Mathematical Modelling and Cluster Analysis in Healthcare Analytics – The Case of Length of Stay Management

Research-in-Progress

Daniel Gartner
School of Mathematics,
Cardiff University,
Cardiff, United Kingdom
gartnerd@cardiff.ac.uk

Rema Padman
The H. John Heinz III College,
Carnegie Mellon University,
Pittsburgh, United States
rpadman@andrew.cmu.edu

Abstract

Length of Stay (LOS) is an important metric of care quality and efficiency in hospitals that has been studied for decades. Longer stays lead to increased costs and higher burdens on patients, caregivers, clinicians and facilities. Understanding the characteristics of such outliers is important for developing actionable steps to address LOS reduction by eliminating the unnecessary variations in treatments that result in the higher LOS. In this context, the increasing availability of detailed inpatient data has the potential to enable the development of data-driven approaches that provide novel insights for the management of Length of Stay. This study examines clustering of inpatients using key clinical and demographic attributes to identify LOS outliers and investigates the opportunity to reduce their LOS by comparing their order sequences with similar non-outliers in the same cluster. Learning from retrospective data on 353 pediatric inpatients admitted for appendectomy, we develop a two-stage procedure that first identifies a typical cluster with LOS outliers. Our second stage analysis compares orders pairwise to determine candidates for switching to make LOS outliers similar to non-outliers. Results indicate that switching orders in homogeneous inpatient sub-populations within the limits of clinical guidelines may be a promising decision support strategy for LOS management. These novel data-driven insights can be offered as suggestions for clinicians to apply new evidence-based, clinical guideline-compliant opportunities for LOS reduction through healthcare analytics.

Keywords: Inpatient Length of Stay, Data Driven Outlier Analysis, Clustering

Introduction

Length of Stay (LOS) is an important quality metric in hospitals that has been studied for decades (Kim and Soeken (2005); Tu et al. (1995)). However, the increasing digitization of healthcare with Electronic Health Records and other clinical information systems is enabling the collection and analysis of vast amounts of data using advanced data-driven methods that may be particularly valuable for LOS management (Gartner (2015); Saria et al. (2010)). When patients are treated in hospitals, information about each individual is necessary to perform optimal treatment and patient scheduling decisions (Gartner and Padman (2017)), with the detailed data being documented in the current generation of information systems. Recent research has highlighted that resource allocation decisions can be improved by scheduling patient admissions, treatments and discharges at the right time (Gartner and Kolisch (2014); Hosseinifard et al. (2014)) while machine learning methods can improve resource allocation decisions and the accuracy of hospital-wide predictive analytics tasks (Gartner et al. (2015a)). Thus, using data-driven analytic methods to understand length of stay (LOS) variations and exploring opportunities for reducing LOS with a specific focus on LOS outliers is the goal of this study.

Using retrospective data on 353 inpatients treated for appendectomy at a major pediatric hospital, we first carry out a descriptive data analysis and test which (theoretical) probability distribution best fits our length of stay data. The results reveal that our data matches observations from the literature. In a first stage cluster analysis, we identify one potential outlier cluster while a descriptive analysis using box plot comparisons of this cluster vs. the union of patients assigned to all other clusters supports this hypothesis. In a second clustering stage, we analyse the patient sub-population who belongs to that outlier cluster and provide order prescription behaviour insights. More specifically, on a pairwise comparison, we describe which orders are likely to be selected in the outlier population vs. ones that are deselected in the non-outlier population and vice versa. Our findings reveal that four order items are not prescribed in the outlier population while in the non-outlier sub-population, these orders were prescribed. On the other hand, 51 orders were prescribed for the outlier patients which are not enabled in the non-outlier population. These novel data-driven insights can be offered as suggestions for clinicians to apply new evidence-based, clinical guideline-compliant opportunities for LOS reduction through healthcare analytics.

