered to clinical staff involved in the care of patients with acute chest pain (emergency or inpatient). Sessions described National Heart Foundation guidelines and fedback data from a baseline survey of hospital performance for the management of unstable angina. Performance data was hospital-specific with comparisons with other de-identified hospitals and guideline recommendations. An additional session was planned for other staff, but was not always implemented. **Feedback data**: *Inhospital medications*: aspirin, beta-blockers, calcuim channel blockers, heparin, IV nitroglycerine; *Procedures*: coronary angiography, ECG; *Follow up*: Rehabilitation **Study outcomes**: same as feedback
Hollenbeak, 2008	**Intensive public reporting**: three or more publically accessible reports on three or more health conditions. **Limited public reporting**: Performance data was used for internal purposes only, or had been included in two or less public reports, or in reports covering two or less conditions, or in reports that were not made publically available. **Feedback data**: not reported **Study outcomes**: in-hospital mortality
Moscucci, 2005	**Interventions**: Public reporting in New York State since 1991 via the Coronary Angioplasty Reporting System. No public reporting in Michigan during the study period. **Feedback data**: *Procedures*, *Complications*, *Inhospital outcomes*: details not reported **Study outcomes**: in-hospital mortality

ACE inhibitor: Angiotensin converting enzyme inhibitor. AMI: Acute myocardial infarction. ARB: Angiotensin II receptor blockers. CABG: Coronary artery bypass graft. CHF: Chronic heart failure. CV: Cardiovascular. ECG: Electrocardiogram. HCQIP: Health Care Quality Improvement Program. IV: Intravenous. LMWH: Low molecular weight heparin. MI: Myocardial infarction. NR: Not reported. NSTEMI: Non-ST-elevation myocardial infarction. OL: Local physician opinion leader. Other AMI: AMI patients not able to be classified as either STEMI or NSTEMI and given diagnosis related code (DRG code) of 410.9. PCI: Percutaneous coronary intervention. QI: Quality improvement. STEMI: ST-elevation myocardial infarction. UA: Unstable angina.

[a] Patients were followed up for between six and eighteen months.

Table 4.3 Included papers: Study details

Paper	Location	Time period	Level of Evidence	Intervention hospitals	Control hospitals	Total patients
Renzi, 2012	Italy	Pre-period: 2006 to 2007 (July) Intervention: end 2007 (meetings), Feb 2008 (public reporting) Post-period: 2008 to 2009 (Sept)	III-2	All Lazio	All Italy, excluding Tuscany	381,053
Renzi, 2014	Italy	Pre-period: Jan 2006 to Dec 2007 Intervention: Jan to Feb 2008 Post-period: Jan to Nov 2009	III-2	All Lazio	All Italy, excluding Tuscany	64,150
Carlhed, 2006	Sweden	Pre-period: Jul 2001 to Jun 2002 Intervention: Nov 2002 to Apr 2003 Post-period: May 2003 to Apr 2004	III-2	19	19	6726 + [a]
Carlhed, 2009	Sweden	Pre-period: Jul 2001 to Jun 2002 w/ long-term outcomes to Dec 2012 Intervention: Nov 2002 to Apr 2003 Post-period: May 2003 to Apr 2004 w/ long-term outcomes to Oct 2004	III-2	19	19	13,362
Carlhed, 2012	Sweden	Pre-period: Jul 2001 to Jun 2002 Intervention: Nov 2002 to Apr 2003 Post-period: May 2003 to Apr 2004 Intervention ceased: 1 May 2004 Post-period: Aug 2004 to Jan 2005	III-2	19	19	10,286 + [b]
Sauaia, 2000	Colorado, USA	Pre-period: Sept 1994 to May 1995 Intervention: Apr 1996 (written feedback), Jun to Aug 1996 (meetings) Post-period: Aug to Dec 1996	II	9	9	1,367
Beck, 2005	Quebec, Canada	Pre-period: 1 Apr 1999 to 31 Mar 2000 with follow up data until 31 Mar 2000 Intervention: May 2002 Post-period: 1 Oct 2002 to 31 Mar 2003 with follow up data until 31 Mar 2003 and 31 Mar 2004	II	38	38	17,077
Berner, 2003	Alabama, USA	Pre-period: 1997 to 1998 Intervention: 3mths prior to post-period Post-period: 1999 to 2000 (9mths)	II	HCQIP: 8 OL: 7	6	2,210

Paper	Location	Time period	Level of Evidence	Intervention hospitals	Control hospitals	Total patients
Heller, 2001	NSW, Australia	Pre-period: 1 Feb to 30 Jun 1996 Intervention: Mar to Jun 1998 Post-period: 1 Jul to 31 Dec 1998	II	17	19	3,240
Hollenbeak, 2008	Pennsylvania & 20 other states, USA	Pre-period: 1997 to 1999 Intervention: 1999 [ongoing annual reports] Post-period: 2000 to 2003	III-2	All Pennsylvanian hospitals	60	46,886 [c]
Moscucci, 2005	Michigan & New York, USA	Study period: 1 Jan 1998 to 31 Dec 1999 [Intervention: in New York since 1991]	III-2	All New York (34)	8	80,422

NHMRC Levels of Evidence: II – randomized controlled trial; III-2 - Interrupted time series with a control group (other than Moscucci - Comparative study with concurrent controls)

[a] Only patient numbers for the post-period were specified. [b] Only patient numbers for the first and second post-periods were reported, baseline patient numbers for dissemination indicators were not reported. [c] Equivalent to 23,443 propensity matched pairs.

