Analysing homogenous patient journeys to assess quality of care for patients admitted outside of their ‘home-ward’

LUA PERIMAL-LEWIS¹, SHAOWEN QIN¹, CAMPBELL H THOMPSON², PAUL HAKENDORF³

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

² Medicine
  The University of Adelaide, South Australia

³ Clinical Epidemiology Unit
  Flinders Medical Centre, South Australia

Abstract

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

Keywords: homogenous in-patient journey analysis, process mining, ‘home-ward’, outliers, inliers, quality of care, cluster analysis

1 Introduction

Australian Public Hospitals are faced with increasing demands for hospital services. This is largely due to the aging Australian population and its associated demand for the usage of acute care facilities. The demand on Australian Emergency Department (ED) has been consistently increasing at an average of 1.8% per annum (FitzGerald, Toloo et al., 2012). Over the last 2 decades Australians’ median age has increased by 4.8 years and population projections suggest increase in the proportion of population over the age of 65 thereby indicating that the demand on ED services will be an ongoing issue (FitzGerald, Toloo et al., 2012).

As with any organisations, hospitals too have to comply with strict measure of operational efficiency and effectiveness by conforming to Key Performance Indicators (KPI). Hospitals have mature processes that collect data on various quality measures in order to report and adhere to these KPIs. One such KPI is the improvement in ED throughput measures such as reducing the time first seen by a doctor, reducing did-not-wait rates and the reduction in ED Length of Stay (LOS) (Shetty, Gunja et al., 2012). Patient LOS is one of the criteria used to measure ED performance and a hospital’s performance in general. Performance is measured as percentage of patients who stayed beyond the established LOS target (Kolker, 2008). ED LOS measurement, although a functional performance indicator, could possibly contribute to the streaming of patients to any available wards regardless of whether the ward is an appropriate ward for the condition of the patient.

The complexity and diversity of hospital processes means that there are also diverse ways to measure the quality of patient care which varies based on the characteristics of the process area being studied. This study investigated the quality of care received by patients who were admitted to their ‘home-ward’ referred to as inliers and patients who were admitted outside of their ‘home-ward’ referred to as outliers. It is a common perception amongst clinicians that outliers have longer overall in-hospital LOS compared to inliers. It is also perceived that quality of care received by outliers is inferior to that of inliers.

At Flinders Medical Centre (FMC) where this study was undertaken, percentage of outlier patients was a regularly reported hospital performance indicator and therefore substantial effort is taken to collect the appropriate data needed in regards to the ‘home-ward’ status of the admitted patient.

The study indentified common variables or attributes used to measure quality of care and assessed how these attributes affect quality of care according to whether a patient was admitted to their ‘home-ward’ or outside of
their ‘home-ward’. The variables identified were ‘discharge summary sent within 2 days of discharge’, ‘in-hospital mortality’, ‘re-admitted within 7 days’, ‘total in-hospital LOS’ and ‘time spent in the ED’.

Discharge summary contains relevant information pertinent to a patient’s care during a hospital admission which is important to be communicated to primary health professionals who will continue a patient’s care or provide future care for a patient after discharge (Li, Yong et al., 2011). The same authors established an association between delayed dissemination or the absence of discharge summary and re-admission rate thus encouraging health professionals to complete discharge summary promptly. Prompt discharge summary dissemination has also been associated with decreased hospital re-admission. Re-admission rate within 3 months decreased when a patient followed-up on continuity of care by seeing a physician who had received the discharge summary (Van Walraven, Seth et al., 2002). The hypothesis was that patients admitted outside of their ‘home-ward’, the outliers will have higher re-admission rate because the discharge summaries for these patients were either not processed or delayed.

Inpatient LOS has become one of the many ways used to measure performance of a hospital. Patient mean LOS has been used to measure quality of care and hospital efficiency in terms of resource usage (Thomas, Guire et al., 1997). Lower than normal LOS could indicate that hospitals are discharging patients early possibly sacrificing quality of care (Thomas, Guire et al., 1997). The hypothesis was that outliers have longer overall LOS as their stay were probably prolonged as a consequence of being admitted outside of their ‘home-ward’ therefore not receiving the required level of care.

