

Fitting Aggregated Phase-Type Distributions to the Length-of-Stay in Intra-Hospital Patient Transfers

Wanlu Gu^a, Neng Fan^{a1}, Haitao Liao^b

^a*Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ, USA*

^b*Department of Industrial Engineering, University of Arkansas, Fayetteville, AR, USA*

Abstract

The patient transfer, as a common seen and necessary healthcare procedure, plays an important role in maintaining efficient treatment and improving the quality of healthcare. Among various factors impacting and indicating the safety and efficiency of the patient transfer, the length-of-stay (LOS), which is not often studied in this field, is worth investigating. Phase-type (PH) distributions, as one of the popular methods of modeling LOS, will be integrated in an aggregated Markov chain to construct a model to describe the sequences of LOS in each hospital unit. In this paper, we model the intra-hospital transfer flow routes by fitting aggregated PH distribution and using Maximum Likelihood Estimation to estimate the parameters. Following the results of distribution fitting, the patients can be divided into different groups according to their LOS in the same unit. By analyzing each group to find out its common characteristics, intra-hospital transfer routes, admission and discharge situations, the associations among significant factors, the LOS and the treatment efficiency are evaluated.

Key words: Phase-type distribution, length-of-stay, intra-hospital transfer, Markov chain

1. Introduction

The patient transfer is defined as transfer of patient within the same facility for any diagnostic procedure or transfer to another facility with more advanced care. It includes types of intra- and inter-hospital transfer, which aim to maintain the continuity of medical care and to improve the existing management of the patient [1]. However, it also brings risks to the patients during transfers. There exist various

¹Corresponding Author. Tel.: +1 (520) 621 6557. Email: nfan@arizona.edu

guidelines for safe patient transfer and all state same key elements for both intra- and inter-hospital transfer [2, 3]. Considering the current practice of transfers, a statistical modeling of the patient transfer is still necessary to optimize the management and to reduce the unexpected delay. In this paper, we focus on the transfer of patients between wards and departments (called “units” in the following) within a hospital.

1.1. Patient transfer

Although the intra-hospital transfer is performed daily, it can put patients at risk and even have life-threatening impacts on patients. The intra-hospital transfers trigger complications and worsen outcomes [4], increase staff burden [5], and higher care costs. The unnecessary transfers, and the delay and wasting time in transfers worsen the situation. In [6–8], it is verified that the inefficient intra-hospital transfer had increased hospital length-of-stay (LOS) and patient mortality. These suggest an urgent need to identify factors associated with wasting time in transfer as well as to improve the efficiency of the intra-hospital transfer system.

Besides the key elements for intra-hospital transfers such as pre-transfer stabilization and preparation [9], accompanying the patient [10], equipment, drugs, and monitoring, optimizing the procedure and reducing waiting time to improve transfer efficiency is essential, since it is revealed only 12% efficiency in the transfer process [11]. The wasting time and delay in transfers can be caused by communication breakdowns, resources (bed, equipment, etc.) and staff assignment problem, lag in updating documentation and information.

The LOS, as an important measure of healthcare efficiency, has been studied to help improve healthcare services and resource allocation. In [12], the factors associated with LOS are investigated through analysis in terms of different units, transfer patterns, diagnosis, etc. The results show that the transfer status and transfer time are two factors correlated with LOS. Therefore, the association between LOS and transfer flow information is worth investigating in terms of hospital units.

Lots of previous works have studied on reducing LOS, delay, and overcrowding problem in the emergency departments (ED). The ED is a specialized medical treatment unit which deals with emergency medicine, the acute care of patients without the need of prior appointment. It is suggested in [13–15] that ED crowding is not an isolated phenomenon and appears to affect hospitals nationwide. Patients are evaluated and treated in the ED, and further treatments depend on their conditions. Longer LOS in ED may cause delay in the emergency treatment to other patients and compromise the quality of healthcare.

In [15], alternative units like Emergency Medicine Ward, Chest pain observation units, and Clinical Decision Units (CDU) have been set up to meet the demands of ED, and to reduce the LOS in ED. In [16], the relationship between the

ED and the CDU, the CDU admissions criteria, and decreasing LOS in both units are investigated. Expeditionary transfer of patients to the CDU may decrease the ED overcrowding, ED waiting time, and avoid the phenomenon of patients leaving the ED without being diagnosed. The CDU may also provide cost savings to both patients and hospitals.