Table 4.4 Included papers: Intervention and study analysis details

	Intervention and data feedback summary						Analysis of effect	
Paper	**Condition**	**Intervention**	**Feedback frequency**	**Data recipient**	**Data analyst**	**Data collection**	**Level of analysis**	**Risk adjust**
Renzi, 2012	AMI	Public reporting + Feedback (onsite)	Public reporting: NR; Feedback: 2 yearly	General public Management Clinical staff	External	Routine	Patient	Yes
Renzi, 2014	AMI: STEMI, NSTEMI, Other	Public reporting + Feedback (onsite)	Public reporting: NR; Feedback: 2 yearly	General public Management Clinical staff	External	Routine	Patient	Yes

	Intervention and data feedback summary						Analysis of effect	
Paper	**Condition**	**Intervention**	**Feedback frequency**	**Data recipient**	**Data analyst**	**Data collection**	**Level of analysis**	**Risk adjust**
Carlhed, 2006	AMI	Feedback + QI	Real time	QI staff	Hospital QI team	Routine	Hospital	No
Carlhed, 2009	AMI	Feedback + QI	Real time	QI staff	Hospital QI team	Routine	Hospital	Yes
Carlhed, 2012	AMI	Feedback + QI	Real time	QI staff	Hospital QI team	Routine	Hospital	No
Sauaia, 2000	AMI	Feedback (written); Feedback (onsite)	Once	Management QI staff, Clinical staff	External	Study specific	Hospital & Patient	Yes
Beck, 2005	AMI	Feedback (written)	Once	Management	External	Routine	Hospital	No [a]
Berner, 2003	UA	HCQIP: Feedback + QI;OL: Feedback + QI led by OL	Once	HCQIP: QI staff OL: QI staff, + OL	External	Study specific	Hospital & Patient	Yes [b]
Heller, 2001	UA	Feedback (onsite)	Once	Clinical staff	External	Study specific	Hospital	Yes
Hollenbeak, 2008	AMI	Public reporting	Variable, NR	General public	NR	NR	Matched patient [c]	Yes
Moscucci, 2005	Received PCI	Public reporting	NR	General public	External	Routine	Patient	Yes

AMI: Acute myocardial infarction. HCQIP: Health Care Quality Improvement Program. NR: Not reported. NSTEMI: Non-ST-elevation myocardial infarction. OL: Local physician opinion leader. Other AMI: AMI patients not able to be classified as either STEMI or NSTEMI and given diagnosis related code (DRG code) of 410.9. PCI: Percutaneous coronary intervention. QI: Quality improvement. STEMI: ST-elevation myocardial infarction. UA: Unstable angina.

[a] Risk adjustment was done for odds ratios only. [b] Risk adjustment had no effect, therefore authors present only the unadjusted results. [c] Patients were propensity matched using patient and hospital level characteristics.

The following text summarises the results and the quality assessment for each of the 11 included studies, and interprets the findings with respect to the direction and strength of evidence regarding the effects of the evaluated method(s) of feeding back comparative hospital performance data.

Renzi and colleagues compared the direct feedback of comparative performance data to public reporting. The data was fedback in face-to-face meetings with general management, clinical directors and clinical staff. The study focused on performance with respect to the receipt of PCI within 48 hours. The main effect was observed in STEMI patients, for whom the proportion of patients receiving PCI within 48 hours increased more in the intervention group (absolute increase 17.4% vs. 6.9% in the control group, p value not reported) [Renzi, 2014]. In an earlier study report, a smaller improvement is reported for a larger population (including patients without clinical outcomes), which was statistically significant [Renzi, 2012]. There was no clinically meaningful improvement in the change in 30-day mortality in the intervention group compared to the control group. Indeed, the largest differences in mortality showed increasing mortality in the intervention group for patients in the 'Other AMI' category, and in STEMI patients who did not undergo PCI.

The Renzi study was not a randomised clinical trial, and the two studies received a moderate quality score. The main quality issues highlighted the possibility of confounding.

The significant improvement in the use of PCI for the patient group at highest risk provides some evidence of a positive intervention effect. However, the mortality data raises the possibility of unintended consequences of any actions that were taken as a result of the intervention, i.e. did the focus on improving use of PCI for STEMI patients lead to reduced quality of care for patients not undergoing PCI? One interpretation of the study findings is that the direct feedback of data to key stakeholders resulted in action being taken, but in the absence of specific support for the improvement process, the implemented changes resulted in adverse unintended consequences.

Carlhed et al evaluated an intervention involving the training of two cardiologists and two nurses in QI methods over six months, with an additional twelve months of ongoing support. The training included instruction on generating realtime comparative performance reports, and the development of local action plans. Ongoing support for QI activities was provided by telephone, video conferencing and site visits.

The first results paper reported meaningful and statistically significant improvements in the intervention group for the mean absolute percentage change in four of five process indicators, and a p-value of 0.065 for the fifth indicator (lipid lowering medications) [Carlhed, 2006]. The second paper reported process indicator outcomes for patients for whom clinical outcomes were collected. Meaningful and statistically significant improvements in the mean absolute percentage were reported for three indicators [Carlhed, 2009].

The intervention was associated with meaningful reductions in adverse clinical events (mortality, CV readmissions, and bleeding complications) relative to the control group, though the effect was only statistically significant for the bleeding outcome [Carlhed, 2009].

The third paper showed that the intervention effects did not extend beyond the reported indicators to other aspects of the AMI process of care, including stress testing, ECG, and appropriate use of reperfusion. The third study report also investigated whether the observed effect on the process indicators was sustained beyond three months following withdrawal of study support. The results show a trend towards convergence in the process indicators between the control and intervention groups (only a statistically significant difference in the use of clopidogrel remained).

This was an observational study, with moderate quality scores, though the paper reporting the analysis of clinical outcomes achieved the highest quality score (23/33). The main area of limitation concerned the potential for confounding.