Related Work

Clustering algorithms and other machine learning approaches are discussed in Baesens et al. (2009); Jain (2010); Meisel and Mattfeld (2010); Olafsson et al. (2008) including an overview of operations research (OR) techniques applied to data mining. Mathematical programming and heuristics for clustering clinical activities in Healthcare Information Systems has been applied in Gartner et al. (2015b) while the identification of similar LOS groups has been studied by El-Darzi et al. (2009). Similar to our problem, the authors study the application of approaches to cluster patient records with similar demographic and clinical conditions. Using a stroke dataset, they compare the performance of Gaussian Mixture Models, k-means clustering and a two-step clustering algorithm. Determining cluster centers for patients in the Emergency Department (ED) is studied by Ashour and Okudan Kremer (2014). Having defined similar patient clusters, they study the improvement on patient routing decisions based on the clusters. Similarly, Xu et al. (2014) focuses their clustering problem on the ED. Their objective is to cluster patients to resource consumption classes determined by length of treatment while patient demographics are taken into consideration.

The approaches proposed in our paper can be categorized and differentiated from the literature of clustering in length of stay management as follows: Using mathematical modelling, we provide a concise description of the problem and develop a heuristic solution approach to solve the problem. In our experimental study, we provide a descriptive data analysis and an overview about the characteristics of our length of stay data. Fitting several distribution types and parameters of the theoretical probability distribution, we underline the skewed property of the probability distribution from our data. In a next step, we define homogeneous patient groups with respect to demographic, clinical attributes and length of stay outliers. Having learned homogeneous groups of inpatients, we evaluate patient orders within the group that potentially contains length of stay outliers and may be responsible for increasing LOS in that group. In conclusion, this study may be considered to be the first to link the discovery of similar clinical and demographic attributes in appendectomy inpatients while, within length of stay outlier clusters, we evaluate order switching possibilities and how they potentially reduce the number of LOS outliers.

Problem Description, Model Formulation and Solution Approach

In what follows, we provide a concise problem description followed by a mathematical model that clusters patients based on clinical and demographic information. Moreover, a heuristic approach to solve the clustering problem is shown. The section closes with an example in which we show the application of the heuristic algorithm.

Problem Description

Let $\mathcal{P}$ denote a set of individuals (hospital inpatients) and let $\mathcal{K}$ denote the set of clusters to which these individuals can be grouped. For each inpatient $p \in \mathcal{P}$, we observe a set of attributes $\mathcal{A}$ during the patient's LOS. Let $\mathcal{V}_a$ denote the set of possible values for attribute $a \in \mathcal{A}$ and let $v_{p,a} \in \mathcal{V}_a$ denote the value of attribute a for inpatient p. We wish to assign each patients' clinical and demographic attributes to clusters such that within the LOS outlier and non-outlier populations homogeneity across patients is maximized.

In the following, we will describe how we label patients as LOS outliers, followed by a mathematical model and a two-stage clustering approach: The first stage assigns patients' attributes to homogeneous clusters while clusters with high likelihood to contain LOS outliers can be identified. In a second stage, we filter patients assigned to these clusters and evaluate which patient orders may be switched to reduce length of stay in the LOS outlier patient sub-population. The section closes with an illustrative example.

Given the observed LOS of patient $p \in \mathcal{P}$, denoted by l_p, the 25 and 75 percentile of the LOS distribution denoted by q^{25} and q^{75}, respectively then we assign a patient the flag "outlier" using the following expression (Pirson et al. (2006)):

$$o_p = \begin{cases} 1, & \text{if } l_p > q^{75} + (1.5 \cdot (q^{75} - q^{25})) \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

Then, subset $\mathcal{P}^{out} \subset \mathcal{P}$ denotes all LOS outlier patients for whom $o_p = 1$.

Mathematical Model

We provide a formal description of the LOS Outlier Detection Problem followed by a mathematical model. Afterwards, we present a heuristic algorithm to solve the problem.