Patients are usually transferred to a general ward (GW) when medical staff decides that they are no longer in danger and can be treated under normal observation. Different types of wards are investigated in [17, 18], and the patient satisfaction, and decreases in LOS in different wards are also compared.

The intensive care unit (ICU) is a place where critically ill patients should be admitted as soon as possible to receive specialized care. It is concluded in [19] that prolonged ED to ICU time is associated with increased hospital mortality after treatment in ICU. And longer LOS in ICU may contribute to higher mortality [20]. Studying the LOS in ICU as well as the ED-ICU transfer plays an important role in making decision on whether or not sending patients to ICU and shortening the waiting time in the transfer.

Among methods of studying LOS, the phase-type (PH) distributions gains popularity over time. The PH distributions, as a popular method applied in queueing theory, reliability theory, and healthcare, describe the absorption time of an evanescent finite-state Continuous Time Markov Chain (CTMC) [21]. Among the special structured PH distributions, the Coxian PH distributions are often used in modeling and investigating the LOS data. A lot of previous research has been conducted on the analysis of LOS of patients with various diseases [22–24], and in these papers, the patients are classified into LOS groups which correspond to the states in the Coxian PH distributions. In each LOS group, patients may share similar characteristics and require similar treatment and resource assignment. However, the LOS distributions along with the unit transfer during the hospital stay has not been widely studied yet.

1.2. Aggregated Markov process

The aggregated Markov process was motivated by the neurophysiological investigation [25, 26]. In these papers, bursts are the observable phenomenon and the time from the start of the first opening to the end of last opening in the burst is measured. The states in the system can be divided into three subsets, and the probability density function for the lifetime of sojourn in each subset is defined and evaluated. The states in these papers can be treated as the units in a hospital.

In [27, 28], the forward and backward recursions and the spectral decomposition are applied respectively to derive and simplify the likelihood function, which is the joint probability density of dwell time sequence, and its derivatives. The Maximum Likelihood Estimation (MLE) of parameters is obtained afterwards.

In [29], both expectation-maximization (EM) algorithm and Bayesian method are used to model the aggregate failure-time data, which is an issue to deal with for reliability estimation. To apply the aggregated Markov processes in healthcare area, Xie et al. [30] develop a CTMC model for the LOS of elderly patients moving within and between residential home care and nursing home care, upon which the aggregated Markov process theory and parameter estimation approach is applied. In [31], both Aggregated Markov model and Coxian PH distributions are used in modeling the elderly flow in the hospital services.

For such kind of analysis, a department or facility represents a unit with embedded states or groups in the aggregated Markov process. In each unit, a Coxian PH distribution can be used to model the LOS of patients during the stay in that unit. Most of previous works on modeling LOS in units have only focused on single unit, and the impacts of delay transfer. It lacks broad applications of aggregated Markov process to investigating transfers among multiple units and consider dwell in each unit simultaneously. A model built upon patient flow among units which can simulate the general daily intra-hospital transfers is in need.

In this paper, we apply statistical modeling to intra-hospital patient transfer information from ED to other units, such as GW and CDU. Our goal is to construct an aggregated PH distribution to model the time spent in each unit, and also assess the transition rates among different units. Patients with similar characteristics may share similar LOS sequence in a series of units, which may provide a better view of the patients arrangement at admission and avoid wasting time in transfer.

For our collected patient flow information (more details given in Section 4.1), we filter the data to only contain the records of geriatric patients being admitted into ED directly, and then being transferred to other units for further surgical procedure or to GW before discharge. The unit transfers include ED-GW, ED-CDU, ED-ICU, ED-Discharge, ICU-GW and ICU-Discharge. One needs to notice that the sojourn time within one unit follows a Coxian PH distribution, while the overall LOS before discharge follows a general PH distribution.

The remainder of this paper is organized as follows. In Section 2, we introduce the continuous time Markov process and how it is extended to the aggregated Markov process. The normal PH distributions and the aggregated PH distributions are then defined in this section. Section 3 explains the details of implementing the MLE in fitting the aggregated PH distribution. In Section 4, the process of data collections and pre-analysis are presented and followed with the model constructing, parameter estimation, and further analysis. Finally, Section 5 concludes the paper.

