IJASRM
International Journal of Advanced Scientific Research and Management, Vol. 1 Issue 8, August 2016.
www.ijasrm.com
ISSN 2455-6378

Modeling and Predicting Patient Length of Stay: A Survey

Aya Awad[1], Mohamed Bader-El-Den[2] and James McNicholas[3]

[1] School of Computing, University of Portsmouth
Portsmouth PO1 3HE, UK

[2] School of Computing, University of Portsmouth
Portsmouth PO1 3HE, UK

[3] Consultant in Critical , Queen Alexandra Hospital
Portsmouth Hospitals NHS Trust, UK

Abstract

Over the past few years, there has been increased interest in data mining and machine learning methods to improve hospital performance. Research has focused on prediction of measurable outcomes, including mortality, risk of complications and length of hospital stay. Length of stay is an important metric both for healthcare providers and patients, and is influenced by numerous factors. In particular length of stay in critical care is of great significance, both to patient experience and the cost of care, and is influenced by factors specific to the highly complex environment of the critical care unit.
This paper examines a range of length of stay applications in acute medicine and the critical care unit. It also focuses on the methods of analysing length of stay prediction. Moreover, the paper provides a classification for the analytical methods to length of stay prediction associated with a grouping of relevant research papers published in the domain. This study focuses on publications from the year 1984 till 2015 related to the domain of survival analysis and length of stay prediction. In addition, the paper highlights some of the gaps and challenges of the domain.

Keywords :

1. Introduction

Healthcare expenditure constitutes a significant share of the gross domestic product (GDP) of many countries. For example, in 2012 healthcare spending in the UK reached nearly a tenth (9.3%) of GDP [17], [84]. Government funding in many countries has fallen behind patient care costs, leaving healthcare institutions face the growing number of patients [63]. Accordingly, cost containment has become one of the most critical challenges in healthcare today. Hospitalization constitutes the principal cost of patient care and is therefore a main focus in healthcare management [10], [28].

Patient hospital length of stay typically refers to the number of days that an inpatient stays in a healthcare facility during a single admission [37]. It is considered one of the major indicators for the consumption of hospital resources [55], [57]. It also provides a better understanding of the flow of patients through a healthcare system which is essential for evaluating both the operational and clinical functions of such systems. Previous research have attempted to group patients by their medical condition, assuming that each disease, illness, or procedure is associated with a recommended LOS [68]. Grubinger et al. in 2010 [1] refer to these systems as diagnosis- related-group(DRG) systems. In addition, a relative value ,namely Case Mix Index (CMI) can be assigned to a diagnosis- relate-group (DRG) of patients in a medical care environment used in determining the allocation of resources to care for and/or treat the patients in the group. However, both studies assumed that all patients who fall within the same diagnosis- related-group are the same. However, length of stay is a complex metric affected by other factors including each in-dividual's demographics, treatment complexity, complications and discharge planning which may stretch the LOS beyond the target range. A model that helps to predict a patients length of stay during a single visit - the time from hospital admission until discharge - can be an effective tool for health care providers to plan for preventive interventions and to improve the utilization of hospital resources [6].

This paper reviews LOS applications in acute medicine and critical care unit. Moreover, the paper classifies the analytical methods available in the literature over the past three decades. In addition, the paper highlights some of the gaps and challenges of the domain.

The rest of this paper is organised as follows; Section II provides a survey of the different applications of LOS in the health care domain. Section 3 provide a classification a classification for the analytical methods to length of stay prediction associated with a grouping of relevant research papers published in the domain. Finally, the conclusion is drawn in Section 4.

2. Applications of LOS

LOS is often used as a surrogate for other outcomes in research, where those outcomes cannot be measured; for example as a surrogate for Hospital Mortality or Intensive Care Unit Mortality. LOS is also a parameter, which has been used to identify severity of illness and healthcare resource utilisation [14], [41], [50]. As a surrogate outcome measure it is not highly regarded, as it is subject to influences, which may not bear upon the outcome of real interest. For example, a diagnosis which simply requires a prolonged period of hospital care, but which confers a low Hospital Mortality, would bias the use of LOS as a surrogate for mortality; a condition which required a series of complex treatment interventions might prolong the stay, without necessarily conferring a high mortality risk; conversely presentations with a high illness severity, as quantified by APACHE II [2], ICNARC score [34], SOFA [78] or MPM [48], might be associated with a short LOS, because of an early decease, but also a high mortality, further undermining the correlation between LOS and mortal ity.

Moreover, LOS in hospital may be affected by factors unrelated to the disease, such as the availability of social care or community nursing support. There is an analogous effect upon discharge from critical care to the ward, in the event that there are insufficient ward beds for timely ICU discharge [5], [64]. Finally LOS may also be influenced by characteristics of the organisation including hospital management style [60], [75].
As an important determinant of both healthcare costs and patient experience, it is a high priority that LOS be optimal; it follows that it is also important therefore to identify any factors which affect it. The following two subsections will examine LOS applications in acute medicine and critical care, highlighting factors affecting LOS prediction.

A. LOS in Acute Medicine

This section presents previous studies on the different applications used in modeling LOS and its association with influencing factors with respect to patient flow. Patient flow typically refers to the progressive movement of a patient through a sequence of processes [20]. Reducing delays and making sure that the patient receives the right care at the right time will have a significant beneficial effect on the quality of service. In turn, this will improve patient outcomes and reduce the cost of care.

In 2012, Freitas et al. [24] studied variables associated with high LOS outliers, together with some hospital characteristics (administrative, economic and teaching characteristics). Re- sults show that age, type of admission and hospital type were significantly associated with high LOS outliers. Moreover the study conducted by Caetano et al. [13] showed that the top 3 influential input attributes were the hospital episode type, the physical service where the patient is hospitalized and the associated medical speciality. However, hospital related factors on their own are not sufficient to accurately predict LOS.