This study provides the best evidence of a positive effect of the feedback of comparative performance data. The most prominent aspects of the intervention include the focus on clinical staff for the feedback of data, and the accompanying package for supporting the conduct of quality improvement. An area in which the intervention may have been improved is via more direct interaction with a broader set of senior stakeholders.

Beck et al evaluated the use of report cards presenting risk-adjusted performance data for twelve QI indicators for AMI, provided in multiple forms (printed, overheads, electronic) to encourage further dissemination to hospital staff. The mean results were in favour of the intervention for seven of the nine process indicators, but the magnitude of improvement was not large, and none of the differences were statistically significant. As for the process outcomes, there were no statistically significant differences in the change in mortality or readmissions over time, though the mean results favoured the intervention group for three of the four reported outcomes.

This was a randomized clinical trial, and achieved the highest quality score (30/33), with a score of 14/15 with respect to internal validity.

Despite the focus on the generation of detailed and robust comparative performance data, the study results do not provide evidence for the effectiveness of the intervention. Key limiting factors may include the non-directed and non-supported feedback of the performance data. Unlike the intervention adopted by Renzi et al, the comparative performance data was not fedback in a face-to-face setting involving senior stakeholders. The intervention also included only a limited form of follow-up - a reminder to encourage dissemination, sent after two months. The results imply that more direct feedback, and more intensive support mechanisms may be required to promote meaningful and significant improvements.

Berner et al conducted a three-arm clinical trial, comparing a no intervention with performance feedback and a QI program, with either an assigned co-ordinator, or with local opinion leader involvement. The evaluation focused on five process indicators. Compared to no intervention, the standard QI program showed no significant difference in any of the indicators, with the mean result showing a relative improvement over time in three indicators, and a relative worsening in two indicators. The opinion leader QI intervention showed a statistically significant relative improvement in the use of antiplatelet medication within 24

hours, compared to both the no intervention arm and the standard QI program (after adjusting for baseline compliance). The mean effect of opinion leader QI was mixed for the other four indicators.

This was a randomized clinical trial, and the study performed well with respect to internal validity. However, there were limitations with respect to the representativeness of the participating hospitals, and the clarity of the reported outcomes.

The authors cite ongoing hospital participation in other QI activities as a potential reason for the lack of a strong and consistent effect of either QI intervention. The relatively low intensity of the QI programs may also have contributed: co-ordinators were instructed in guidelines, monitoring data, and developing a QI plan, but no ongoing support and follow-up of QI activities was included.

Heller et al evaluated the effectiveness of the feedback of comparative performance data alongside opinion leader led educational sessions. The mean results imply that the intervention hospitals demonstrated improved relative performance in three of the nine indicators, whilst the control hospitals had better outcomes for the remaining six indicators. The *p* values for the time x intervention interaction terms, in multivariate regression-based analyses of nine process indicators, show that none of the terms reach significance at the 0.05 level (though the *p* value for beta blocker use was 0.07, and in favour of the intervention hospitals).

This was a randomized clinical trial, for which the quality was judged to be high (27/33). Internal and external validity was high (13/15), and the main area in which the study paper could have been improved concerned the reporting of the details of the intervention and increased clarity regarding the outcomes.

The QI activities undertaken alongside the feedback of performance data were limited, and the authors hypothesise that the lack of intervention effect may have been due to the "failure to follow-up the initial educational session" (p220).

Hollenbeak et al compared alternative forms of public reporting of comparative performance data, defined as intensive vs. limited public reporting. The intensive form of the reporting was defined as comparative data being available in three or more publicly accessible reports. The control group included hospitals receiving privately reported comparative data (that was not presented in three public reports), though details on the number of such hospitals were not reported.

The paper does not present direct estimates of the difference in differences between the pre-period (when all hospitals were classified as limited public reporting), and the post-period (when some States moved to an intensive public reporting system). Indirectly, the odds ratio for in-hospital mortality in AMI patients between limited reporting hospitals in Pennsylvania and other States in the pre-period is approximately 0.79 (95% CI 0.68 to 0.9), whilst the odds ratio between intensive reporting hospitals in Pennsylvania and limited reporting hospitals in other States in the post-period is approximately 0.64 (95% CI 0.55 to 0.73). These two results

indicate a non-statistically significant trend in favour of improved performance associated with intensive public reporting.

This was an observational study that achieved a moderate quality score (21/33), with the main concerns around potential confounding. The main limitation of this study is the reliance on a single measure of effect – in-hospital mortality, which has been criticised as having "low sensitivity (most quality problems do not cause death) and low specificity (most deaths do not reflect poor-quality care)" [p645, Scott et al, 2011]. Another potentially important limitation with the study design concerns the Statewide inclusion of Pennsylvanian hospitals, but the self-selected sample of non-Pennsylvanian hospitals, and in particular, the 35% reduction in the number of participating non-Pennsylvanian hospitals in the post-period.

This study supports the potential effectiveness of a more intensive approach to reporting comparative performance data, but more research is needed to confirm these findings using more meaningful measures of outcome, and more robust study designs.

Sauaia et al compared the on-site presentation of feedback data with the provision of mailed out data. All hospitals were also invited to submit a written quality improvement plan. Univariate patient-level analyses imply improved relative performance in the intervention group with respect to the receipt of thrombolytics within 1 hour, and reduced relative performance with respect to smoking cessation advice. Although the data were not presented, the text states that the multivariate analysis "mostly confirmed the results of the univariate analysis" and that the 'time x intervention' interaction term was only statistically significant for the thrombolytics within 1-hour indicator (in favour of the intervention).

The results suggest a focus on a single quality indicator - receipt of thrombolytics within 1 hour – at the intervention hospitals, which produced a significant improvement in that area of performance. However, there was a small sample size for this indicator, and large differences in baseline rates of 'thrombolytics within 1 hour' (55% in intervention, 84% in control).