2. Aggregated Markov Process

2.1. Continuous time Markov process

Let $\{X(t)\}_{t \geq 0}$ be a continuous time Markov process with state space $S = \{1, 2, \dots, m, 0\}$, where $m \geq 1$ is a finite number, such that the states $1, \dots, m$ are transient states and state 0 is an absorbing state. The sojourn in each transient state i ($i = 1, \dots, m$) follows an exponential distribution with parameter λ_i . Then the transition rate from state i ($i = 1, \dots, m$) to state j ($j = 0, 1, \dots, m$) can be denoted by $\lambda_{ij} = \lambda_i p_{ij}$, where λ_i is interpreted as the average rate of moving out of state i , and p_{ij} is the probability of transferring from state i to state j with $\lambda_i > 0, p_{ij} \geq 0, \lambda_{ij} \geq 0$.

Let the initial distribution of such CTMC as $\pi = (\pi_1, \dots, \pi_m)$ with $\sum_{i=1}^m \pi_i = 1$. The transition rate matrix or infinitesimal generator matrix $\mathbf{Q}$, describing that the rate a CTMC moves between states, can be written in block-matrix form as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \mathbf{q} \\ \mathbf{0} & 0 \end{pmatrix}$$

Here,

$$\mathbf{T} = \begin{bmatrix} -\lambda_1 & \lambda_1 p_{12} & \cdots & \lambda_1 p_{1m} \\ \lambda_2 p_{21} & -\lambda_2 & \cdots & \lambda_2 p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_m p_{m1} & \lambda_m p_{m2} & \cdots & -\lambda_m \end{bmatrix}_{m \times m} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} \lambda_1 p_{1,0} \\ \lambda_2 p_{2,0} \\ \vdots \\ \lambda_m p_{m,0} \end{bmatrix}_{m \times 1}.$$

The matrix $\mathbf{T}$ describes the moving of the CTMC among transient states, and the vector $\mathbf{q}$ consists of transition rates from transient states to the absorbing state, and $\mathbf{0}$ is a $1 \times m$ vector of 0. Note that $\mathbf{q} = \mathbf{T}\mathbf{e}$ where $\mathbf{e}$ is a $m \times 1$ vector of ones, and this property will be further used in derivative correction in Section 3.3.

We use a random variable Y to denote the time to absorption, and it said to have a continuous PH distribution. Then the probability that the CTMC is not in the absorbing state at time y is $P(Y > y) = \mathbf{p}(y)\mathbf{e} = \pi \exp(\mathbf{T}y)$. Similarly, the cumulative distribution function (CDF) of Y is $P(Y \leq y) = F(y) = 1 - \pi \exp(\mathbf{T}y)\mathbf{e}$, and the probability density function (PDF) is $f(y) = F'(y) = -\mathbf{T}\pi e^{\mathbf{T}y}\mathbf{e} = \mathbf{q}\pi \exp(\mathbf{T}y)$.

One special type of the PH distribution is Coxian PH distribution, in which the CTMC can only move sequentially or enter the absorbing state directly from any of the m transient states. It has been previously shown that Coxian PH distributions are ideal methods in fitting the LOS of patients (see [22–24]).

2.2. Aggregated Markov chain

In studying the LOS in hospital or the sojourn in one unit, the CTMC is always used to interpret the whole healthcare treatment process. If we consider each state

in the CTMC as an aggregate of several substates or groups, the aggregated Markov Chain is then introduced to measure the transition within and among the states. For example, the transfer from short-term stay group to long-term stay group inside a hospital unit and the intra-transfer among units in a hospital form an aggregated Markov Chain.

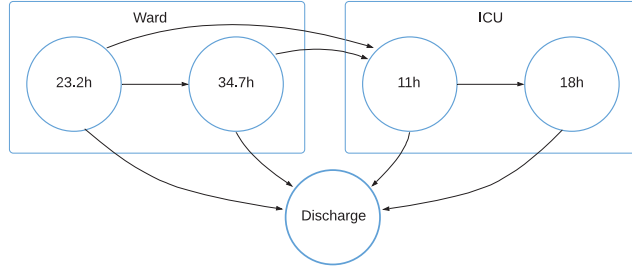


Figure 1: 2-Unit Hospital Patient Flow

The Fig. 1 illustrates a simple example for a patient flow in a hospital. A group of patients being admitted into the hospital stay in the ward for 23.2 hours averagely. A part of the group may keep staying in the ward, and another part are probably transferred to the ICU for further treatment while the rest discharged from the hospital. For the second group of patients with an average stay of 34.7 hours, there also exists a potential that some of them enter into the ICU and the rest are discharged directly. In the ICU, two similar LOS groups are identified with mean LOS as 11 hours and 18 hours separately. The patient flow process in Fig. 1 is a two-state aggregated Markov process. Each state in the aggregated Markov process represents a unit in a hospital and we assume that there exist two LOS groups in each unit. The initial distribution denotes how the patients enter into the hospital, and more specifically is the admission probability into each unit after being admitted. It is impossible for a patient belonging to the longer LOS group to be transferred to the shorter LOS group, and patients will finally leave the group they belong to. Thus, the time spent in each unit follows a Coxian PH distribution.