An important variable associated with LOS prediction and common is several studies is the nutritional status of a patient prior to admission. Research conducted in [3], [16], [21], [39], [65], [85], examined the effect of the variable malnutrition on patient LOS. In the study of Robinson et al. in 1987 [65], on average LOS was 15.6 days for a malnourished patient group versus 10 days for the well nourished group. However, in 1997 Chima et al. [16], showed that LOS for the two patient groups was 6 days for the at-risk for malnutrition population and 4 days for the not-at-risk for malnutrition population. The significant decrease in LOS may reflect the fact that all health institutions are under pressures of payment and reviews by government and other third-party payers. In addition, according to Correia et al. [39], length of hospital stay is shorter in the well-nourished patients; median of 6 days versus 9 days for the malnourished. Warnold and Lundhon [85], studied the clinical significance of preoperative nutri- tional status in 215 non-cancer patients. The variables inves- tigated included weight loss, weight-for-height index, serum protein levels (serum albumin, transferrin, prealbumin, retinol-binding protein), delayed hypersensitivity skin testing, arm circumference, and triceps skinfold thickness. Of the markers evaluated, weight-for-height index, arm muscle circumference, serum albumin level, and weight loss correlated significantly to post-surgery outcome. In addition, Epstein et al. [21] also emphasized that underweight patients have 40% higher LOS than normal weight patients. Also according to Burritt et al. [3], low serum albumin level is the most sensitive single nutrition-related variable in the prediction of complications and length of stay.

Another important variable in a different clinical domain that was also associated with an increase in LOS was Serum Creatinine (SCr). Chertow et al. [15] evaluated the marginal effects of acute kidney injury (AKI) on mortality, LOS, and costs. Changes in serum creatinine (SCr) was used as a determinant for adverse outcomes. Results show that AKI was consistently associated with an independent increase in LOS. Larger increases in SCr were associated with longer relative increases in hospital LOS.

B. LOS in Critical Care

There are significant potential benefits from quantification and optimisation of LOS in critical care: specifically, these relate to cost containment and clinical quality. The provision of critical care is of necessity expensive, deploying complex interventions and requiring a high intensity of clinician input to a relatively small group of patients. Greater LOS requires more critical care resource and greater cost. As critical care facilities experience increasing pressure and economic resources are more constrained, the priority given to improvements in the timeliness and efficiency of critical care, is rising [62]. Clinical quality in the critical care unit may also be affected by extended LOS. Prolonged LOS gives rise to capacity pressure; this may lead to the cancellation of elective surgery, which is both costly and harmful; it may increase the pressure to decline or delay emergency admission, which could potentially have an adverse effect upon outcome; it may dilute the attention given to the most seriously sick individuals [4].

The critical care unit is also an environment which is well suited to exploiting data for mathematical modelling and prediction, both because of analytical experience and data availability: there exist well developed methodologies for performance benchmarking; the increasing use of electronic Clinical Information Systems, means that computer analy- sis can now be performed directly on the patient record, rather than after specific-to-purpose hand data extraction; the physiological and laboratory data sets are relatively large by comparison with other patient groups.

In the England, Wales and Northern Ireland, the bench- marking of critical care unit performance is conducted by the Intensive Care National Audit and Research Centre (IC- NARC), by means of its Case Mix Programme [34]. The CMP uses rigorous methods to ensure data are complete, valid and reliable [33], [61]; admissions are scored for severity using an in house scoring system and also the APACHE II model, and then a predicted hospital mortality for admissions is calculated. Comparison is made with actual mortality and a Standardised Mortality Ratio is generated quarterly [33], [61]. A Another example of a non-commercial database of this kind is that held by the Australia and New Zealand Intensive Care Society, which contains data on over 900,000 ICU stays [71].

Several research groups have investigated LOS in the ICU as it has been felt to be a suitable target for improvement [79], [80]. LOS has been linked to mortality; research in 2006 showed significantly greater ICU, hospital and long-term mortality in patients with an ICU stay longer than than 3 days, in comparison with those who have a stay of 3 days or less. Others have sought to develop models which predict LOS; Buchman et al [61] predicted chronicity in a surgical intensive care unit by classifying patients LOS in accordance with a seven day norm. Levin et al [49] developed a model to produce real-time, updated forecasts of patients intensive care LOS using naturally generated provider orders. The model was designed to be integrated within a computerised decision sup- port system to improve patient flow management. The study compared the predicted LOS to the actual LOS based on: fixed variables, such as age, source of admission and readmission status; temporal variables, such as current LOS, day of the week, time of the day; and order-based predictor variables grouped by medication, ventilation, laboratory, diet, activity, foreign body and extra-corporeal membrane oxygenation.

LOS prediction would help with capacity planning. At present, LOS prediction tools are not used in mainstream critical care practice. Surges in demand are managed reac- tively, requiring considerable staffing flexibility and variability in the balance between demand and capacity. It is possible that accurate prediction of LOS would help to align these quantities in critical care, and improve resource allocation, in particular staffing resource. According to Celi et al [4], healthcare delivery has worked as well as it has to date because clinicians are bright, hard-working, and well-intentioned, not because systems are well designed nor data systematically harnessed.

It follows that the presence of complete, highly detailed critical care databases is essential, if the potential benefits of modelling and prediction are to be fully realised [26], [80]. Several commercial ICU databases have been developed, archiving patient demographics and aggregating information such as underlying disease, severity of illness, and hospital- specific information such as LOS, mortality and readmis- sion. For example, among the commercial ICU databases is APACHE Outcomes, created at Cerner by merging APACHE [2] with Project IMPACT [18], and includes data from about 150,000 ICU stays since 2010. The commercial Philips eICU, a telemedicine intensive care support provider, archives data from participating ICUs; Philips eICU is estimated to maintain a database of over 1.5 million ICU stays, and is adding 400,000 patient records per year from over 180 subscribing hospitals in the US. More ambitious still is the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II database established in October 2003. Developed by an interdisciplinary team from academia (MIT), industry (Philips Medical Systems) and clinical medicine (Beth Israel Deaconess Medical Center), the database incorporates two different types of medical data: clinical data is stored in a relational database and bedside monitoring is stored in flat binary files. There are over 25,000 patients in the MIMIC II relational database, which permits the systematic capture, analysis and integration of information contained within the massive quantity of data generated by each critical care admission. Clearly this kind of data set could be used to investigate LOS; it is not difficult to see that it could also be exploited for broader research activity.