This was a randomized controlled trial, for which the quality criteria indicate a relative high standard (25/33). The main limitations concerned the reporting in the study paper, and the potential for confounding. Despite the conduct of analysis to adjust for potential confounding, it is not clear if the adjustment analyses were sufficient to remove the effects of the large differences in the baseline characteristics of the control and intervention hospitals.

This study did not include a control group that did not receive feedback, but assessed alternative forms of presenting the comparative data. Compared to the mail out of the data, the one-off on-site presentation of the data was associated with little, if any effect. The authors state that "compared with the number of nurses and quality managers, few physicians attended the presentations", and identify the limited ability of attendees at the presentations to act on the information as a factor in the lack of intervention effect. An additional factor may have been the lack of follow-up beyond the presentation of the data.

Moscucci et al evaluated the effects of public reporting on case selection for PCI. The unadjusted results show statistically significantly lower in-hospital mortality in New York

(with public reporting) compared to Michigan (without public reporting). However, the intervention effect disappears in the multivariate analysis. Based on their analyses, the authors suggest that public reporting had an unintended effect on case selection – with a reduced likelihood of intervening on higher risk patients.

This was a cross-sectional study (i.e. no baseline data), which achieved a moderate quality score (21/33). The limiting factors were distributed across issues relating to reporting, and internal and external validity.

The focus on patient undergoing PCI, and the use of in-hospital mortality as the only measure of effect, provides limited information on the effect of public reporting of comparative performance data, but the results imply no positive effect on outcomes (and a hypothesis regarding negative effects associated with case selection).

4.3 Discussion

The review identified moderate evidence of effect for some of the evaluated interventions involving the feedback of comparative performance data, but significant uncertainty remains around the magnitude of effect, and the optimal approach to the feeding back comparative performance data.

From the review, we highlight three potentially key components of such interventions: quality of the comparative data; form and focus of feedback; and ongoing support for quality improvement activities.

None of the reviewed studies evaluated an intervention that addressed all of these components, though individual studies were identified that provided examples of good practice with respect to the individual components:

- generation of detailed and robust comparative performance data (Beck et al)
- face-to-face feedback of the performance data to general management, clinical directors and clinical staff (Renzi et al)
- supported development of local action plans and ongoing support for QI activities (Carlhed et al)

In the absence of strong evidence for the effectiveness of any of the evaluated interventions, we would suggest that future approaches to the feedback of comparative performance data incorporate each of the above three design components.

Chapter 5 Conclusions

For too long, investment in health care has been synonymous with new technologies and services. Variation in clinical practice has long been recognised [Wennberg & Gittelsohn, 1973], and despite successes [Kennedy et al, 2010], investment in this area remains limited. This research report has described three connected studies that highlight the huge potential for quality improvement activities to improve patient outcomes at lower cost than achieved through investing in new technologies.

In Chapter 2, we reported on an investigation of the use of process mining techniques to the analysis of health care data. Process mining involves the application of data mining techniques to process information. It was first applied to the analysis of industrial processes, and has only recently been tested with respect to health care processes, which are generally less uniform and deterministic than industrial process. The following findings were derived from the review of existing studies, and a direct application of process mining techniques to health care data:

- Mining based analyses of workflow and performance are useful tools for representing processes at individual hospitals, and for initiating conversations re:comparative practices,
- The cognitive load of comparing complex processes across multiple sites, whilst controlling for potential differences in casemix, was large,
- Cluster analyses (the mining based identification of process clusters at each hospital) have the potential to highlight process differences within common casemix groupings, but these analyses were hampered by the need to represent continuous timing variables (e.g. time in the ED).

As a result of these findings, the applied comparative analysis of processes of care for patients presenting at the ED with chest pain applied statistical models of key process indicators along the clinical pathway, controlling for relevant patient and hospital level covariates directly.

The study reported in Chapter 3 demonstrated the feasibility and relevance of the analysis of routinely collected data to identify important variation in costs, outcomes, and processes of care across hospitals. A dataset of de-identified administrative and clinical data was compiled for over 15,000 patients who presented at the emergency department of four hospitals with chest pain. The data were linked to inform costs and outcomes over a 12-month follow-up period.

Significant, casemix-adjusted differences were observed in costs and 12 month clinical outcomes (cardiovascular admissions and mortality) across providers, which resulted in the identification of an apparent outlier (non-benchmark) hospital that was incurring higher costs and achieving poorer patient outcomes than other hospitals.

The analyses of costs and outcomes were supported by analyses of process variables representing harder process variables (e.g. inpatient admission rates, use of invasive

procedures, and length of stay) than those more commonly reported process indicators, such as appropriate medication on presentation and discharge. These analyses found that the non-benchmark hospital had significantly lower rates of admission and use of invasive diagnostic and management procedures in patients with negative diagnostic tests on presentation and no pre-existing circulatory conditions. These results suggest that clinical pathways for low risk patients might be reviewed at this hospital. Another potential area for investigation concerned the inpatient length of stay for higher risk patients, which was significantly longer at the non-benchmark hospital.

The analyses reported in Chapter 3 combined administrative and clinical data. Administrative hospital data is collated by hospitals for reporting to a centralized body (e.g. the National Hospital Cost Data Collection in Australia). The advantages of such data for informing hospital performance include that they are routinely collected in a standardized manner by all hospitals. Disadvantages include the potentially limited nature of the data collected, and the lag between data collection and availability of the data for analysis.

Clinical data comprises data recorded during the process of managing patients, and may include details of care provided, such as the timing and results of diagnostic tests and investigations, and interventions. The advantages of clinical data include the additional detail on the clinical pathway, and the potential for quicker access to the data. The main disadvantage is the non-routine collection and reporting of clinical data, which means that primary data extraction has to be undertaken. However, with increasing use of electronic data systems, the extraction of large amounts of clinical data is more feasible than previously, when such data could only be extracted manually from paper-based clinical records. In South Australia, the Clinical Reporting Repository provides centralized and real-time access to clinical data from six metropolitan public hospitals, as well as being linked to population-based mortality data.