Transitions among LOS groups and intra-transfers among units are both transition processes, and they denote a change of status, activity, or stage characteristics. Figuring out the distributions of such transition processes and capturing the shared characteristics of patients with similar transfer experiences are meaningful to improve the healthcare efficiency and to fill the gap due to the differences in altered level of care during unit transfer.

In real applications, the number of groups in each unit can take any arbitrary positive value. Assuming there are m_1 groups in the first unit, m_2 groups in the

second unit, and so on. There is one absorbing unit as discharging from hospital at the end of the hospital visit. Once entering into the first unit with initial distribution π_1 , the system can move to the next group in the same unit with rate as $\lambda_{11}^{i,j}$, $i, j = 1, \dots, m_1, i \neq j$ or to the first group in the second unit directly with transition rate $\lambda_{12}^{i,1}$, $i = 1, \dots, m_1$. It is also possible to exit the system directly from the first unit with absorbing rates λ_{10}^i , $i = 1, \dots, m_1$. The subscript of transition rates denotes the current unit and the next unit it is going to enter into, and the superscript tells the same information of the LOS groups.

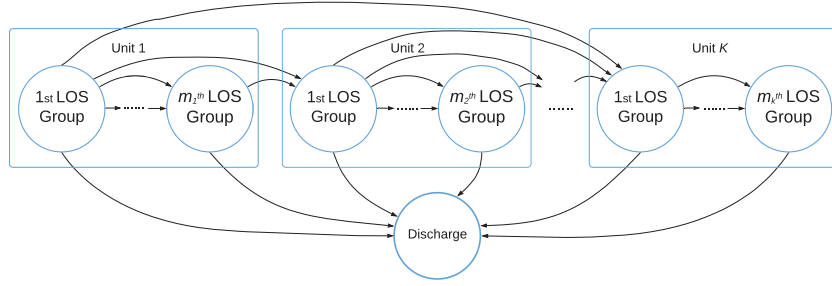


Figure 2: Hospital patient flow

Therefore, when considering K units in a hospital (Fig. 2), if assuming there are m_k groups in the unit k , $i = 1, \dots, K$, the initial distribution becomes

$$\pi = (\pi_1, \dots, \pi_K) = (\pi_1^1, \dots, 0, \pi_2^1, \dots, 0, \pi_K^1, \dots, 0)_{1 \times (\sum_{k=1}^K m_k)},$$

the absorbing vector becomes

$$\mathbf{q} = (\mathbf{q}_{10}, \mathbf{q}_{20}, \dots, \mathbf{q}_{K0})^T = (\lambda_{10}^1, \dots, \lambda_{10}^{m_1}, \dots, \lambda_{K0}^1, \dots, \lambda_{K0}^{m_K})_{(\sum_{k=1}^K m_k) \times 1}^T,$$

and the transition matrix $\mathbf{T}$ becomes a $K \times K$ -blocks partitioned matrix

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1K} \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2K} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{T}_{K1} & \mathbf{T}_{K2} & \cdots & \mathbf{T}_{KK} \end{pmatrix}_{(\sum_{k=1}^K m_k) \times (\sum_{k=1}^K m_k)},$$

where $\mathbf{q} = -\mathbf{T}\mathbf{e}$, and here $\mathbf{e}$ represents an $(\sum_{k=1}^K m_k) \times 1$ column vector with every element being 1. As shown in Fig. 2, the outer level is a general PH distribution consists of K states and one absorbing states, representing K units in hospital and discharging from hospital, respectively. The inner level for state k ($k = 1, 2, \dots, K$) is a Coxian PH distribution with m_k transient substates, denoting m_k LOS groups,

and one absorbing substate, representing leaving the unit to other units or discharging from hospital directly.