3. Analytical Methods to LOS Prediction

This section explores the methods used in the field of calculating and predicting patient LOS. After surveying the previous literature, LOS prediction methods were categorized into 4 subgroups as shown in Figure III. A classification of the reviewed papers based on LOS prediction methods is shown in table I.

A. Arithmetic and Statistical Approaches to LOS Prediction

Despite the complex nature of the metric LOS, simple arithmetic methods still exist for the calculation of LOS [31]. Arithmetic methods usually compute the average length of stay or the median as shown. However, this is a very simple way to measure length of stay as it assumes that LOS is normally distributed, however, typically LOS has an exponential distri- bution. Also Vasilakis et al. 2003 [56] illustrates how average LOS can be a misleading measure; the research proposes alternative statistical techniques- survival analysis, on stroke patients, aging 65 years and over. Survival Analysis [19] is a branch of statistics that typically uses LOS data to study the effect of different patient attributes on survival time.

In addition, Figure 1 highlights a special type of statistical method which includes the analysis of covariates - Regression Analysis. Covariates are defined in the context of LOS as the patient's characteristics and external factors which possibly predict LOS. Within this type are found linear regression and logistic regression, which is a special case of survival models [31]. The models developed often include the patient's diag- noses, procedures, gender and age [24], [27]. Moreover, Freitas et al. [24] used regression models to examine the association of some administrative variables from inpatient episodes in public acute care hospitals in the Portuguese National Health Service with high LOS outliers. The variables include year of discharge, comorbidities, age, adjacent DRG complexity (A-DRG), readmission, admission and DRG type, discharge status, distance from residence to hospital and hospital type. Results show that age, type of admission and hospital type were significantly associated with high LOS outliers.

A hospital is a complex stochastic system, therefore simple deterministic approaches for planning and managing such a system is considered inadequate to provide a complete and accurate analysis [10], [42], [58]; also the resulting models which are mostly based on simple rules modelled with re- gression trees, are usually further adjusted manually according to medical knowledge, decreasing the predictive accuracy of successive models [22] [46] [45] [47]. Grubinger et al. [1] argue that any minor change in the data of such simple models can lead to a completely different tree, although all of these trees can be statistically accurate. As a result, the work presented in their research used the bootstrap-based model method bumping [11] to build diverse regression tree models through systematic re-sampling (uniform randomness) of the data. Bootstrap methods are most commonly based on the idea of combining and averaging models to reduce prediction error. Examples of such methods include Bagging [11], Boosting [25] and Random Forests [12].

A data-driven approach [7], [76], [81], [82], which will be discussed thoroughly in subsection B, can be used to predict which patients seem likely to experience an extended LOS by analyzing survival data using decision trees (also called survival trees), artificial neural networks, ensemble methods...etc. Usually these approaches are used to predict categorical survival outcomes (dead or alive) for a given set of patient attributes, or used to measure patient length of stay above or below a certain threshold.

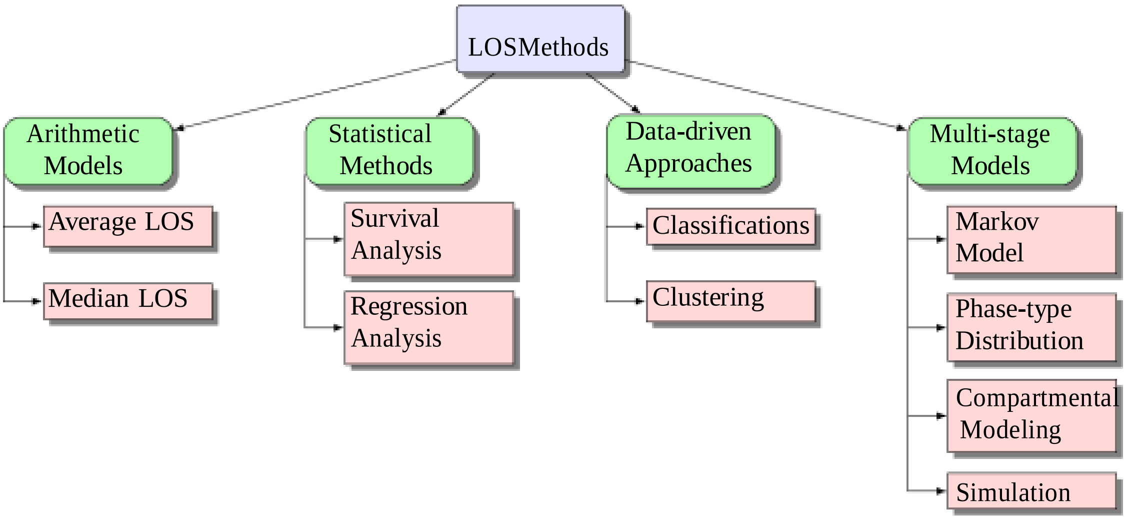

Fig. 1. The classification of the LOS prediction methods.

Table 1 :A SUMMARY OF RESEARCH PAPERS GROUPED BY ANALYTICAL METHODS TO LOS PREDICTION

Method	references	applications	Common evaluation methods
Data-driven & Data Min-Ing	[36], [1], [27], [13], [28], [62], [44], [32], [52]	Stroke unit, Intensive Medicine, Appen-dectomy, General	Sensitivity, Specificity, Harmonic mea-sure (F-measure, precision and recall average) and RBF kernel
Statistical methods	[24], [49], [56], [35], [15], [39], [16]	Acute care, intinsive care, stroke pa-tients, ICU cardiac surgery, nutrition and acute kidney	Mean, median, std. deviation, skewness, kurtosis, min-max, 25th percentile, 75th percentile, confidence interval and area under the receiver-operating character-istic curve (AUROC).
Multi-stage methods	[22], [30] and [83]	healthcare of elderly patients, Obstet-ric Unit (pregnancy and childbirth) and Emergency Departments	Sensitivity analysis, simulation mod-els, generalized Erlang, hyperexponen-tial and Coxian.