The delays in gaining access to the data precluded the planned application of the findings to inform quality improvement activities in the study hospitals. Timely access to routinely collected data that can be analysed to provide robust estimates of variation in costs, outcomes, and processes is a key prerequisite for the use of comparative performance data to inform quality improvement in health care.

Important ongoing analyses will assess the relative value of the administrative and clinical data. Given the importance of timeliness of the analysis to inform investment in quality improvement, and the time lag in accessing administrative data, the available dataset provides a key opportunity to test the validity of analyzing hospital performance using clinical data alone.

As reported in Chapter 4, the literature review of interventions that have incorporated the feedback of comparative performance data to improve services for acute coronary syndromes provided moderate evidence of effect. The findings suggest that the reporting of comparative performance data is not necessarily sufficient to motivate significant and sustained improvements in patient outcomes. From the review, we identified three components of the evaluated approaches to the feedback of comparative performance data that appear to be important:

- quality of the comparative data;
- form and focus of feedback;
- ongoing support for quality improvement activities.

The findings from the review support the potential value of the comparative performance data generated for this research project, whilst emphasising the need for such data to be used in the context of broader, well designed quality improvement systems that target key stakeholders and provide ongoing support to those stakeholders to implement and maintain quality improvement activities.

5.1 Next steps

As noted in Chapter 1, there is a funding bias in favour of new technologies, ahead of strategies to improve existing services. In Australia, the conduct of quality improvement initiatives are generally funded at the level of the hospital, i.e. with funds diverted directly from the provision of services. This inevitably means that quality improvement is assigned a low priority.

A report from the King's Fund in the UK recommended a program of work to identify causes of variation at specific local levels, and to prioritise those variations and causes that have the most important impact on equity, effectiveness, efficiency and patient health outcomes [Appleby et al, 2011]. The analytic methods presented in this report provide robust estimates of the consequences of variation in a format that is comparable to the data used to inform funding decisions for new technologies, as well as supporting evidence on the likely causes of the variation.

The reported analyses of comparative performance with respect to costs, outcomes, and processes should also be validated in a range of clinical areas, using data that is accessible within a timeframe that meets the expectations of a live quality improvement system. Once validated, analyses of the consequences and causes of variation in clinical practice could be undertaken across a broad range of clinical areas on an ongoing basis. The results could be part of a transparent process of selecting priority clinical areas and providers for quality improvement intervention.

In line with the results of our systematic review, the King's Fund report suggested that the reporting of unwarranted variations may not sufficiently incentivise action. They suggest that providers be required to respond to evidence of significant variation from benchmark performance, and that it may be necessary to "explore the development of harder-edged, locally focused incentives to encourage action to deal with unwarranted variation".

A national body, such as NHS IQ or the ACSQHC could manage the allocation of an improvement fund. Following the identification of priority areas for improvement, these bodies would fund and monitor the improvement process. Alternatively, such bodies could have a more direct role, for example, co-ordinating quality improvement collaboratives, or developing and assigning experienced quality improvement teams to lead improvement projects.

Based on the reviewed literature around the use of comparative performance data as part of a quality improvement process, such intervention requires careful design and evaluation. The findings from the review should inform the development of a generic quality improvement system, that is designed around the availability of ongoing comparative performance data on costs, outcomes, and processes of care.

References

Acute Coronary Syndrome Guidelines Working Group (ACSGWG). Guidelines for the management of acute coronary syndromes. Medical Journal of Australia. 2006;184 (8 Suppl):S1-S32.

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

Appleby J, Raleigh V, Frosini F, Bevan G, Gao H, Lyscom T, Variations in health care: The good, the bad and the inexplicable, The King's Fund 2011, www.kingsfund.org.uk/publications/variations-health-care

Aroney C, Aylward P, Kelly A, Chew D, Clune E. on behalf of the Acute Coronary Syndrome Guide- lines Working Group. 2006. Guidelines for the Management of Acute Coronary Syndromes 2006. The Medical Journal of Australia 184 (8 Suppl) (2006), S1–S32.

Beck, CA, H. Richard, et al. 2005. "Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, a cluster randomized trial." JAMA 294(3): 309-317.

Ben-Tovim DI, Bassham JE, Bolch D, Martin MA, Dougherty M, Szwarcbord M, Lean thinking across a hospital: redesigning care at the: Flinders Medical Centre. Australian Health Review 2007: 31(1): 10-15

Berner ES, Baker CS, et al. 2003. "Do local opinion leaders augment hospital quality improvement efforts? A randomized trial to promote adherence to unstable angina guidelines." Med Care 41(3): 420-431.

Binder M, Dorda W, Duftschmid G, Dunkl R, Froschl KA, Gall W, et al. 2012. On Analyzing Process Compliance in Skin Can- cer Treatment: An Experience Report from the Evidence-Based Medical Compliance Cluster (EBMC2). In LNCS 7328, J. Ralyte (Ed.). Springer, 398–413.

Birkhead JS, Walker L, et al. 2004. "Improving care for patients with acute coronary syndromes: initial results from the National Audit of Myocardial Infarction Project (MINAP)." Heart 90(9): 1004-1009.

Blum T, Padoy N, Feuner H, Navab N. 2008. Workflow mining for visualisation and analysis of surgeries. International Journal of Computer Assisted Radiology and Surgery 3, 5 (2008), 379–386.

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

Carlhed R, Bellman C, et al. 2012. "Quality improvement in coronary care: analysis of sustainability and impact on adjacent clinical measures after a Swedish controlled, multicenter quality improvement collaborative." J Am Heart Assoc 1(4): e000737.