The diagonal blocks $\mathbf{T}_{kk}$, $k = 1, \dots, K$, as shown below, are transition matrices of Coxian PH distribution since there only exist transfers from the previous LOS group to the next LOS group or to the absorbing state (exiting the unit).

$$\mathbf{T}_{kk} = \begin{pmatrix} -\lambda_{kk}^1 & \lambda_{kk}^{1,2} & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & -\lambda_{kk}^{m_{k-1}} & \lambda_{kk}^{m_{k-1}, m_k} \\ 0 & 0 & \dots & 0 & -\lambda_{kk}^{m_k} \end{pmatrix}_{m_k \times m_k}.$$

The off-diagonal blocks $\mathbf{T}_{kl}$, $k, l = 1, \dots, K$, $k \neq l$, as shown below, only contain the transition rates from the all of the groups in the k th state to the first group in the l th state

$$\mathbf{T}_{kl} = \begin{pmatrix} \lambda_{kl}^{1,1} & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \lambda_{kl}^{m_k,1} & 0 & \dots & 0 \end{pmatrix}.$$

Then the probability of spending time y_k in state k , and then moving into another state l is

$$g_{kl}(y_k) = \exp(\mathbf{T}_{kk}y_k)\mathbf{T}_{kl},$$

where the (i, j) th element of the matrix $g_{kl}(y_k)$ is the conditional probability that $X(t)$ enters state k , stays for a period of y_k , and exits from the i th group to the j th group of state l . Here, $i = 1, \dots, m_k$, $j = 1, \dots, m_l$, $k, l = 1, \dots, K$, and $k \neq l$.

Similarly, the probability of spending time y_k in unit k , and then exiting the system is

$$g_{k0}(y_k) = \exp(\mathbf{T}_{kk}y_k)\mathbf{q}_{k0}$$

where the i th element of the vector $g_{k0}(y_k)$ is the conditional probability that $X(t)$ spends time y_k in state k , and exit from the i th group to the absorbing state, which should be treated as having only one group. Here, $i = 1, \dots, m_k$, $k = 1, \dots, K$.

Let $\mathbf{y}$ be an independent and identically distributed sample from a population with PDF as $f(\mathbf{y})$, denoting a sequence of sojourn in each state the observation visits before absorption. We can use a R -dimensional vector $(w_1, w_2, \dots, w_R)$ to represent the sojourn vector as $\mathbf{y} = (y_{w_1}, y_{w_2}, \dots, y_{w_R})$, where $(w_1, w_2, \dots, w_R)$ are the states it visits and R is the number of jumps the observation $\mathbf{y}$ has. Then the PDF of the observation $\mathbf{y}$ is

$$f(\mathbf{y}) = \pi_{w_1} \exp(\mathbf{T}_{w_1 w_1} y_{w_1}) \mathbf{T}_{w_1 w_2} \exp(\mathbf{T}_{w_2 w_2} y_{w_2}) \mathbf{T}_{w_2 w_3} \dots \exp(\mathbf{T}_{w_R w_R} y_{w_R}) \mathbf{q}_{w_R} \mathbf{e}, \quad (1)$$

where $1 \leq w_1, w_2, \dots, w_R \leq K$.

The parameters to be estimated in (1) include the transition rates among the LOS groups in unit k , $k = w_1, w_2, \dots, w_R$, and the transition rates from all m_k groups in unit k to the unit l , where $w_1 \leq k, l \leq w_R$.

Then, the proportion of the i th LOS group in the unit k can be calculated by integrating the transition rates in $\mathbf{T}_{kk}$, i.e.,

$$\begin{aligned} P_k^i &= \frac{\lambda_{kk}^{1,2}}{\lambda_{kk}^1} \times \frac{\lambda_{kk}^{2,3}}{\lambda_{kk}^2} \times \dots \times \frac{\lambda_{kk}^{i-1,i}}{\lambda_{kk}^{i-1}} \times \left(1 - \frac{\lambda_{kk}^{i,i+1}}{\lambda_{kk}^i}\right), \quad i = 1, \dots, m_k - 1, \\ P_k^{m_k} &= \frac{\lambda_{kk}^{1,2}}{\lambda_{kk}^1} \times \frac{\lambda_{kk}^{2,3}}{\lambda_{kk}^2} \times \dots \times \frac{\lambda_{kk}^{m_k-1,i}}{\lambda_{kk}^{m_k-1}}, \quad k = 1, \dots, K. \end{aligned} \quad (2)$$

Similarly, the probability of the transfer from the i th LOS group in unit k to the unit l can be obtained using elements in both $\mathbf{T}_{kk}$ and $\mathbf{T}_{kl}$, $P_{kl}^i = \frac{\lambda_{kl}^{i,1}}{\lambda_{kk}^i - \lambda_{kk}^{i,i+1}}$, $i = 1, \dots, m_k - 1$, and $P_{kl}^{m_k} = \frac{\lambda_{kl}^{m_k,1}}{\lambda_{kk}^{m_k}}$, $k = 1, \dots, K$, $l = 0, \dots, K$, $k \neq l$.