There are various studies establishing an association between ED overcrowding and in-hospital mortality. Richardson (2006) reported increased in-hospital mortality at 10 days amongst patients presenting at the ED during high ED occupancy. The hypothesis was that more patients would end up in an outlier ward during ED overcrowding due to the pressure to reduce ED congestion. As a consequence of inferior quality of care received by outliers, it was perceived that this group might have higher in-hospital mortality rate due to the delay in receiving treatment.

Health Care data analysis is traditionally done using various statistical techniques in order to report and hopefully forecast health care performances. New approaches in health care modelling and data analysis are emerging where more than one technique and approach are used to discover hidden information that might not be easily discovered from one approach. Combinations of techniques are used to complement the strength in each technique will give a better in-sight. This study aims to investigate the relationship between the quality of care attributes in regards to the patient’s inlier or outlier status by applying cluster analysis combined with statistical techniques. An in-depth evaluation of the patient flow processes using data from the Patient Journey Database was used to aid in identifying the relationships hidden within statistics alone.

2 Study Setting and Data

The analysis was undertaken on in-patient records for patients admitted to and discharged by the General Medicine (GM) service at Flinders Medical Centre (FMC). FMC is a public teaching hospital in South Australia and it attends to approximately 62,000 patients per annum. The GM service controlled about 100 in-patient beds out of about 500 beds in FMC as a whole. The analysis was carried out on in-patient records of the GM service only; that is, on those patients whose in-patient care had been allocated to a GM team. The wards that were ‘home-wards’ for this service were clearly defined. A home-ward is a ward that is equipped with the appropriate medical team and specialised equipment to treat the patient’s primary disease. Patients who were not allocated a ‘home-ward’ of the GM unit responsible for their care were defined as being an outlier and staying in an outlier ward.

The Patient Journey Database from FMC contains information on in-patients or officially admitted patients only and records detailed information on the journey or movements of a patient from the time of admission to the time of discharge. An individual patient could have multiple admissions at different points in time and each admission will be allocated with a unique journey number that remains the same until discharge. Each movement of the patient from one ward to another ward is recorded with a timestamp, so at any point the “start time” in a ward and the “end time” in a ward are known together with the name of the ward. Each ward occupied by a patient is appropriately marked to reflect whether the ward occupied was an inlier or an outlier ward. Patient admitted to an inlier ward is admitted to their ‘home-ward’. Timestamp for Admission is the combination the “Date” field and the “Admission Time” field. Timestamp for Discharge is the combination of “Date” field and “Discharge Time” field. Timestamp is a derived field. The individual patients are not identifiable at any point.

The original data set contained about 1.9 million records spanning from January 2003 to September 2009. To reduce the heterogeneous nature of the types of patients, various levels of record filtering were applied to reduce the dimension of the data set. The final record set which was used for the analysis only consisted of patient journeys that had been exclusively cared by the GM service from admission to discharge. If a patient’s journey
was under the care of a combination of GM service and non GM service, the journey was excluded. This level of filtering reduced the record set to about 24,439 patient journeys.

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

3 Methodology – Process Mining – Case Perspective

Process Mining uses event logs to discover organisational processes, control data, social and organisation structure (van der Aalst, Reijers et al., 2007). According to the same authors, processes could be analysed from the process perspective, the organisational perspective and the case perspective. The use of process mining techniques in the healthcare industry is becoming increasingly widespread. The complex nature of healthcare industry and varied processes makes the use of process mining techniques a viable method to gain insights into these processes (Perimal-Lewis, Qin et al., 2012). Mans, Schonenberg et al., (2008) used process mining techniques to identify bottleneck and to better understand the different clinical pathways taken by various groups of patients. Rebuge and Ferreira (2012) concluded that despite the proven success of process mining techniques, the complexity and the ad hoc nature of health data calls for the identification of right algorithm to handle noise in the data. It is common knowledge that healthcare is rich in data which presents a challenging task for researchers trying to discover knowledge using data from this domain. As with any knowledge discovery, gaining meaningful insight from data has to be accompanied with the knowledge rendered by domain experts to understand the intricacies behind complex health care decision making processes.