In contrast with these data-driven approaches listed above [7], [76], [81], [82], Caetano et al. [13] do not perform a classification task to LOS, instead a more information pure regression approach is adopted which predicts the actual number of LOS days and not classes. The study describes 14 input covariates to the LOS target variable. Six regression tech- niques were tested and compared: Average Prediction (AP), Multiple Regression (MP), Decision Trees (DT), Artificial Neural Network (ANN) ensemble, Support Vector Machines (SVM) and Random Forests (RF). The best results were obtained by the Random Forest model to reveal high impact of inpatient clinical process attributes, instead of the patient's characteristics. Effective predictions can aid healthcare insti- tutions and clinicians to improve their decisions about patient managements and resource allocations [30], [32], [52], [66].

Despite such attempts, Marshall et al. and Garg et al. [28], [57] argue that data-driven methods among the other statistical models fail to address the inherent uncertainty, complexity and heterogeneity in health processes. To address such issues, a more reliable way is to model patient flow as it presents the temporal dimension as well as the structural dimension
of the system [57].Numerous probabilistic models have been proposed to address the issue of LOS, namely Markov models, phase-type distributions, conditional phase-type distributions, compartmental and simulation modeling [22], [27], [28], [30], [77]. Such models may be used for planning health services for both acute and chronic patients. These

IJASRM International Journal of Advanced Scientific Research and Management, Vol. 1 Issue 8, August 2016.
www.ijasrm.com
ISSN 2455-6378

models will be discussed thoroughly in subsections C, D and E.

B. A Data-Driven Approach to LOS Prediction

Whereas most previous research examines LOS numerically [8], [9], [40], several studies take a data-drive approach to LOS prediction. A data-driven approach refers to a predictive model that is based on data-mining techniques, such as classification, clustering...etc. Such techniques are used to discover useful patterns in large datasets by showing novel and interesting relationships among data variables. Data Mining techniques facilitate the creation of knowledge and support clinical decision making, in what is known as medical data mining [69], [70].

The data-driven approach classification is used to generate early alerts with respect to a target LOS range for a specific diagnosis related group (DRG). For example, Buchman et al. [72] predict chronicity in a surgical intensive care unit by classifying patients LOS in accordance with a recommended seven-day norm. In response to the need for effective resource planning and cost containment, Mobley et al. [8] predict the LOS of patients receiving post-coronary care over the range of 120 days. Frye et al. [43] use a technique to predict whether the LOS of patients suffering from burns will fall within a one-week period. Cheng et al. in 2009 [36] introduce a study that examines the LOS management of appendectomy patients by building and empirically evaluating an automatic prediction system to identify those patients whose LOS will likely exceed the recommended five-day period. Hachesu et al. [32] apply three classification algorithms namely, decision tree, support vector machines (SVM), and artificial neural network (ANN) to draw an accurate model to predict the LOS of heart patients. Thirty six input variables were used to predict the target variable, LOS. The findings demonstrated that the SVM was the best fit. There was a significant tendency for LOS to be longer in patients with lung or respiratory disorders and high blood pressure. One of the interesting findings was that most single patients (64.3%) had a LOS 5 days, whereas 41.2 % of married patients had an LOS greater than 10 days. The most significant variables affecting LOS were drug categories, such as nitrates and anticoagulants as well as coronary artery disease (CAD) diagnosis. Co-morbidity is also a strong predictor of prolonged LOS. Gender was significant in predicting LOS since men had longer LOS than women. Age played a notable role as well since analysis revealed that patients aged less than 50 and greater than or equal 80 statistically had increased mean LOS.

Moreover, Rowan et al. [67] implemented a software pack- age demonstrating that artificial neural networks (ANNs) could be used as an effective LOS stratification instrument in postoperative cardiac patients. In [6], Azari et al. propose an approach for predicting hospital length of stay using a multi-tiered data mining approach. They form training sets, using groups of similar claims identified by k-means clustering and perform classification using ten different classifiers. They consistently found that using clustering as a precursor to form the training set gives better prediction results as compared to non-clustering based training sets. Binning the LOS to three groups of short, medium and long stays, their method identifies patients who need aggressive or moderate early interventions to prevent prolonged stays.

Liu et al. [52] applied two classifiers: decision tree C4.5 & its successor R-C4.5s, Naive Bayesian classifier (NBC) and its successor NBCs to a geriatric hospital dataset, called Clinics Dataset, containing 4722 patient records including patient demographic details, admission reasons, discharge details, out- come and LOS, to predict inpatient LOS for long stay patients. According to [51], C4.5 is one of the classifiers, which has the best combinations in terms of error rate and speed. Also, R- C4.5s combines branches with little classification contribution and thus resulted in building more robust and smaller trees [87].In addition, NBC is robust and insensitive to missing data as stated in the work of Liu et al. [52].

In addition, phase-type survival trees and mixed distribution survival trees are used to cluster stroke-related patients into clinically meaningful groups with respect to length of stay where partitioning is based on covariates, such as gender, age at time of admission, primary diagnosis code, treatment outcome and discharge destination [27], [28]. Moreover, Kudyba et al. in 2010 [44] utilize the method of neural networks to analyze data describing inpatient cases to examine the effect of the independent variables of patient demographics, primary payer, admission and discharge dates, physician specialty, and detailed radiology procedural variables (including the sum of radiology hours) on the dependent variable of length of stay excess per patient case for a major New Jersey based healthcare provider. Also, artificial neural networks, decision trees and ensemble methods are used in developing an intelligent decision support system- INTCare for intensive medicine in the ICU of the Hospital Santo Antonio (HAS) in Porto, Portugal [62]. In addition, the bootstrap-based method bumping is used by [1] to build diverse and more accurate regression tree models for DRG systems in Austria. Eight datasets are used consisting of patient's main diagnosis, secondary diagnoses, procedures, number of diagnoses, number of procedures, gender and age as well as patients' LOS.