Furthermore, by separating the patients in unit k into LOS groups according to the P_k^i , one can obtain the actual LOS range of each group, and further analyze the shared characteristics of patients in the same group which can help detect the potential longer-stay patients. When a patient stays in unit k long enough to be assigned to group i , then there is a probability of P_{kl}^i for the patient to be transferred to unit l or discharge (if $l = 0$) directly.

3. Maximum Likelihood Estimation

In order to simplify the likelihood function and the derivative of it over the unknown parameters, we introduce forward vectors and backward vectors in Section 3.1. Moreover, the spectral decomposition is used in Section 3.2 to get rid of the exponential matrix terms in the likelihood function. The derivatives of likelihood function over transition matrix and absorbing vectors, as well as the MLE are computed in Sections 3.3, 3.4.

3.1. Forward and backward recursions

Since the likelihood function of a series of sojourn time consists of the product of the $g_{ab}(t)$ function [28], we can define the forward vectors as $\alpha_{w_r}^T = \pi_{w_1} \prod_{i=1}^r g_{w_i w_{i+1}}(y_{w_i})$, where $r = 1, 2, \dots, R$. Then by forward recursion, $\alpha_{w_r}^T = \alpha_{w_{r-1}}^T g_{w_{r-1} w_r}(y_{w_{r-1}})$, where $r = 1, 2, \dots, R$ and $\alpha_{w_0}^T = \pi_{w_1}$. The last forward vector gives the likelihood function

$$f(\mathbf{y}) = \alpha_{w_R}^T = \pi_{w_1} \prod_{i=1}^R g_{w_i w_{i+1}}(y_{w_i}) \mathbf{1}. \quad (3)$$

As expressed in (1) and (3), the likelihood function given the observation $\mathbf{y}$ is just the sum of the last forward vector α_R , which denotes the total probability of exiting the whole system from the last group, the m_R group, in the unit w_R the observation $\mathbf{y}$ visited before exiting.

In a similar way, the backward vectors β_{w_k} can be defined as $\beta_{w_r} = \prod_{i=r}^R g_{w_i w_{i+1}}(y_{w_i})$, where $r = 1, 2, \dots, R$. Again the backward recursion forms $\beta_{w_r} = g_{w_r w_{r+1}}(y_{w_r}) \beta_{w_{r+1}}$, where $r = R, R-1, \dots, 1$, and $\beta_{R+1} = \mathbf{1}$.

3.2. Spectral decomposition

In [26], the equilibrium probabilities of I_t , denoted by w_i . No matter the forward or the backward recursion involves calculating the probability function $f_{kl}(y_k)$, which depends on the matrix exponential of both $\mathbf{T}_{kk}$ and $\mathbf{T}_{kl}$, $k = 1, \dots, K$, $l = 0, \dots, K$. An efficient way to calculate the matrix exponential is using spectral representation of matrix. The spectral decomposition of matrix $\mathbf{T}_{kk}$ is

$$\mathbf{T}_{kk} = \sum_{i=1}^{m_k} \lambda_i^k \mathbf{A}_i^k, \quad (4)$$

where λ_i^k is the i th eigenvalue of $\mathbf{T}_{kk}$, and $\mathbf{A}_i^k$ is the projector onto eigenvalue λ_i^k , a matrix determine from the $\sum_{i=1}^{m_k} \mathbf{A}_i^k = \mathbb{I}$, and

$$\mathbf{A}_i^k \mathbf{A}_j^k = \begin{cases} \mathbf{A}_i^k, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

The formula for the projector is $\mathbf{A}_i^k = \prod_{\lambda_j^k \neq \lambda_i^k} \frac{\mathbf{T}_{kk} - \lambda_j^k \mathbb{I}}{\lambda_i^k - \lambda_j^k}$. According to the definition of the spectral decomposition, we have the following property:

If the spectral decomposition of $\mathbf{T}$ is $\mathbf{T} = \sum_{i=1}^n \lambda_i \mathbf{A}_i$, then it follows that for any polynomial f , one has

$$f(\mathbf{T}) = \sum_{i=1}^n f(\lambda_i) \mathbf{A}_i = f(\lambda_1) \mathbf{A}_1 + \dots + f(\lambda_n) \mathbf{A}_n.$$

Then the probability densities of the system staying in unit k for a duration y , and then transfer to another unit $k+1$ is as follows

$$g_{k,k+1}(y) = \sum_{i=1}^{m_k} \mathbf{A}_i^k \exp(\lambda_i^k y) T_{k,k+1}.$$

The one-dimensional density of entering into the system through unit k and exiting from it is

$$f_k(y_k) = \pi_k \exp(\mathbf{T}_{kk} y_k) \mathbf{q}_k = \sum_{i=1}^{m_k} \pi_k \mathbf{A}_i^k \mathbf{q}_k \exp(\lambda_i^k y_k). \quad (5)$$

The two-dimensional density of starting from unit k , transferring to unit r , and finally exit from unit r is

$$f_{k,l}(y_k, y_l) = \pi_k \exp(\mathbf{T}_{kk} y_k) \mathbf{T}_{kl} \exp(\mathbf{T}_{ll} y_l) \mathbf{q}_l = \sum_{i=1}^{m_k} \sum_{j=1}^{m_l} \pi_k \mathbf{A}_i^k \mathbf{T}_{kl} \mathbf{A}_j^l \mathbf{q}_l \exp(\lambda_j^k y_k + \lambda_i^l y_l). \quad (6)$$

3.3. Derivative of likelihood function

By applying the chain rule of differentiation to (1) and making use of α_{w_r} and β_{w_r} , we have

$$\frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{w_a w_b}} = \pi_{w_1} \prod_{r=1, r \neq a}^R g_{w_r w_{r+1}}(y_{w_r}) \mathbf{1} \times \frac{\partial g_{w_a w_b}(y_{w_a})}{\partial \mathbf{T}_{w_a w_b}} = \frac{\partial \alpha_{w_{a-1}}^T g_{w_a w_b}(y_{w_a}) \beta_{w_b}}{\partial \mathbf{T}_{w_a w_b}}. \quad (7)$$

The key to calculating the equations (7) is the expression $\alpha_{w_{a-1}}^T g_{w_a w_b}(y_{w_a}) \beta_{w_b}$, $a, b = 1, \dots, R+1$. Furthermore, the expression $\alpha_{w_{a-1}}^T g_{w_a w_b}(y_{w_a}) \beta_{w_b}$ only depends on $\mathbf{T}_{w_a w_a}$ and $\mathbf{T}_{w_a w_b}$. By equation (4) and equation (5), we have the spectral representation as follows

$$g_{w_a w_b}(y_{w_a}) = \sum_{i=1}^{n_{w_a}} \mathbf{A}_{i, w_a} \mathbf{T}_{w_a w_b} \exp(\lambda_{i, w_a} y_{w_a}).$$

It is easy to find that the function $g_{w_a w_b}(y_{w_a})$ linearly depends on $\mathbf{T}_{w_a w_b}$, thus the derivative can be represented as

$$\frac{\partial \alpha_{w_{a-1}}^T g_{w_a w_b}(y_{w_a}) \beta_{w_b}}{\partial \mathbf{T}_{w_a w_b}} = (\alpha_{w_{a-1}}^T \exp(\mathbf{T}_{w_a w_a} y_{w_a}))^T \beta_{w_b}^T = \sum_{i=1}^{n_{w_a}} \mathbf{A}_{i, w_a} \alpha_{w_{a-1}} \beta_{w_b}^T \exp(\lambda_{i, w_a} y_{w_a}).$$