Carlhed R, Bojestig M, et al. 2006. "Improved adherence to Swedish national guidelines for acute myocardial infarction: the Quality Improvement in Coronary Care (QUICC) study." Am Heart J 152(6): 1175-1181.

Carlhed R, Bojestig M, et al. 2009. "Improved clinical outcome after acute myocardial infarction in hospitals participating in a Swedish quality improvement initiative." Circ Cardiovasc Qual Outcomes 2(5): 458-464.

Chen, J C. Public reporting of health system performance: Review of evidence on impact on patients, providers and healthcare organisations: An Evidence Check rapid review brokered by the Sax Institute (http://www.saxinstitute.org.au) for the Bureau of Health Information, 2010

Claxton K, Briggs A, Buxton MJ, Culyer AJ, McCabe C, Walker S, Sculpher MJ, Value based pricing for NHS drugs: an opportunity not to be missed? BMJ. Feb 2, 2008; 336(7638): 251–254

Cox J, Johnstone D, et al. 2008. "Optimizing healthcare at the population level: results of the improving cardiovascular outcomes in Nova Scotia partnership." Healthc Q 11(2): 28-41.

Cox JL. 1999. "Optimizing disease management at a health care system level: the rationale and methods of the improving cardiovascular outcomes in Nova Scotia (ICONS) study." Can J Cardiol 15(7): 787-796.

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

Downs SH and Black N. 1998. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. Journal of Epidemiology Community Health, 52, 377-384

Duckett, S, Coory, M, Kamp, M, Collins, J, Skethcher-Baker, K & Walker, K. 2008. VLADs for Dummies, ed. Clinical Practice Improvement Centre, Wiley Publishing Australia Pty Ltd, Milton, QLD

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

Eckard N, Janzon M, Levin LA. 2011. Compilation of cost-effectiveness evidence for different heart conditions and treatment strategies. Scandinavian Cardiovascular Journal 45, 2 (April 2011), 72–6.

Eijkenaar F, Emmert M, Scheppach M, Schoffski O, Effects of pay for performance in health care: A systematic review of systematic reviews, Health Policy 2013; 110(2):115-30

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

Ellerbeck EF, Kresowik TF, et al. 2000. "Impact of quality improvement activities on care for acute myocardial infarction." Int J Qual Health Care 12(4): 305-310.

Fernandez-Llatas C, Meneu T, Benedi J, Traver V. 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).

Goodman SG, Huang W, et al. 2009. "The expanded Global Registry of Acute Coronary Events: baseline characteristics, management practices, and hospital outcomes of patients with acute coronary syndromes." Am Heart J 158(2): 193-201 e191-195.

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

Gunther C. 2009. Process mining in flexible environments. Ph.D Dissertation. Technische Universiteit Eind- hoven.

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

Han B, Jiang L, Cai H. 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.

Heller RF, D'Este C, et al. 2001. "Randomised controlled trial to change the hospital management of unstable angina." Med J Aust 174(5): 217-221.

Hickey A, Scott I, et al. 2004. "Using clinical indicators in a quality improvement programme targeting cardiac care." Int J Qual Health Care 16 Suppl 1: i11-25.

Hollenbeak CS, Gorton CP, et al. 2008. "Reductions in mortality associated with intensive public reporting of hospital outcomes." Am J Med Qual 23(4): 279-286.

Hoot NR, Aronsky D, Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions, Annals of Emergency Medicine 2008; 52(2): 126-36

Hornix PTG. 2007. Performance Analysis of Business Processes through Process Mining. Master's thesis. Eindhoven University of Technology.

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

Impellizzeri FM, Bizzini M, Leunig M, Maffiuletti NA, Mannion AF. 2009. Money matters: exploiting the data from outcomes research for quality improvement initiatives. European Spine Journal 18, Suppl 3 (2009), 348–59.

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

Karnon J, Caffrey O, Pham C, Grieve R, Ben-Tovim D, Hakendorf P, Crotty M. Applying risk adjusted cost-effectiveness (RAC-E) analysis to hospitals: estimating the costs and consequences of variation in clinical practice. Health Econ. 2012; doi: 10.1002/hec.2828.

Kennedy PJ, Leathley CM, Hughes CF. Clinical practice variation. Med J Aust. 2010;193(8 Suppl):S97-9.

Kim SM, Jang WM, et al. 2012. "Korean National Health Insurance value incentive program: achievements and future directions." J Prev Med Public Health 45(3): 148-155.

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

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

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

Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet. 2004;363(9415):1147-54

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

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

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

Mans R, van der Aalst W, Vanwersch R, Moleman A. 2012. Process Mining in Healthcare: Data

challenges when answering frequently posed questions. In Proceedings of BPM 2012 Workshops.

McGregor C, Catley C, James A. 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.

Moscucci M, Eagle KA, et al. 2005. "Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases." J Am Coll Cardiol 45(11): 1759-1765.

Moscucci M, Rogers EK, et al. 2006. "Association of a continuous quality improvement initiative with practice and outcome variations of contemporary percutaneous coronary interventions." Circulation 113(6): 814-822.

National Health and Medical Council. 2009. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Accessed on 2014, August 3. Accessed at: https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf

Olivari Z, Steffenino G, et al. 2012. "The management of acute myocardial infarction in the cardiological intensive care units in Italy: the 'BLITZ 4 Qualita' campaign for performance measurement and quality improvement." Eur Heart J Acute Cardiovasc Care 1(2): 143-152.

Partington A. Wynn MT, Suriadi S, Ouyang C, Karnon J. 2014. Process Mining of Clinical Processes: Comparative Analysis of Four Australian Hospitals, ACM Transactions on Management Information Systems (in press)

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

Perimal-Lewis L, Qin S, Thompson C, Hakendorf P. 2012. Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach. In HIKM 2012 (CRPIT), Vol. 129. ACS, 59–66.?