Decision Variables and Model Formulation

Using the binary decision variables

$$x^{clin}_{a,k} = \begin{cases} 1, & \text{if clinical attribute } a \in \mathcal{A}^{\text{clin}}\text{'s value is true in cluster } k \in \mathcal{K} \\ 0, & \text{otherwise,} \end{cases}$$

$$x^{dem}_{a,k} = \begin{cases} 1, & \text{if demographic attribute } a \in \mathcal{A}^{\text{dem}}\text{'s value is true in cluster } k \in \mathcal{K} \\ 0, & \text{otherwise,} \end{cases}$$

$$x^{outlier}_{k} = \begin{cases} 1, & \text{if cluster } k \text{ is chosen as the one that most likely contains LOS outliers} \\ 0, & \text{otherwise,} \end{cases}$$

$$z_{p,k} = \begin{cases} 1, & \text{if patient } p \text{ is assigned to cluster } k \\ 0, & \text{otherwise,} \end{cases}$$

the problem of clustering clinical and demographic information of patients reads as follows where α, β and γ are objective function weights:

$$\text{Minimize} \quad \alpha \cdot \sum_{p \in \mathcal{P}} \sum_{a \in \mathcal{A}^{\text{clin}}: v_{a,p}=1} \sum_{k \in \mathcal{K}} |x_{a,k}^{clin} - z_{p,k}| +$$
$$\beta \cdot \sum_{p \in \mathcal{P}} \sum_{a \in \mathcal{A}^{\text{dem}}: v_{a,p}=1} \sum_{k \in \mathcal{K}} |x_{a,k}^{dem} - z_{p,k}| +$$
$$\gamma \cdot \sum_{p \in \mathcal{P}^{out}} \sum_{k \in \mathcal{K}} |x_{k}^{outlier} - o_p| \tag{2}$$

subject to

$$\sum_{k \in \mathcal{K}} z_{p,k} = 1 \qquad \forall p \in \mathcal{P} \tag{3}$$
$$\sum_{k \in \mathcal{K}} x_{k}^{outlier} = 1 \tag{4}$$
$$x_{a,k}^{dem} \in \{0,1\} \qquad \forall a \in \mathcal{A}^{\text{dem}}, k \in \mathcal{K} \tag{5}$$
$$x_{a,k}^{clin} \in \{0,1\} \qquad \forall a \in \mathcal{A}^{\text{dem}}, k \in \mathcal{K} \tag{6}$$
$$x_{k}^{outlier} \in \{0,1\} \qquad \forall k \in \mathcal{K} \tag{7}$$
$$z_{p,k} \in \{0,1\} \qquad \forall p \in \mathcal{P}, k \in \mathcal{K} \tag{8}$$

Objective function (2) seeks to minimize three terms: First, the absolute deviation between patients' clinical attribute values and the clinical attribute values present in the cluster to which the patients are assigned. The second term is similar to the first term while here, demographic attributes are considered. Finally, the deviation between non LOS outlier patients in the LOS outlier cluster are minimized. All terms are weighted. Constraints (3) ensure that each patient is assigned to exactly one cluster. Constraints (4) ensure that exactly one LOS outlier cluster is selected. Expressions (5)–(8) describe the decision variables and their domains.

Clustering problems belong to the category of NP-hard problems (Meisel and Mattfeld (2010)) and solving real-world data containing hundreds of patients using an optimal approach (e.g. using our mathematical model) is unrealistic. We therefore developed a two-stage heuristic clustering approach which will be discussed next.

A Heuristic Clustering Approach

Our objective function 2 is similar to finding cluster centers of patient attributes in order to minimize deviations of each patient's attribute values with the ones of the cluster centres. One algorithm that pursues this objective is the k-means clustering algorithm (Jain (2010)). The algorithm is a method of vector quantization. It seeks to partition observations into clusters in which each observation belongs to the cluster with the nearest mean which serves as a prototype of the cluster.

Once we have found patients with similar clinical, demographic and LOS characteristics, we wish to separate patients within the cluster that has the highest likelihood to contain LOS outliers. In this stage, we extract patients with these attributes and evaluate the order prescription behaviour for these patients between outliers and the false positively clustered outliers which actually belong to the group of non-outliers. Orders prescribed by clinicians to patients are, for example, the application of drugs, examinations and therapies.

The heuristic is described in Algorithm 1.