The notion of efficient patient care providing patient-centred approach has seen the emergence of various Health Information Systems (Vezyridis, Timmons et al., 2011). Electronic Patient Management or Tracking Systems have all become not only common but essential systems for any hospital. These information systems store invaluable information that can be used for knowledge discoveries.

Process mining enables the discovery of knowledge regarding a process. Process mining uses event or process logs to extract information regarding a process as it has taken place (van der Aalst, Reijers et al., 2007). These processes / event logs do not have to necessarily originate from a Workflow Management System. A process log could be derived from a dataset that contains an order of events which could be used to construct a process model that portrays the activity of the subject matter (van der Aalst, Reijers et al., 2007). In this study the event log was constructed from information collated from the Patient Journey Database. The process mining activities discussed in this paper is from the case perspective. The concept of event log as introduced by van der Aalst, Reijers et al., (2007) referred to as "history", "audit trail" and "transaction log" shapes the foundation of the event log used in this study. The individual patient journey is comparable to the concept of process instance introduced by the same authors. The information collated from the patient journey database is a derived event log. After constructing the event log, the event log was analysed as described and discussed in the following sections. The focus of the study is on gaining insight into the process of streaming patients to an outlier or an inlier ward and its effect on quality of care received by these patients.

Each patient journey is the process instance or the case being studied in relation to the activities on the patient journey. An activity is equivalent to a ward occupied by the patient which correlate to either an inlier or an outlier ward. Each case which in this study is the patient journey can be characterised by the values of the corresponding data elements (van der Aalst, Reijers et al., 2007). Data elements are the quality of care variables/attributes. The patient journeys are analysed to establish the relationship between the quality of care attributes of the journey against the amount of time a patient stayed in an outlier or an inlier ward. Table 1 shows a snippet of the dataset from the patient journey database.

![Table 1: Snippet of data used for process mining](image)

The pre-processed dataset from the patient journey database as discussed in this section forms the source of data for the rest of the analysis.

3.1 Data analysis to define outlier patients and inliner patients

The next task was an explorative analysis to discover a meaningful way to categorise the population of the GM patients into 2 distinct categories of outlier patients and inliner patients. At any given time, it is possible to establish from the original dataset whether a patient stayed in an outlier or inliner ward. Majority of patients had stayed in a combination of outlier and inliner wards. The patient journeys were categorised according to those who had overall in-hospital LOS of "0-3 Days", "4-7 Days", "8-30 Days" and "> 30 Days". The percentage of time spent in an outlier ward and the percentage of time spent in an inliner ward were derived for each journey to show the distribution of the percentage of outlier time versus the percentage of inliner time.

It was discovered that the distribution of percentage of time spent in an outlier ward was very similar across all 4 categories of LOS. Figure 1 shows the distribution of outlier hours for the overall GM patient journeys.
spending time in the ED waiting for an in-patient bed to become available after decision to admit. Naturally, as far as the data recorded, the time in ED after decision to admit is considered as outlier time. The presentation of the above information assisted the domain experts to further deliberate on how the ED time should be classified in regards to the overall definition of the outlier and inlier status for the overall patient journeys, which was important to address as this might confound the findings.

The distribution for the different LOS categories is not presented here because the trend across the groups was similar. Based on the information discovered from this analysis together with the insight from the domain experts, it was decided that the best classification of inlier patient journeys will be those journeys that spent “≥ 70% Inlier Hours” in their ‘home-ward’ and the best classification for outlier patient journeys will be those journeys that spent “≥ 70% Outlier Hours” outside their ‘home-ward’.