Pham C, Caffrey O, Ben-Tovim D, Hakendorf P, Crotty M, Karnon J. Evaluating the effects of variation in clinical practice: a risk adjusted cost-effectiveness (RAC-E) analysis of acute stroke services. BMC Health Serv Res. 2012;12(1):266.

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

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

Productivity Commission. 2005. Impacts of Advances in Medical Technology in Australia. Research report. Productivity Commission, Melbourne.?

Quaglini S. 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.?

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

Renzi C, Asta F, et al. 2014. "Does public reporting improve the quality of hospital care for acute myocardial infarction? Results from a regional outcome evaluation program in Italy." Int J Qual

Health Care.

Renzi C, Sorge C, et al. 2012. "Reporting of quality indicators and improvement in hospital performance: the P.Re.Val.E. Regional Outcome Evaluation Program." Health Serv Res 47(5): 1880-1901.

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

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

Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JI, Coiera EW, et al, CareTrack: assessing the appropriateness of health care delivery in Australia, Med J Aust 2012; 197 (2): 100-105

Sauaia A, Ralston D, et al. 2000. "Influencing care in acute myocardial infarction: a randomized trial comparing 2 types of intervention." Am J Med Qual 15(5): 197-206.

Schade CP, Cochran BF, et al. 2004. "Using statewide audit and feedback to improve hospital care in West Virginia." Jt Comm J Qual Saf 30(3): 143-151.

Scobie S, Thomson R, McNeil J, Phillips P. 2006. Measurement of the safety and quality of health care. MJA 184 (2006), S51–S55.

Scott I, Denaro C, Hickey A, Bennett C, Mudge A, Sanders D, et al. 2004. Optimising care of acute coronary syndromes in three Australian hospitals. Int. Jour. for Quality in Health Care 16, 4 (2004), 275–284.

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

Scott IA, Darwin IC, et al. 2004. "Multisite, quality-improvement collaboration to optimise cardiac care in Queensland public hospitals." Med J Aust 180(8): 392-397.

Siegel B, Sears V, et al. 2012. "A quality improvement framework for equity in cardiovascular care: results of a national collaborative." J Healthc Qual 34(2): 32-42; quiz 42-33.

Snow RJ, Engler D, et al. 2003. "The GDAHA hospital performance reports project: a successful community-based quality improvement initiative." Qual Manag Health Care 12(3): 151-158.

Song M, Gunther C, van der Aalst W. 2009. Trace Clustering in Process Mining. In BPMWorkshops. 109–120.

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

Sueta CA, Schenck A, et al. 2001. "Hospital interventions improve care of North Carolina Medicare patients with acute myocardial infarction." N C Med J 62(2): 69-73.

Suriadi S, Mans RS, Wynn MT, Partington A, Karnon J. 2014. Measuring Patient Flow Variations: A Cross-organisational Process Mining Approach. In The Proceedings of the 2nd Asia Pacific Conference on Business Process Management (Lecture Notes in Business Information Processing), C. Ouyang and J.Y. Jung (Eds.), Vol. 181 (in press). Springer.

Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M, Reduced mortality with hospital pay for performance in England, N Engl J Med 2012; 367: 1821-1828

van der Aalst W, Adriansyah A, de Medeiros AKA, Arcieri F, Baier T, Blickle T, et al. 2012.

Process Mining Manifesto. In BPM Workshops 2011 (LNBIP), F. Daniel, K. Barkaoui, and S. Dustdar (Eds.), Vol. 99. Springer, Heidelberg, 169–194.

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

vom Brocke J, Simons A, Niehaves B, Niehaves B, Reimer K, Plattfaut R, Cleven A. 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.

Weijters AJMM, van der Aalst W, Alves de Medeiros AK. 2006. Process Mining with the Heuristics Miner Algorithms. BETA Working Paper Series WP 166 (2006).

Wennberg J, Gittelsohn. Small area variations in health care delivery. Science. 1973;182(4117):1102-8

Zhang H, Alexander JA, et al. 2005. "Data feedback and clinical process improvement in acute myocardial infarction." Am Heart J 149(5): 856-861.

Appendix 1 Literature Search strategy (PubMed)

1	acute coronary syndrome[mh] OR myocardial infarction[mh] OR angina, unstable[mh]
2	acute coronary syndrome*[tiab] OR myocardial infarction[tiab] OR STEMI[tiab] OR segment elevation[tiab] OR NSTEMI[tiab] OR non-STEMI[tiab] OR NSTEACS[tiab] OR unstable angina[tiab] OR AMI[tiab]
3	**(#1) OR (#2)**
4	Feedback[mh]
5	Outcome and process assessment (health care)/organization & administration[mh] OR Quality assurance, health care[mh] OR Quality improvement[mh] OR Quality indicators, health care[mh] OR Program evaluation[mh]
6	feedback[tiab]
7	(measure[tiab] OR measures[tiab]) AND (performance[tiab] OR activity[tiab] OR activities[tiab] OR outcome[tiab])
8	(data analys*[tiab]) AND (continuous[tiab] OR systematic[tiab] OR rapid-cycle[tiab] OR real-time[tiab])
9	(indicator*[tiab]) AND (clinical[tiab]) OR quality[tiab])
10	(monitoring[tiab]) AND (real-time[tiab] OR continuous[tiab])
11	reporting[tiab] OR report card[tiab]
12	**(#4) OR (#5) OR (#6) OR (#7) OR (#8) OR (#9) OR (#10) OR (#11)**
13	hospitals[mh] OR cardiology service, hospital[mh] OR Emergency Medical Services[mh] OR Emergency Service, Hospital[mh] OR Regional health planning[mh] OR Rural health services[mh] OR Intensive Care Units[mh]
14	hospital*[tiab]) OR cardiology[tiab] OR emergency medical services[tiab] OR emergency department[tiab] OR ambulance[tiab] OR paramedic[tiab] OR prehospital[tiab] OR regional health network[tiab] OR regional network[tiab] OR rural health service[tiab] OR rural network[tiab]
15	**(#13) OR (#14)**
16	**(#3) AND (#12) AND (#15)**
17	**(#3) AND (#12) AND (#15) AND ("2000/01/01"[PDat] : "2014/04/31"[PDat])**
18	review[Publication Type] OR guideline[Publication Type] OR editorial[Publication Type] OR letter[Publication Type] OR systematic review[tiab] OR meta-analysis[tiab]
19	**(#3) AND (#12) AND (#15) AND ("2000/01/01"[PDat] : "2014/04/31"[PDat]) NOT (#18)**
20	(#3) AND (#12) AND (#15) AND ("2000/01/01"[PDat] : "2014/04/31"[PDat]) AND (#18)

