Modeling and Predicting Patient Length of Stay: A Survey

Aya Awad¹, Mohamed Bader-El-Den² and James McNicholas³

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-related 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 individual’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 it.

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 ofpayment 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 nutrition status in 215 non-cancer patients. The variables investigat ed 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 analysis 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 benchmarking 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 support 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 reactively, 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 readmission. 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 distribution. 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 diagnosis, 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 regression 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.
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 techniques 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 institutions 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

---

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

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

<table>
<thead>
<tr>
<th>Method</th>
<th>references</th>
<th>applications</th>
<th>Common evaluation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-driven &amp; Data Min-Ing</td>
<td>[36], [1], [27], [13], [28], [62], [44], [32], [52]</td>
<td>Stroke unit, Intensive Medicine, Appendectomy, General</td>
<td>Sensitivity, Specificity, Harmonic measure (F-measure, precision and recall average) and RBF kernel</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>[24], [49], [56], [35], [15], [39], [16]</td>
<td>Acute care, intensive care, stroke patients, ICU cardiac surgery, nutrition and acute kidney</td>
<td>Mean, median, std. deviation, skewness, kurtosis, min-max, 25th percentile, 75th percentile, confidence interval and area under the receiver-operating characteristic curve (AUROC).</td>
</tr>
<tr>
<td>Multi-stage methods</td>
<td>[22], [30] and [83]</td>
<td>healthcare of elderly patients, Obstetric Unit (pregnancy and childbirth) and Emergency Departments</td>
<td>Sensitivity analysis, simulation models, generalized Erlang, hyperexponential and Coxian.</td>
</tr>
</tbody>
</table>
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 continuous 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 compartmental 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 admitted 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 trade-offs 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, Enntal Baptist Hospital in Lexington, KY, uses a discrete-event simulation model to evaluate 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


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