Figure 1 also shows the distribution of patient journeys with and without including the time spent waiting in the ED or ward “FMC” in the outlier hour calculation. This study was carried out on in-patient journeys, and in theory ward “FMC” or ED time should be zero however, this was not the case for many patient journeys. This indicates that many in-patients were spending time in the ED waiting for an in-patient bed to become available after decision to admit. Naturally, as far as the data recorded, the time in ED after decision to admit is considered as outlier time. The presentation of the above information assisted the domain experts to further deliberate on how the ED time should be classified in regards to the overall definition of the outlier and inlier status for the overall patient journeys, which was important to address as this might confound the findings.

Table 2: Breakdown of patient journeys according to LOS and time in outlier and inlier ward

<table>
<thead>
<tr>
<th>LOS</th>
<th>Number of journeys and (%)</th>
<th>Number and % of journeys with “≥ 70% Inlier Hours”</th>
<th>Number and % of journeys with “≥ 70% Outlier Hours”</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 3 Days</td>
<td>12012 (51.25)</td>
<td>3542 (29.45)</td>
<td>7273 (60.74)</td>
</tr>
<tr>
<td>4 - 7 Days</td>
<td>5637 (24.05)</td>
<td>758 (13.45)</td>
<td>4656 (82.85)</td>
</tr>
<tr>
<td>8 - 30 Days</td>
<td>5324 (23.79)</td>
<td>522 (16.13)</td>
<td>4223 (81.87)</td>
</tr>
<tr>
<td>Above 30 Days</td>
<td>652 (2.71)</td>
<td>34 (8.49)</td>
<td>533 (109.52)</td>
</tr>
</tbody>
</table>

Table 2 shows the breakdown of patient journeys with “≥ 70% Inlier Hours” and “≥ 70% Outlier Hours” for each LOS category. This classification of the outlier and inlier group captured about 90% of the GM patient journeys.

The dataset was further filtered to remove patient journeys with 100% ED time as these patient journeys might confound LOS analysis for the outlier patient group. According to the domain experts these patients’ health might have improved while waiting for a bed to become available and discharged from the ED as an outlier patient with short LOS. The other set of patient journeys that could also confound the outcome were those patient journeys who had stayed more than 30 days. According to the domain experts the longer a patient stays in the hospital, the more likely these patients would eventually end up in a ‘home-ward’. Prolonged LOS for these patients are normally not related to medical issues but more likely related to finding appropriate care outside of the hospital.

After excluding patient journeys that were discharged from the ED, those staying more than 30 days and patient journeys with missing attributes the final sample size derived for the outlier group was 2592 records and for the inliner group was 15213 records.

The rest of the analysis is based on investigating the quality of care received by these 2 groups of patient journeys. The patients who stayed “<70%” of their in-hospital stay in an outlier or an inlier ward were not included in this analysis as the aim of this study was focussing on the outliers and the inliers.

3.2 Cluster analysis

Acknowledging the diversity of GM patients, as well as the complexity and variability embedded in each patient journey, it was important to reduce the heterogeneity of the patient journeys to gain better insight from the data. Disregarding patient heterogeneity can mask the discovery of meaningful patterns in patient characteristics which can lead to misleading results (Armstrong, Zhu et al., 2011). The aim of cluster analysis is to group cases, which in this study are the patient journeys, into homogenous groups based on the natural structure of data (Tan, Steinbach et al., 2005). Cluster analysis is an exploratory technique which aims to group cases into clusters based on their similarities and dissimilarities (Luke 2005). Cases in the same cluster share similar characteristics and very dissimilar to cases belonging to other clusters (Mooi and Sarstedt 2011). Applying statistical methods to a homogenous cluster of patients would be much more meaningful in revealing insights that are otherwise hidden due to heterogeneity.