Algorithm 1 Clustering of LOS outliers and identification of candidates for switching orders

Determine LOS outliers using Equation (1).
Set $K := 1$.
$x_k^{\text{outlier}} = 0 \quad \forall k = 1, \ldots, K$.
while $x_k^{\text{outlier}} = 0 \quad \forall k = 1, \ldots, K$ **do**
  $K := K + 1$.
  Run k-means algorithm.
  Calculate $x_k^{\text{outlier}} = 0$ based on cluster centroids.
end while
for all $k = 1, 2, \ldots, K$ **do**
  if $x_k^{\text{outlier}} = 1$ **then** Determine non-outlier reference patients $p^*(p)$ with respect to outliers $p \in \mathcal{P}_k^{\text{out}}$
    Determine $\mathcal{A}_k^{\text{o, off} \rightarrow \text{on}}$.
    Determine $\mathcal{A}_k^{\text{o, on} \rightarrow \text{off}}$.
  end if
end for

We first determine whether or not patients are LOS outliers using Equation (1). Afterwards, we initialize $K := 1$ and increment it until k-means algorithm includes the LOS outlier flag $x_k^{\text{outlier}} = 1$ where the subset of outlier patients in cluster k are denoted by $\mathcal{P}_k^{\text{out}} \subset \mathcal{P}^{\text{out}}$. Having determined patients with high likelihood of belonging to the group of outliers, we introduce a set $\mathcal{A}_k^{\text{o, off} \rightarrow \text{on}}$ for cluster $k \in \mathcal{K}$ which allows experts to evaluate orders which were switched off for outlier patients and were switched on for non-outliers. For each outlier cluster k, the set is determined by $\mathcal{A}_k^{\text{o, off} \rightarrow \text{on}} := \{a \in \mathcal{A}^{order} | v_{a,p^*(p)} - v_{a,p} = 1 \quad \forall p \in \mathcal{P}_k^{out}\}$ in which $p^*(p)$ denotes the reference non-outlier patient given outlier patient p. Similarly, for each outlier cluster k, we introduce the set $\mathcal{A}_k^{\text{o, on} \rightarrow \text{off}} := \{a \in \mathcal{A}^{order} | v_{a,p^*(p)} - v_{a,p} = -1 \quad \forall p \in \mathcal{P}_k^{out}\}$ to analyze which orders were given to LOS outlier patients while the non-outlier reference patient didn't receive the order.

A Motivating Example

Table 1 shows a sample data set with $\mathcal{P} = \{1, 2, \ldots, 10\}$ patients, $\mathcal{A}^{\text{dem}} = \{1, 2, 3\}$ attributes and $\mathcal{A}^{\text{order}} = \{1, 2, \ldots, 5\}$ orders. Demographic attribute values for patient $p \in \mathcal{P}$ are given by $v_{1,p}, \ldots, v_{3,p}$, length of stay is given by l_p and orders are given by $v_{4,p}, \ldots, v_{8,p}$.

$p \in \mathcal{P}$	1	2	3	4	5	6	7	8	9	10
$v_{1,p}$ (gender = male)	1	1	0	1	0	1	1	0	0	0
$v_{2,p}$ (age $\geq$ 6 months)	0	1	1	0	1	1	1	1	0	0
$v_{3,p}$ (diagnosis = sepsis)	0	1	1	1	0	0	0	0	0	0
l_p (in hours)	25	28	17	29	75	14	70	40	21	13
o_p	0	0	0	0	1	0	1	0	0	0
$v_{4,p}$ (blood test 1)	0	0	1	0	1	1	0	1	0	1
$v_{5,p}$ (blood test 2)	1	0	1	1	1	1	0	0	0	0
$v_{6,p}$ (medication 1)	1	0	0	0	1	1	1	1	1	0
$v_{7,p}$ (medication 2)	1	0	0	0	0	1	0	0	0	1
$v_{8,p}$ (medication 3)	0	1	1	1	1	0	0	1	0	0
$p^*(p)$	–	–	–	–	6	–	6	–	–	–
K_p	2	1	1	1	2	2	2	2	2	2

Table 1. Patients, attributes, attribute values and lengths of stay

The table reveals that, for example, patient $p = 2$ is a male patient with age greater or equals 6 months and was diagnosed with a sepsis. His length of stay is 28 hours and only medication 3 was ordered.