C. Markov Model and Phase-type Distributions

Markov and semi-Markov chain models are models that assume sub-groups of patients are homogeneous and events occur at equally spaced intervals of time; queueing models and deterministic models of the transition of patients between states. These techniques are useful for examining patient flow in large population groups where Markov assumptions can be made [20]. Phase-type distributions describe the time to absorption of a finite Markov chain in continuous time when there is a single absorbing state and the stochastic process starts in a transient state [31].

The first probabilistic approach describes a special type of Markov model known as the Coxian phase-type distribution and its further development into the conditional phase-type distribution. The Coxian phase-type distribution, allows the representation of the continuous duration of stay of patients in hospital as a series of sequential phases which the patients progress through until they leave the hospital completely [23]. It is possible to expand the theory of Coxian phase-type distributions to include a network of additional interrelated variables (such as patient characteristics) that may interact to influence patient LOS Conditional phase-type distribution. This approach allows the incorporation of discrete and contin- uous variables representing causality. Marshall et al. [54] [53] [57] [29] uses conditional phase-type distribution to model the LOS of elderly patients in hospital. The approach illustrates data on hospital processes for a number of geriatric patients along with personal details, admissions reasons, dependency levels and destination (the causal network). The final model represents patient LOS in terms of five of the most significant patient variables in the dataset, namely patient age, gender, admission method into hospital, Barthel grade (dependency score) and destination on departure from hospital.

D. Compartmental modeling

The second general approach described is the compartmen- tal model. Compartmental modeling of patient flow is a type of mathematical model used for describing the way patients are transmitted among the compartments of a healthcare system. Each compartment is assumed to be a homogeneous entity within which the entities being modelled are equivalent. For instance, in a pharmaceutical model, the compartments may represent different sections of a body within which the concentration of a drug is assumed to be uniformly equal. Another example, in a healthcare facility, the compartments may represent the different stages that patient goes through- acute, long-stay and death.

Haigeng Xie et al. [86] present a model-based approach to extract from an administrative social care dataset, high- level length of stay patterns of residents in long-term care (LTC). A continuous-time Markov model, a residents stay in both residence care (RC) and nursing care (NC) is modeled as consisting of a short-stay and a long-stay phase, was used to show the flow of residents within and between RC and NC, as well as discharge from RC and NC. The model has been extended to incorporate residents features, such as gender. The final model showed that gender has a significant influence on transition rates.

Irvine et al. [38] describes the development of a two-stage continuous-time Markov model that describes the movement of patients through geriatric hospitals. Patients are initially ad- mitted to the acute state from which they transfer to the long- stay state or leave the hospital completely through discharge or death state. McClean et al. [59] extends the stochastic Markov model presented in [38] to a three-stage one and attaches different costs to each stage thus taking cost into account. Taylor et al. [73] uses a continuous time Markov model and applies it to the case of a four compartmental model, where the four stages are acute, long-stay, community, and dead. The model estimates the expected number of patients at any time t in each stage. Taylor et al. [74] extends these models to contain six stages. Garg et al. 2012 [28] proposed a novel distribution, multi-absorbing state phase-type distribution, as a generalization of the single absorbing state Coxian phase-type distribution for representing a Markov process having more than one absorbing states. The approach effectively forecasts the bed requirements in a care unit considering the effect of several factors, such as patient demography age and gender, as well as treatment outcome based on diagnosis and patients expected destination after discharge, which may also affect a patients LOS in hospital.

E. Simulation modeling

Simulation-based models simulate scenarios which replicate real life in an attempt to understand the complex health processes and their interactions [28]. Vasilakis et al. 2003 [56] illustrate how average LOS can be a misleading measure. The research proposes alternative statistical techniques, such as survival analysis, the application of mixed exponential and phase-type distributions demonstrated in two dynamic models of patient flow compartmental model (small, medium and long stay) and discrete event simulation model, introducing capacity constraints in the various stages of the model, such as bed blockage and refuse-admission rates.

In addition, Griffin et al. 2011 [30], developed a simulation model using a path-based approach for an obstetric unit to study tradeoffs in blocking and system efficiency. The model focuses on patient flow, considering patient classification, blocking effects, time dependent arrival and departure patterns, and statistically supported distributions for LOS. Moreover, the study conducted by Wang et al. [83] in the Emergency department at a community hospital, Entral Baptist Hospital in Lexington, KY, uses a discrete-event simulation model to eval- uate patient outcome, identify the impact of critical resources and procedures, conduct "What if" analysis for various staffing and operational scenarios, and provide recommendations for hospital management.

Discrete event simulation models allow patients to have individual attributes and to interact with resource provision but they are more time consuming to test and run. They are particularly suitable for models of systems of patient care where the constraints on resource availability are important. They may also be used on unconstrained population models with several thousands of patients. A significant development in simulation is the facility to model entities so that they can participate in more than one activity simultaneously and interrupt each other. The credibility of any model is dependent on reliable data which are not always readily available in the British health service [20].

4. Conclusion

This paper has presented a comprehensive review of the applications of LOS in acute medicine and critical care. An introduction to LOS theory and the main drivers behind the interest in such research topic/metric was given in the opening section. In addition, several applications in LOS was demonstrated both in acute medicine and in the intensive care unit environment in particular, highlighting the challenges facing both physicians and information engineers today. An analysis of the various methods for LOS prediction were presented and compared. Moreover, the paper provides a classification for some state-of-the-art literature based on each paper's analytical method utilised to predict LOS. The four main categories include: (1) Arithmetic methods, (2) Statistical methods, (3) data-driven methods and (4) multi-stage methods. The classification presents a brief summary of the analytical method, dataset and the evaluation method utilised. Given that LOS is a relatively complex metric as it is influenced by various external uncontrollable factors, there is no one good- for-all technique that serves its prediction. At present, in most cases several algorithms are tested, tweaked based on some domain knowledge or some performance criteria to enhance the accuracy of prediction. However, it is clear that much research remains to be done. Also the presence of databases allows for investigation.