In this study, the patient journeys were clustered using the two-step cluster analysis in SPSS. Two-step cluster analysis was chosen because of its ability to handle both continuous and categorical variables (SPSS 2001). From the automatic number of clusters derived by this clustering procedure an optimal number of clusters were derived using exploratory method while taking into consideration of the practicality of having large or small number of clusters against the ratio between clusters and the goodness of fit for the derived model. In the 1st step, the automatic number of clusters is determined using Clustering Criterion by choosing either Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). The number of clusters derived for this data set using BIC and AIC were similar. SPSS computes the BIC and AIC for J clusters respectively as per equation (1) and equation (2) below (IBM 2011).
In equation (1), N stands for total number of records in the data set. In equation (2), $m_j$ is calculated as shown in equation (3). $K_a$, $K_b$ and $L_k$ in equation (3) stands for ‘total number of continuous variables used in the procedure’, ‘total number of categorical variables used in the procedure’ and ‘number of categories for the kth categorical variable’ respectively (IBM 2011). In the 2nd step, the initial number of clusters derived in the 1st step is further refined. This is done by finding the largest increase in the distances between the 2 closest clusters (IBM 2011). The distance between 2 clusters is calculated by using log-likelihood distance measure, which is the decrease in log-likelihood as the clusters are combined into 1 cluster.

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

(1)

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

(2)

The patient journeys were clustered based on the quality of care variables and their outlier or inlier status. Patient journeys with outlier status were journeys with “≥ 70% Outlier Hours” and patient journeys with inlier status were journeys with “≥ 70% Inlier Hours”. The variables chosen have been assessed for collinearity between variables to ensure that they were unique in identifying distinct clusters. Table 3 shows the clustering results of the 2 homogenous clusters.

The characteristics of patients in both clusters are listed in Table 4 above. The number of female patients is higher in both the clusters. Charlson co-morbidity Index (CI) is the most widely used clinical index for the evaluation of co-morbidities (Simon, Beland et al., 2012). CI is a pre-calculated variable for every patient admission and was supplied with the data set. Patients in cluster 2 had a higher Charlson co-morbidity Index (CI) score suggesting that these patients were sicker than those in cluster 1. Age differences between patients in both clusters were small and the difference is not clinically significant.

Table 3: Patient journey composition in the 2 clusters

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size, %</td>
<td>9858 (50.5%)</td>
</tr>
<tr>
<td>Ratio size between cluster 1 and cluster 2</td>
<td>1.27</td>
</tr>
<tr>
<td>Average Silhouette *</td>
<td>0.06 (Good)</td>
</tr>
</tbody>
</table>

* Measure of cluster cohesion and separation

Table 4: Patient characteristics

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Cluster 1 (n=9968)</th>
<th>Cluster 2 (n=7837)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlson Index</td>
<td>1.39 (1.88) *</td>
<td>1.55 (2.63) *</td>
</tr>
<tr>
<td>Sex, n, (%)</td>
<td>Female, 5831, (58.5%) **</td>
<td>Female, 4474, (57.1%) **</td>
</tr>
<tr>
<td>Age, years</td>
<td>72.83 (18.32) *</td>
<td>71.61 (18.47) *</td>
</tr>
</tbody>
</table>

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

The ratio between the smallest and largest cluster is 1.27 which is a good ratio as the larger cluster is less than 2 times larger than the smaller cluster. Between the 0.60 average Silhouette and the ratio, the model is a good fit for the purpose of this study where all the quality of care variable identified had to be included in the model to give the insight required.

Table 5 summarises the quality of care attributes and their relative importance in deriving the 2 clusters. The predictor importance for each quality of care attribute is calculated as per Equation (5) where Ω is the set of predictor and evaluation fields and ‘sig’ is the p-value (IBM 2012). The values are relative; therefore the sum of values for all attributes is 1. An attribute with a value close to 1 is the most important attribute in deriving the cluster and a value close to 0 is the least important attribute.

$$\sqrt{I_j} = \frac{-\log_{10}(\text{sig}_j)}{\max_{j \in \Omega}(-\log_{10}(\text{sig}_j))}$$

(5)

The most significant quality of care attribute for deriving the 2 homogenous clusters was ‘discharge summary sent within 2 days of discharge’ with the relative importance of 1.0. Patient journeys in cluster 1 consists of patients where the discharge summaries were sent within 2 days for the entire, 100% of the cluster population as from the case to the centroid of every other cluster which the case belongs to’ and ‘B is the minimal distance from the case to the centroid of every other cluster’ (IBM 2012).
opposed to 94.8% of patient journeys who did not have their discharge summaries sent within 2 days suggesting inferior quality of care received by patients in Cluster 2.