In the first stage (see Algorithm 1), we calculate the LOS outlier flags o_p for all patients $p \in \mathcal{P}$. The lower and upper quartiles come up to 18 and 37.25 hours, respectively and the threshold to determine whether a patient can be considered an outlier comes up to $q^{75} + (1.5 \cdot (q^{75} - q^{25})) = 66.125$ hours. Accordingly, two patients are assigned the LOS outlier flag. We now run k-means clustering algorithm starting with setting $K = 2$ such that the set of clusters is $\mathcal{K} := \{1, 2\}$.

Assume, $k-$means algorithm assigns the outlier flag to a cluster and that demographic attributes $a = 0, 1$ and $a = 2$ are clustered in cluster $k = 1$ while attributes $a = 0$ and a =1 are clustered in cluster $k = 2$. The subsets of clustered patients for $k = 1$ and $k = 2$ come up to $\mathcal{P}_1 = \{2, 3, 4\}$ and $\mathcal{P}_2 = \{1, 5, 6, 7, 8, 9, 10\}$, respectively. The cluster assignment of each patient is presented in Table 1 by the corresponding K_p-value which gives the cluster index of patient p. We now determine the most similar non-outlier reference patient $p^*(p)$ with respect to the demographic and clinical attributes of the outlier patient p which is given in Table 1. '–' means that the patient is not an LOS outlier and therefore has no reference patient. One can observe that for cluster $k = 2$, patient 6 is always chosen as non-outlier reference patient for the two outliers. Another observation is that for outlier patient $p = 7$, the non-outlier reference patients ties, i.e. $p^*(7) = 6$ and $p^*(7) = 10$. This is because for patients 6 and 10 it occurs 3 times that the binary vectors of attribute values are different. In this case, the tie is broken by minimum patient index. From all reference patients, the order switchings come up to: $\mathcal{A}_1^{\text{o, on} \rightarrow \text{off}} = \emptyset$, $\mathcal{A}_1^{\text{o, off} \rightarrow \text{on}} = \emptyset$, $\mathcal{A}_2^{\text{o, on} \rightarrow \text{off}} = \{8\}$ and $\mathcal{A}_1^{\text{o, off} \rightarrow \text{on}} = \{4, 5, 7\}$. This means that switching medication 3 from on to off should be evaluated by physicians for patients having demographic and clinical attributes from cluster $k = 2$. On the contrary, we can observe that switching blood test 1, 2 and medication 2 from off to on should also be evaluated by physicians.