References

[1] Grubinger, Thomas, Kobel, Conrad and Pfeiffer. Karl-Peter Regression tree construction by bootstrap: Model search for drg-systems applied to austrian health-data), MC Medical Informatics and Decision Making, 10(1),9,(2010).

[2] RR Abbott, M Setter, S Chan, and K Choi. Apache ii: prediction of outcome of 451 icu oncology admissions in a community hospital. Annals of Oncology, 2(8):571–574, 1991.

[3] Burritt MR¿ Anderson CG, Moxness K, Meister, J. The sensitivity and specificity of nutrition-related variables in relationship to the duration of hospital stay and the rate of complications. In Mayo Clin Proc., pages477–485, 1984.

[4] Leo Anthony Celi, Roger G Mark, David J Stone, and Robert A Montgomery. big data in the intensive care unit. closing the data loop. American journal of respiratory and critical care medicine,187(11):1157–1160, 2013.

[5] P Appelros. Prediction of length of stay for stroke patients. ActaNeurologica Scandinavica, 116(1):15–19, 2007.

[6] Ali Azari, Vandana P Janeja, and Alex Mohseni. Healthcare data mining: Predicting hospital length of stay (phlos). International Journal of Knowledge Discovery in Bioinformatics (IJKDB), 3(3):44–66, 2012.

[7] Ali Azari, Vandana P Janeja, and Alex Mohseni. Predicting hospital length of stay (phlos): A multi-tiered data mining approach. In Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on, pages 17–24. IEEE, 2012.

[8] L. Davidson B. A. Mobley, R. Leasure. Artificial nerual network predictions of lengths of stay in a post-coronary care unit. Heart Lung,24(3):251–256, 1995.

[9] K. H. Hennecke B. Zernikow, K. Holtmannspotter, E. Michel, F.Hornschuh, K. Groote. Predictin glength

of stay in preterm neonates.Eur. J. Pediatr., 158(1):59–62, 1999.

[10] Christopher A Bain, Peter G Taylor, Geoff McDonnell, and Andrew Georgiou. Myths of ideal hospital occupancy. Medical Journal of Australia, 193(5):311, 2010.

[11] Leo Breiman. Bagging predictors. Machine learning, 24(2):123–140,1996.

[12] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[13] Nuno Caetano, Raul Laureano, and Paulo Cortez. A data-driven approach to predict hospital length of stay: a portuguese case study.2014.

[14] Ku-Chou Chang, Mei-Chiun Tseng, Hsu-Huei Weng, Yin-Hui Lin, Chia- Wei Liou, and Teng-Yeow Tan. Prediction of length of stay of first-evr ischemic stroke. Stroke, 33(11):2670–2674, 2002.

[15] Glenn M Chertow, Elisabeth Burdick, Melissa Honour, Joseph V Bon- ventre, and David W Bates. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. Journal of the American Society of Nephrology : JASN, 16(11):3365–70, November 2005.

[16] Cinda S Chima, Kathy Barco, MARCI LA DEWITT, Michelle Maeda, J Carlos Teran, and Kevin D Mullen. Relationship of nutritional status to length of stay, hospital costs, and discharge status of patients hospitalized in the medicine service. Journal of the American Dietetic Association,97(9):975–978, 1997.

[17] Adelina Comas-Herrera, Raphael Wittenberg, and Linda Pickard. Euro- pean study of long-term care expenditure. Long-Term Care Expenditure in an Ageing Society, page 111, 2003.

[18] Suzanne F Cook, Wendy A Visscher, Connie L Hobbs, Rick L Williams, Project IMPACT Clinical Implementation Committee, et al. Project impact: results from a pilot validity study of a new observational database. Critical care medicine, 30(12):2765–2770, 2002.

[19] D. Oakes D. R. Cox. Analysis of Survival Data. U.K.: Chapman & HAll, 1984.

[20] R Davies and HTO Davies. Modelling patient flows and resource provision in health systems. Omega, 22(2):123–131, 1994.

[21] Hoefer M. Epstein AM., Red JL. The relation of body weigth to length of stay adn charges for patients undergoing elective surgery: as study of two procedures. Am J Public Health, pages 993–997, 1987.

[22] Mark Fackrell. Modelling healthcare systems with phase-type distribu- tions. Health Care Management Science, 12(1):11–26, June 2008.

[23] MJ Faddy and SI McClean. Analysing data on lengths of stay of hospital patients using phase-type distributions. Applied Stochastic Models in Business and Industry, 15(4):311–317, 1999.

[24] Alberto Freitas, Tiago Silva-Costa, Fernando Lopes, Isabel Garcia- Lema, Armando Teixeira-Pinto, Pavel Brazdil, and Altamiro Costa- Pereira. Factors influencing hospital high length of stay outliers. BMC health services research, 12(1):265, January 2012.

[25] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232, 2001.

[26] Reed M Gardner, William L Hawley, Thomas D East, Thomas A Oniki, and Hsueh-Fen W Young. Real time data acquisition: recommendations for the medical information bus (mib). International journal of clinical monitoring and computing, 8(4):251–258, 1991.

[27] Lalit Garg, Sally Mcclean, Brian J Meenan, and Peter Millard. Phase- Type Survival Trees and Mixed Distribution Survival Trees for Cluster- ing Patients Hospital Length of Stay. 22(1):57–72, 2011.

[28] Lalit Garg, Sally I McClean, Maria Barton, Brian J Meenan, and Ken Fullerton. Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 42(6):1332–1345, 2012.

[29] NMS Golu¨ ke, CJA Huibers, SC Stalpers, DG Taekema, SE Vermeer, and PAF Jansen. An observational, retrospective study of the length of stay, and its influencing factors, among elderly patients at the emergency department. European Geriatric Medicine, 2015.

[30] Jacqueline Griffin, Shuangjun Xia, Siyang Peng, and Pinar Keskinocak.Improving patient flow in an obstetric unit. Health care management science, 15(1):1–14, March 2012.

[31] M Guzman Castillo. Modelling patient length of stay in public hospitals in Mexico. PhD thesis, University of Southampton, 2012.