The next quality of care attribute used to derive the 2 clusters was ‘in-hospital mortality’ with relative importance value of 0.61. None of the patients in cluster 1 died during their hospital admission. 8.3% of patients in cluster 2 died.

The next quality of care attribute in order of importance used to derive the 2 clusters was ‘readmission within 7 days’ with relative importance value of 0.4. Once again none of the patients in cluster 1 were re-admitted within 7 days; however 5.4% of patients in cluster 2 were re-admitted within 7 days.

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

The final quality of care attribute was ‘time spent in the ED’ with relative importance value of 0.03. Patients in cluster 1 spent slightly longer time in the ED (5.7 hours) compared to patients in cluster 2 with mean time of (5.18 hours).

The next important step in this study was to investigate if there were any significant differences between the quality of care attributes and patient characteristics in both clusters for those patients in the outlier and the inlier groups defined earlier. Table 6 and Table 7 below summarises the quality of care attributes and patient characteristics for the outliers and inliers in cluster 1 (n=9968) and cluster 2 (n=7837) respectively. The ‘Sig.’ column shows the p-value where significance level α < 0.05 is considered significant. Mann-Whitney U test was used for significance level test for continuous variables. Chi-square test was used for significance level test for proportions.

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

<table>
<thead>
<tr>
<th>Quality of Care Variables</th>
<th>Cluster 1 (n=9968)</th>
<th>Cluster 2 (n=7837)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean / Mode *</td>
<td>Predictor Importance **</td>
</tr>
<tr>
<td>Discharge Summary sent within 2 days of discharge</td>
<td>Yes (100%)</td>
<td>5</td>
</tr>
<tr>
<td>Readmitted within 7 days</td>
<td>No (100%)</td>
<td>0.81</td>
</tr>
<tr>
<td>Total in-hospital LOS, days</td>
<td>6.14</td>
<td>0.04</td>
</tr>
<tr>
<td>Time spent in the ED, hours</td>
<td>5.7</td>
<td>0.03</td>
</tr>
</tbody>
</table>

* Mean for continuous variables; Mode for binary variables (indicating the most frequent category); ** Relative importance of each quality of care variable/attributes in estimating the model

Table 5: Summary of quality of care variables/attributes

In cluster 2, there were 20.1% of outlier patient journeys and the rest were inliers. (see Table 7). Charlson Index (CI) was not statistically significant between the outliers and the inliers. This was similar to patients in cluster 1. Age differences between the outliers and the inliers were statistically significant (p=0.000; Mann-Whitney U test), however as noted before this is not of clinical significance. Contrary to patient journeys in cluster 1, although outliers spent slightly longer time in the ED compared to the inliers this was not statistically significant (p=0.778; Mann-Whitney U test) for patients in cluster 2. Similar to outliers in cluster 1, outliers in cluster 2 had shorter overall in-hospital LOS compared with the inliers and this was statistically significant (p=0.000; Mann-Whitney U test). The main differences between patients in cluster 1 and cluster 2 is in relation to the 3 quality of care attributes; ‘in-hospital mortality’, ‘readmitted within 7 days’ and ‘discharge summary sent within 2 days of discharge’. All patients with inferior quality of care in relation to these 3 attributes were in
cluster 2. In-hospital mortality between outliers and inliers in this cluster was not statistically significant. Outliers were re-admitted more than the inliers and this was statistically significant (p=0.022; chi-square test) suggesting that quality of care for outliers were inferior to those who were inliers. Less outliers had their discharge summaries sent within 2 days of discharge compared to the inliers and this was statistically significant (p=0.000; X² test). This again suggests an inferior quality of care for the outliers.