Results

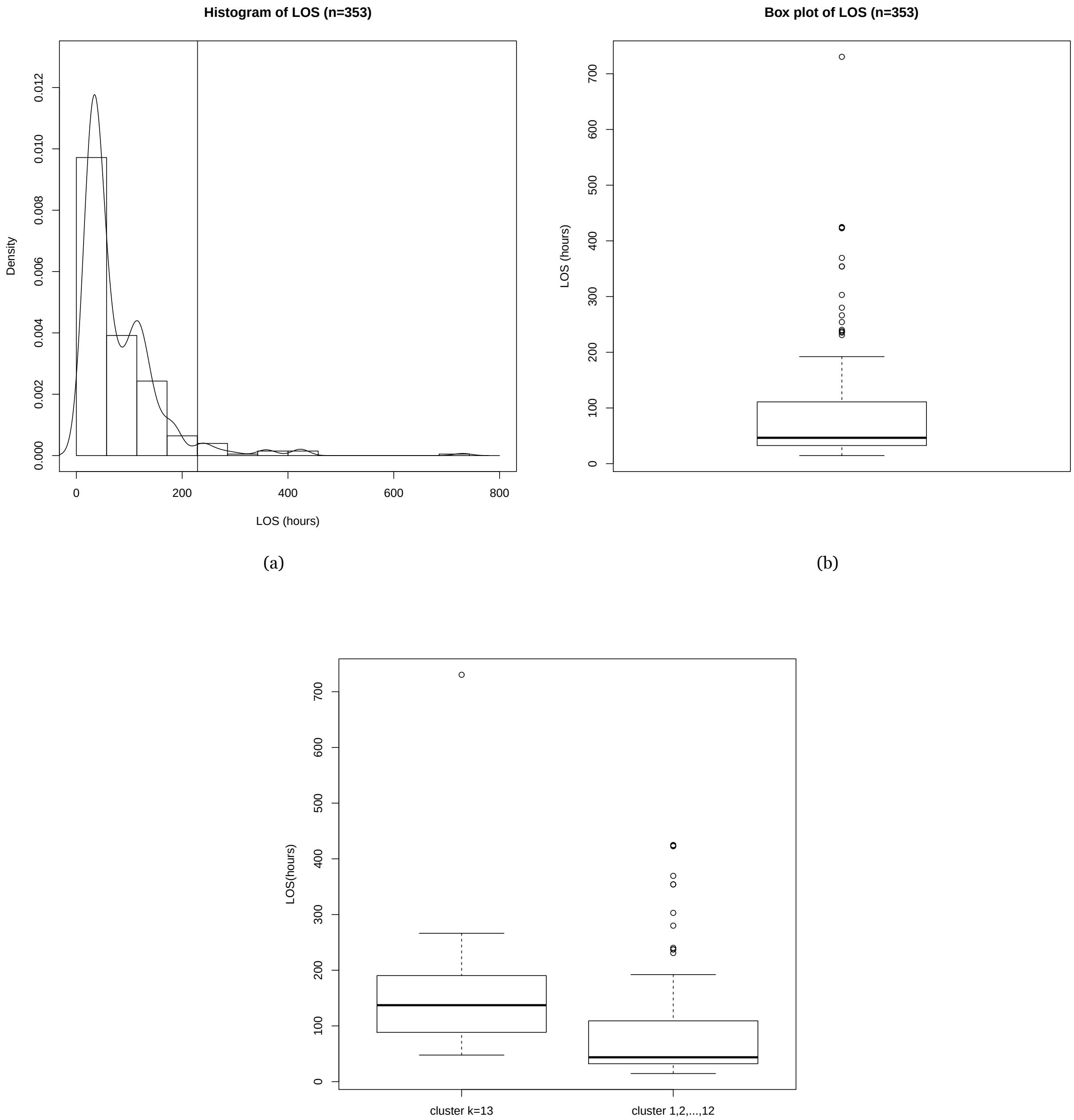
The data for this study were obtained from a pediatric hospital in the U.S. In total, $|\mathcal{P}| = 353$ appendectomy patients were hospitalized for, on average, 78.968 hours. Important variables extracted from the data warehouse include, among others, diagnosis codes, gender, age and 636 unique orders that were entered using Computerized Physician Order Entry. All patient-identifiable health information was removed to create a de-identified dataset for this study. Table 2 provides an overview about the attributes assessed.

Attribute	Data type	Distinct attribute values or bins
Diagnosis code	nominal	14 (e.g. ICD 9 code '540.0')
Outlier	nominal	2 (yes, no)
Emergency type	nominal	3 (e.g. '7 – very urgent')
Gender	nominal	2 (male, female)
Age in days at admission	nominal	9 (e.g. '0–739 days')
APR DRG Severity	nominal	4 (e.g. 'moderate')
Laparoscopic appendectomy	nominal	2 (yes, no)
636 unique order IDs	nominal	2 (yes, no)

Table 2. Attributes assessed for clustering and classifying LOS outliers

Descriptive Analysis of Length of Stay

A histogram of the LOS distribution including a Gaussian kernel density curve is shown in Figure 1(a). We used Equation (1) to determine the outlier LOS threshold o_p which is 229.140 hours. The figure reveals a skewed distribution with a density maximum at the first interval. Another observation is a large proportion of patients after the outlier LOS threshold. A boxplot of the LOS is shown in Figure 1(b). One can observe that the median is very close to the first quartile and some LOS outliers can be observed after the 95 percentile.