mh: mesh heading. tiab: title/abstract. PDat: publication date. *: wildcard (truncated term).

Appendix 2 Quality Criteria Checklist

The checklist used to assess study quality was based on Downs and Black's *Checklist for Measuring Study Quality* (1998). It was adapted for use with hospital level interventions, where potential quality issues may arise at both patient and hospital levels.

For each question, a positive answer was awarded one point. Negative answers or instances where the information was not reported were awarded zero points. Points were summed to give sectional and total scores for each paper.

Quality Criteria Checklist

Reporting (total out of 12) [a]

1. There was a clear statement of the hypothesis, aims or objectives.
2. The main outcomes were described in the introduction or methods.
3. Patient characteristics were clearly described (e.g. inclusion/exclusion criteria).
4. Hospital characteristics were clearly described (e.g. inclusion/exclusion criteria).
5. The intervention of interest was clearly described (the whole intervention and the feedback component specifically).
6. The distribution of the principal patient-level confounders between the intervention and control groups were listed and/or clearly described.
7. The distribution of the principal hospital-level confounders between the intervention and control groups were listed and/or clearly described.
8. The main findings were clearly described (i.e. numerators /denominators).
9. Estimates of the random variability for the main outcomes were provided (e.g. interquartile ranges, standard errors, standard deviations or confidence intervals).
10. The characteristics of patients who were lost to follow-up were described. [This was interpreted as a description provided of the patients who were excluded from study due to missing data]. Answer yes if none or very small numbers and no if not reported.
11. The characteristics of hospitals lost to follow-up were described. Answer yes if none or very small and no if not reported.
12. The actual probability values were presented for the main outcomes of interest (i.e. not $p<0.05$, with the exception of $p<0.001$).

External validity (total out of 4) [b]

13. Those patients asked to participate were representative of the entire population (e.g. the source population was identified and how patients were selected was specified; e.g. via consecutive or random sampling).
14. Those hospitals asked to participate were representative of the entire population (e.g. the source population was identified and how hospitals were selected was specified; e.g. via random sampling).
15. Those patients agreeing to participate were representative of entire population (e.g. provided response rates, validated the distribution of patient characteristics between the sample and the population).
16. Those hospitals agreeing to participate were representative of entire population (e.g. provided response rates, validated the distribution of hospital characteristics between the sample and the population).

Internal validity - bias (total out of 8)

17. An attempt was made to blind study participating patients (i.e. patients had no way of knowing which intervention they received).
18. An attempt was made to blind study participating hospitals (i.e. hospitals had no way of knowing which intervention they received).
19. An attempt was made to blind those measuring the main outcomes (this includes blinding the researchers who conducted the analysis of the dataset).
20. All analyses were planned at the study outset. Any results based on data dredging were clearly identified.
21. The length of the follow-up periods were the same, or adjustment was made for differing lengths of follow-up.
22. The statistical tests used for the main outcomes were appropriate.
23. There was reliable compliance with the intervention (i.e. there was no non-compliance, contamination, or cross-over, or if these occurred, the misclassification bias was toward the null).
24. The main outcome measures were accurate (i.e. valid and reliable, for example the measures were clearly described or authors referred to other papers to demonstrate their accuracy).

Internal validity – confounding (total out of 8)

25. Patients in the intervention and control groups were recruited from the same population (i.e. using the same data source and screening criteria).
26. Hospitals in the control and intervention groups were recruited from the same population (the 'population' of interest was as defined by the authors).
27. The intervention and control groups were recruited over the same time period (i.e. the pre-intervention measurement period covers the same block of time for both groups; the post-intervention measurement period covers the same block of time for both groups).
28. Hospitals were randomised to the intervention or control group.
29. Randomisation was concealed from hospitals until after hospital recruitment was complete and irrevocable.
30. There was adequate adjustment for confounders in the analysis of the main findings (e.g. analysis by intention to treat; listing, investigating and accounting for the effects of all likely and/or important confounders).
31. The number of patients lost to follow up were reported and too small to make a difference. This included the number of patients excluded from the study due to missing data.
32. The number of hospitals lost to follow up were reported and were too small to make a difference.

Power (total out of 1)

33. A sample size calculation was reported.

Total score (out of 33)

[a] The following question was removed from the reporting section of the Downs and Black's Checklist: *All important adverse events were measured and reported.* This was not applicable for all included studies.
[b] The following question was removed from the external validity section of the Downs and Black's Checklist: *The staff, places and facilities where treatment was delivered to patients were representative of the treatment of the majority of patients.* Given the interventions were delivered at a hospital level, and additional questions were added to the checklist to address hospital level representativeness, this was already captured by previous questions in the checklist.