Table 7: Quality of care attributes comparison for inliers and outliers in cluster 2

Another set of analysis was carried out to compare the differences between characteristics of patients in cluster 1 and cluster 2 (table not shown). Apart from Age with (p=0.066; Mann-Whitney U test), CI and Sex were significantly different between patients in cluster 1 and cluster 2 with (p=0.000; Mann-Whitney U test) respectively. All quality of care variables were significantly different between patients in cluster 1 and patients in cluster 2 with (p=0.000; Mann-Whitney U test).

4 Discussion

The main differences between patients in cluster 1 and cluster 2 relates to the quality of care attributes. Patients in cluster 2 had inferior quality of care compared to those in cluster 1 regardless of whether they were outliers or inliers. In cluster 1, there were no in-hospital mortality, none were re-admitted within 7 days and discharge summaries were sent within 2 days of discharge for all the patients. Analysing patients in cluster 2 (those who had inferior quality of care) in regards to the outlier and inlier status revealed meaningful in-sight as the comparison was done on a cluster of patients with similar characteristics and quality of care attributes. One of the major challenges of the study was the considerable effort that went into exploring the data to discover the best way to derive the outlier and the inlier population. Over the period of 6 years, investigating the spread of time spent in an outlier ward and the spread of time spent in an inlier ward lead to the dichotomisation of this variable into "≥ 70%" of outlier or inlier time. The method used and the dichotomisation of this variable was believed to be the best approach for this data set to discover the effect of being a ward outlier or ward inlier on the quality of care received by these 2 groups of patients.

It was also necessary to investigate the effect of the quality of care attributes on the outliers and inliers status based on a homogenous group of patients. The relationships discovered based on analysing the quality of care attributes on homogenous clusters were different when the patients were not clustered. This study demonstrates the complexity of analysing hospital data and the need to identify group of patients with more similar characteristics from the raw data. Although outliers in both clusters were younger and the association was statistically significant, it was not a clinically significant association. This emphasised the importance of involving domain experts to make meaningful conclusion.

Patient co-morbidity, (CI) did not have a significant association on whether the patient was admitted in a “home-ward” or outside of a “home-ward”. This result was also obtained when the analysis was carried out without clustering the patients into 2 homogenous clusters.

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

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

Point-biserial was used to further analyse the correlation between total in-hospital LOS and in-hospital mortality for patients in cluster 2. The correlation showed lower in-hospital LOS was associated with patients who did not die whilst in-hospital (r = -0.139, p = 0.000). This finding calls for further research into the nature of this relationship. The finding reveals that outliers’ short LOS is not associated with in-hospital mortality.

Using point-biserial for patients in cluster 2, lower in-hospital LOS was associated with patients who were not readmitted within 7 days of discharge (r = -0.042, p = 0.000) suggesting that re-admission might not necessarily be linked with shorter LOS or the outliers. Again, applying advanced modelling and analysis would reveal further in-sight to this association.

5 Conclusion & Future Work

In conclusion, patients in cluster 2 had significant association with inferior quality of care attributes. Outliers had shorter in-hospital LOS contrary to the hypothesis. Also, contrary to the results of un-clustered
patient journeys, there were no significant association between being outlier and in-hospital mortality. Discharge summaries were not sent as promptly for the outliers compared to the inliers compromising continuation of care after discharge for the outlier group. Higher percentage of outliers was re-admitted within 7 days again suggesting inferior quality of care and conforming to the hypothesis.

Future work in regards to the inlier and outlier group of patients will include undertaking process mining to discover the patterns of patient movement and their correlation with LOS. Preliminary work on ward movement for the 2 groups has been initiated. Adapting process mining techniques to discover the control flow and ward movement for the two groups of patients will reveal further in-sight into the process of ward allocation in relation to quality of care.

Further analysis is needed to discover the reasons behind the longer ED time for outliers in Cluster 1 and why this is not a significant association for patients in cluster 2. Further study is also needed to discover the cause of shorter LOS for the outlying patients. According to domain experts, although LOS is a measure of efficiency, further analyses are needed to conclude the association with quality of care attributes. Additional in-sight is needed to understand the association between shorter in-hospital LOS and lower in-hospital mortality.

6 References