(a) (b)

(c)

Figure 1. LOS distribution (a), LOS box plot (b) clustered patients' LOS (c)

Fitting Distributions

To investigate whether a parametric model may be used to fit the data, we ran experiments with 9 distributions such as Beta, Log-normal, Weibull and Erlang. Our results revealed that the Beta distribution results in the best fit with respect to the squared error between the empirical and the best theoretical distribution. The log-normal distribution fits second best and its results of the fitting process will be analysed in more detail: Both the Chi-Square (CS) and the Kolmogorov-Smirnov (KS) test resulted in $p < 0.01$ while the CS-test run with 7 intervals and 4 degrees of freedom resulted in a $p < 0.005$. The optimal parameters of the (theoretical) log-normal distribution's expected value and variance come up to $\mu = 72$ and $\sigma^2 = 123$, respectively with a LOS-intercept of 14 (hours) based on the empirical minimum LOS value. Using this distribution to fit our data, the squared error comes up to 0.137. The result that the log-normal distribution fits very well is not surprising and confirms assumptions from the literature (Min and Yih (2010)).

Patient Clustering and Order Switching Results

In our first stage clustering, we varied the number of k until we reached a cluster in which the outlier flag was present. The first cluster was $k = 13$. The clinical attributes present in the outlier cluster are shown in Table 3(a) and a summary statistics of the outlier cluster is shown in Table 3(b). The table shows that ICD-9 code 540.1 – 'Acute appendicitis with peritoneal abscess', an emergency type of 4, a moderate APR DRG severity and 'laparoscopic appendectomy' are the attributes in which outlier patients are most likely to be present. Figure 1(c) shows a LOS boxplot of patients of which the demographic and clinical attributes belong to this cluster vs. all other patients. In a second stage, we determined the order switching patterns which revealed that the number of orders more than doubled when comparing the outlier cluster with the non-outlier cluster. One explanation for this phenomenon is that the length of stay is longer and therefore more orders are likely to be prescribed to patients. Now, comparing both clusters, we observed $|\mathcal{A}_{13}^{\text{o, on} \rightarrow \text{off}}| = 52$ occurrences with a switch from on-off while a off-on was only observed $|\mathcal{A}_{13}^{\text{o, off} \rightarrow \text{on}}| = 4$ times. In the latter case, we predominantly observed order switches in drug and diet prescriptions.

$\mathcal{A}$	v_a
Diagnosis code	540.1
Emergency type	4
APR DRG Severity	Moderate
Laparoscopic appendectomy	yes

(a)

Number of Data Points	17
Min Data Value	47.7
Max Data Value	730
Sample Mean	178.9
Sample Std Dev	156.5
1^{st} quartile	88.5
2^{nd} quartile	137.2
3^{rd} quartile	190.4

(b)

Number of Data Points	336
Min Data Value	14.5
Max Data Value	424.9
Sample Mean	73.9
Sample Std Dev	66.7
1^{st} quartile	32.2
2^{nd} quartile	43.7
3^{rd} quartile	109.1

(c)

Table 3. Clinical attributes present in outlier cluster $k = 13$ (a), its summary statistics (b) and summary statistics of patients not belonging to it (c)

Discussion and Implications

Assuming that the patient population in the outlier cluster could be moved towards the non-outlier cluster through order switching, we can determine a lower LOS bound. Applied to our dataset, the total length of stay could be reduced from 78.97 to 76.11 hours which equals to a 3.8% LOS reduction. In practice and to create a decision support tool which involves clinicians, similar reference patients may be presented to a clinician when treating each particular patient. A clinician may then decide to what extent order switching is appropriate within the limits of clinical guidelines.