[32] Peyman Rezaei Hachesu, Maryam Ahmadi, Somayyeh Alizadeh, and Farahnaz Sadoughi. Use of data mining

techniques to determine and predict length of stay of cardiac patients. Healthcare informatics research, 19(2):121–129, 2013.

[33] David A Harrison, Anthony R Brady, and Kathy Rowan. Case mix, outcome and length of stay for admissions to adult, general critical care units in england, wales and northern ireland: the intensive care national audit & research centre case mix programme database. Critical Care,9(Suppl 3):S1, 2004.

[34] David A Harrison, Gareth J Parry, James R Carpenter, Alasdair Short, and Kathy Rowan. A new risk prediction model for critical care: The intensive care national audit & research centre (icnarc) model*. Critical care medicine, 35(4):1091–1098, 2007.

[35] Ortrud Vargas Hein, Ju¨ rgen Birnbaum, Klaus Wernecke, Michael Eng- land, Wolfgang Konertz, and Claudia Spies. Prolonged intensive care unit stay in cardiac surgery: risk factors and long-term-survival. The Annals of thoracic surgery, 81(3):880–5, March 2006.

[36] P.J.-H. Hu. A Data-Driven Approach to Manage the Length of Stay for Appendectomy Patients. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 39(6):1339–1347, November 2009.

[37] Dale A Huntley, Dong Won Cho, Jane Christman, and John G Csernan- sky. Predicting length of stay in an acute psychiatric hospital. Psychiatric Services, 2014.

[38] P. H. Irvine, V., McClean, S. I., Millar. Stochastic models for geriatric inpatient behavior. IMA JOurnal of Mathematics Applied in Medicine and Biology, pages 207–216, 1994.

[39] M Isabel T. D. Correia. The impact of malnutrition on morbidity, mor- tality, length of hospital stay and costs evaluated through a multivariate model analysis. Clinical Nutrition, 22(3):235–239, June 2003.

[40] J. Hershberger J. Grigsby, R. Kooken. Simulated neural networks to pre- dict outcomes, costs, and length of stay among orthopedic rehabilitation patients. ARch. Phys. Med. Rehabil., 75(10):1077–1081, 1994.

[41] Xiaochun Jiang, Xiuli Qu, and Lauren Davis. Using data mining to analyze patient discharge data for an urban hospital. In DMIN, pages139–144, 2010.

[42] Rodney P Jones. Myths of ideal hospital size. Medical Journal ofAustralia, 193(5):298–300, 2010.

[43] A.Luterman K. E. Frye, S. D. Izenberg, M. D. Williams. Simulated biologic intelligence used ot predict length of stay and survival of bums. J. Burn Care Rehabil., 17:540–546, 1996.

[44] Stephan Kudyba and Thomas Gregorio. Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics. Health informatics journal, 16(4):235–45, December 2010.

[45] P. H. Millard L. Garg, S. I. McClean, B. J. Meenan. Non-homogeneous Markov models for sequential pattern mining of heatlhcare data. IMA J. Management Math., 20(4):327–344, 2009.

[46] P. H. Millard L. Garg, S. I. McClean, B. J. Meenan. A non-homogeneous discrete time Markov model for admission scheduling and resource planning in a care system. Health Care Management Science, 13(2):155–169, 2010.

[47] K. Fullerton L. Garg, S. I. McClean, M. Barton, B. J. Meenan.Forecasting hosptial bed requirements and cost of care using phase type survival trees. In IEEE Conference Intell, pages 185–190, 2010.

[48] Stanley Lemeshow, Daniel Teres, Janelle Klar, Jill Spitz Avrunin,Stephen H Gehlbach, and John Rapoport. Mortality probability models (mpm ii) based on an international cohort of intensive care unit patients.Jama, 270(20):2478–2486, 1993.

[49] Scott R Levin, Eric T Harley, James C Fackler, Christoph U Lehmann, Jason W Custer, Daniel France, and Scott L Zeger. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. Critical care medicine, 40(11):3058–64, November2012.

[50] Apiradee Lim and Phattrawan Tongkumchum. Methods for analyzing hospital length of stay with application to inpatients dying in southern thailand. Global Journal of Health Science, 1(1):p27, 2009.

[51] Tjen-Sien Lim, Wei-Yin Loh, and Yu-Shan Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine learning, 40(3):203–228,2000.

[52] Peng Liu, Lei Lei, Junjie Yin, Wei Zhang, Wu Naijun, and Elia El-Darzi.Healthcare data mining: predicting inpatient length of stay. 2006.

[53] McClean S. I. Marshal, A. H. Conditional phase-type distributions for modelling patient length of stay in hospital. International Transactions in Operational Research, pages 567–576, 2003.

[54] P. H. Marshal, A. H., McClean, S. I., Shapcott, C. M., Millard. Modeling patient duration of stay to facilitate resource management of geriatric hospitals. Helath care management science, pages 313–319, 2002.

[55] Adele Marshall, Christos Vasilakis, and Elia El-Darzi. Length of stay- based patient flow models: recent developments and future directions. Health Care Management Science, 8(3):213–220, 2005.

[56] Adele Marshall, Christos Vasilakis, and Elia El-Darzi. Length of stay- based patient flow models: recent developments and future directions. Health Care Management Science, 8(3):213–220, 2005.

[57] P. H. Marshall, A. H., McClean, S. I., Shapcott, C. M., Hastie, I. R., Millard. Developing a Bayesian belief network for th emanagement of geriatric hospital care. Healthcare Management Science, pages 25–30,2001.

[58] Sally M McCarthy. Hospital capacity: what is the measure and what is the goal? The Medical Journal of Australia, 193(5):252–253, 2010.

[59] P. H. McClean, S. I., McAlea, B., Millard. Using a Markov reward model to estimate spend-down costs for a geriatric department. Journal of Operational Research Society, pages 1021–1025, 1998.

[60] R McMullan, B Silke, K Bennett, and S Callachand. Resource utilisa- tion, length of hospital stay, and pattern of investigation during acute medical hospital admission. Postgraduate medical journal, 80(939):23–26, 2004.