Summary and Conclusions

In this paper, we have developed a mathematical program and heuristic clustering approach of patients for the management of length of stay outliers for pediatric appendectomy. We provided a two-stage clustering method to cluster patients based on similar clinical, demographic and length of stay characteristics and applied it to a data set including more than 350 patients. We retrieved a cluster of patients in which LOS outliers are likely to occur. In a second stage, we compared order prescription for LOS outliers with the ones for patients who have similar clinical and demographic characteristics but are non-outlier patients. Future work will develop an integer programming-based improvement heuristic by using our heuristic as initial solution. Another next step is to evaluate whether our work can be extended towards clustering patients having chronic conditions such as asthma and for managing readmissions instead of LOS. We also plan to incorporate evidence from clinical guidelines into our model and methods.

References

Ashour, O., and Okudan Kremer, G. 2014. "Dynamic patient grouping and prioritization: a new approach to emergency department flow improvement," *Health Care Management Science* pp. 1–14.

Baesens, B., Mues, C., Martens, D., and Vanthienen, J. 2009. "50 years of data mining and OR: upcoming trends and challenges," *Journal of the Operational Research Society* (60:Supplement 1), pp. 16–23.

El-Darzi, E., Abbi, R., Vasilakis, C., Gorunescu, F., Gorunescu, M., and Millard, P. 2009. "Length of stay-based clustering methods for patient grouping," in *Intelligent Patient Management*, S. McClean, P. Millard, E. El-Darzi, and C. Nugent (eds.), Springer, pp. 39–56.

Gartner, D. 2015. *Scheduling the hospital-wide flow of elective patients – early classification of diagnosis-related groups through machine learning*, Springer, Lecture Notes in Economics and Mathematical Systems ed., Heidelberg.

Gartner, D., and Kolisch, R. 2014. "Scheduling the hospital-wide flow of elective patients," *European Journal of Operational Research* (223:3), pp. 689–699.

Gartner, D., Kolisch, R., Neill, D. B., and Padman, R. 2015a. "Machine Learning Approaches for Early DRG Classification and Resource Allocation," *INFORMS Journal on Computing* (27:4), pp. 718–734.

Gartner, D., and Padman, R. 2017. "Mathematical Programming and Heuristics for Patient Scheduling in Hospitals: A Survey," in *Handbook of Research on Healthcare Administration and Management*, N. Wickramasinghe (ed.), Hershey, Pennsylvania: IGI Global, pp. 627–645.

Gartner, D., Zhang, Y., and Padman, R. 2015b. "Workload Reduction Through Usability Improvement of Hospital Information Systems – The Case of Order Set Optimization," *Proceedings of the International Conference on Information Systems (ICIS)* Fort Worth, TX.

Hosseinifard, S., Abbasi, B., and Minas, J. 2014. "Intensive care unit discharge policies prior to treatment completion," *Operations Research for Health Care* (3:3), pp. 168–175.

Jain, A. 2010. "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters* (31:8), pp. 651–666.

Kim, Y.-J., and Soeken, K. 2005. "A meta-analysis of the effect of hospital-based case management on hospital length-of-stay and readmission," *Nursing Research* (54:4), pp. 255–264.

Meisel, S., and Mattfeld, D. 2010. "Synergies of operations research and data mining," *European Journal of Operational Research* (206:1), pp. 1–10.

Min, D., and Yih, Y. 2010. "Scheduling elective surgery under uncertainty and downstream capacity constraints," *European Journal of Operational Research* (206:3), pp. 642–652.

Olafsson, S., Li, X., and Wu, S. 2008. "Operations research and data mining," *European Journal of Operational Research* (187:3), pp. 1429–1448.

Pirson, M., Dramaix, M., Leclercq, P., and Jackson, T. 2006. "Analysis of cost outliers within APR-DRGs in a Belgian general hospital: two complementary approaches," *Health Policy* (76:1), pp. 13–25.

Saria, S., Rajani, A., Gould, J., Koller, D., and Penn, A. 2010. "Integration of early physiological responses predicts later illness severity in preterm infants," *Science Translational Medicine* (2:48), pp. 48–65.

Tu, J., Jaglal, S., and Naylor, C. 1995. "Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery," *Circulation* (91:3), pp. 677–684.

Xu, M., Wong, T., and Chin, K. 2014. "A medical procedure-based patient grouping method for an emergency department," *Applied Soft Computing* (14:Part A), pp. 31–37.