[61] JP Nolan, SR Laver, CA Welch, DA Harrison, V Gupta, and K Rowan.Outcome following admission to uk intensive care unit after cardiac arrest: a secondary analysis of the icnarc case mix programme database*. Anaesthesia, 62(12):1207–1216, 2007.

[62] Filipe Portela, Manuel Filipe Santos, lvaro Silva, Fernando Rua, Antnio Abelha, and Jos Machado. Adoption of pervasive intelligent information systems in intensive medicine. Procedia Technology, 9:1022 – 1032,2013. {CENTERIS} 2013 - Conference on {ENTERprise} Informa- tion Systems / ProjMAN 2013 - International Conference on Project MANagement/ {HCIST} 2013 - International Conference on Health and Social Care Information Systems and Technologies.

[63] Adam Roberts, Louise Marshall, and Anita Charlesworth. A decade of austerity? the funding pressures facing the nhs from 2010/11 to 2021/22. London: Nuffield Trust, 2012.

[64] Gordon H Robinson, Louis E Davis, and Richard P Leifer. Prediction of hospital length of stay. Health services research, 1(3):287, 1966.

[65] Levine GM. Robinson G, Goldstein M. Impact of nutritional status onDRG length of stay. In JPEN, pages 49–51, 1987.

[66] Michael Rowan, Thomas Ryan, Francis Hegarty, and Neil OHare. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. Artificial Intelligence in Medicine, 40(3):211–221, 2007.

[67] Michael Rowan, Thomas Ryan, Francis Hegarty, and Neil OHare. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. Artificial Intelligence in Medicine, 40(3):211 – 221, 2007.

[68] P. D. Clayton S. Shea, R. V. Sideli, W. DuMouchel, G. Pulver, R.R. Arons. Computer-generated informational messages directed to physicians: Effect on length of hospital stay. J. Amer. Med. Info. Assoc.,2(1):58–64, 1995.

[69] A´ lvaro Silva, Paulo Cortez, Manuel Filipe Santos, Lopes Gomesand Jose´Neves. Mortality assessment in intensive care units via adverse events using artificial neural networks. Artificial Intelligencein Medicine, 36(3):223–234, 2006

[70] A´ lvaro Silva, Paulo Cortez, Manuel Filipe Santos, Lopes Gomes, and Jose´ Neves. Rating organ failure via adverse events using data mining in the intensive care unit. Artificial intelligence in medicine, 43(3):179–193, 2008.

[71] Peter J Stow, Graeme K Hart, Tracey Higlett, Carol George, Robert Herkes, David McWilliam, Rinaldo Bellomo, ANZICS Database Man- agement Committee, et al. Development and implementation of a high- quality clinical database: the australian and new zealand intensive care society adult patient database. Journal of critical care, 21(2):133–141,2006.

[72] M. J. Siegforth T. G. Buchman, K. L. Kubos, A. J. Seidler. A comparison of statistical and connectionist models for th

eprediction of chronicity in a surgical intensive care unit. Crit. Care Med., 22(5):750–762, 1994.

[73] P. H. Taylor, G. J., McClean, S. I., Millar. Continuous-time Markov model for geriatric patient behaviour. Applied Stochastic Models and Data Analysis, pages 315–323, 1998.

[74] P. H. Taylor, G. J., McClean, S. I., Millard. Stochastic Models of Geriatric Patient Bed Occupancy Behaviour. Journal of the royal statistical society series, pages 1–10, 2000.

[75] R Vahidi, H Kushavar, and R Khodayari. Factors affecting coronary artery patients hospital length of stay of tabriz madani hospital 2005-2006. Journal of Health Administration, 9(25):63–68, 2006.

[76] V. Vapnik. The Nature of Statistical Learning Theory. 1995.

[77] Christos Vasilakis and Adele H Marshall. Modelling nationwide hospital length of stay: opening the black box**. Journal of the Operational Research Society, 56(7):862–869, 2005.

[78] J-L Vincent, Rui Moreno, Jukka Takala, Sheila Willatts, ArnaldoDe Mendonc¸a, Hajo Bruining, CK Reinhart, PeterM Suter, and LG Thijs The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive care medicine, 22(7):707–710, 1996.

[79] Jean-Louis Vincent. Is the current management of severe sepsis a septic shock really evidence based? PLoS medicine, 3(9):e346, 2006.

[80] Jean-Louis Vincent and Mervyn Singer. Critical care: advances and future perspectives. The Lancet, 376(9749):1354–1361, 2010.

[81] S. D. Izenberg W. E. Pofahl, S. M. Walczak, E. Rhone. Use of an artificial neural network to predict length of stay in acute pancreatitis. Mer. Surgeon, 64(9):868–872, 1998.

[82] H. R. Park W. O. Kim, H. K. Kil, J. W. Kang. Prediction on lengths of stay in the postanesthesia care unit following general anesthesia: Pre- liminary study of the neural network and logistic regression modeling. J. Korean Med. Sci., 15(1):25–30, 2000.

[83] Junwen Wang, Jingshan Li, Kathy Tussey, and Kay Ross. Reducing length of stay in emergency department: A simulation study at a community hospital. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 42(6):1314–1322, 2012.

[84] Y Claire Wang, Klim McPherson, Tim Marsh, Steven L Gortmaker, and Martin Brown. Health and economic burden of the projected obesity trends in the usa and the uk. The Lancet, 378(9793):815–825, 2011.

[85] Lundhom K. Warnold I. Clinical significance of preoperative nutritional status in 215 non cancer patients. In Ann Surg., pages 299–305, 1984.

[86] H. Xie, T.J. Chaussalet, and P.H. Millard. A Model-Based Approach to the Analysis of Patterns of Length of Stay in Institutional Long-Term Care. IEEE Transactions on Information Technology in Biomedicine,10(3):512–518, July 2006.

[87] Zheng Yao, Peng Liu, Lei Lei, and Junjie Yin. R-c4. 5 decision tree model and its applications to health care dataset. In Services Systems and Services Management, 2005. Proceedings of ICSSSM'05. 2005International Conference on, volume 2, pages 1099–1103. IEEE, 2005.