According the findings in [28], the derivative of likelihood function over transition matrix and absorbing vectors are as follows

$$\begin{aligned} \frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kl}} &= \sum_{i=1}^{m_k} \mathbf{A}_{i, k}^T \left[\sum_{r=1, w_r=k, w_{r+1}=l}^R \bar{\alpha}_{w_{r-1}} \bar{\beta}_{w_{r+1}}^T \exp(\lambda_{i, k} y_k) \right], \\ \frac{\partial f(\mathbf{y})}{\partial \mathbf{q}_k} &= \sum_{i=1}^{m_k} \mathbf{A}_{i, k}^T \left[\sum_{r=1, w_r=k, w_{r+1}=0}^R \bar{\alpha}_{w_{r-1}} \bar{\beta}_{w_{r+1}}^T \exp(\lambda_{i, k} y_k) \right], \\ \frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kk}} &= \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} \mathbf{A}_{i, k}^T \left[\sum_{r=1, w_r=k}^R \bar{\alpha}_{w_{r-1}} \bar{\beta}_{w_{r+1}}^T \mathbf{T}_{k, w_{r+1}} \gamma_{i, j, k}(y_k) \right] \mathbf{A}_{j, k}^T, \end{aligned} \quad (8)$$

where

$$c_k = \frac{1}{|\alpha_k|}, \quad \bar{\alpha}_k = \left(\prod_{i=0}^k c_i \right) \alpha_k, \quad \bar{\beta}_k = \left(\prod_{i=k}^R c_i \right) \beta_k,$$

$$\gamma_{i,j,k}(y_k) = \begin{cases} y_k \exp(\lambda_{i,k} y_k), & i = j \\ \frac{\exp(\lambda_{i,k} y_k) - \exp(\lambda_{j,k} y_k)}{\lambda_{i,k} - \lambda_{j,k}}, & i \neq j \end{cases}$$

It should be taken into account the relationship between $\mathbf{T}_{kl}$ and $\mathbf{T}_{kk}$ that $\sum_{l=0}^K \mathbf{T}_{kl} = \mathbf{0}$, $k = 1, \dots, K$, that is the rows of the $\mathbf{Q}$ should sum to zero. The adjusted partial derivative of likelihood function over $\mathbf{T}_{kl}$ should be

$$\frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kl}} + \frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kk}} \frac{\partial \mathbf{T}_{kk}}{\partial \mathbf{T}_{kl}} = \frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kl}} - \mathbf{1} \frac{\partial f(\mathbf{y})}{\partial \mathbf{T}_{kk}}, \quad l \neq k, \quad k, l = 1, \dots, K.$$

Suppose we have N observations, the set $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N)$ consists of units vectors, where $\mathbf{w}_n = (w_1^n, \dots, w_{R_n+1}^n)_{1 \times (R_n+1)}$ represents the units, including the absorbing unit, the observation n visits sequentially, including the absorbing states. Similarly, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$, $\mathbf{y}_n = (y_{w_1^n}, \dots, y_{w_{R_n}^n})_{1 \times R_n}$ denotes the LOS time in each unit, excluding the absorbing unit, the n th observation visits. Timing up and taking log of the likelihood function given n th observation, we can obtain the final log-likelihood function as

$$\log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}, \mathbf{q}) = \sum_{n=1}^N \log(\pi_{w_1^n} \prod_{r=1}^{R_n} g_{w_r^n, w_{r+1}^n}(y_{w_r^n})).$$

Then the final derivative of log likelihood function over matrix $\mathbf{T}$ is as follows

$$\begin{aligned} \frac{\partial \log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}, \mathbf{q})}{\partial \mathbf{T}} &= \sum_{n=1}^N \frac{\partial \log \pi_{w_1^n} \prod_{r=1}^{R_n} g_{w_r^n, w_{r+1}^n}(y_{w_r^n})}{\partial \mathbf{T}} \\ &= \sum_{n=1}^N \frac{\partial \log f(\mathbf{y}_n)}{\partial \mathbf{T}} = \sum_{n=1}^N \frac{1}{f(\mathbf{y}_n)} \frac{\partial f(\mathbf{y}_n)}{\partial \mathbf{T}}. \end{aligned} \quad (9)$$

3.4. Maximum likelihood estimation

In Section 3.3, we already compute the expression of likelihood and its first order derivatives with respect to elements of $\mathbf{Q}$, then the next step is to maximize the likelihood. Actually, generating the first order derivatives by the forward and backward recursion is already time consuming, neither is computing the Jacobian matrix of the gradient or the Hessian matrix of the likelihood. Hence, the Quasi-Newton method can be used here, in which the Hessian matrix is not computed. Instead, the Hessian matrix is replaced with its approximation.

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, as one of the Quasi-Newton algorithms, uses Sherman-Morrison formula to obtain the updated Hessian matrix and is one of the most popular method of the same class. The Wolfe conditions are used in the line search procedure to compute an acceptable step length that reduces the objective function. The steps in BFGS are shown in Algorithm 1, [which is implemented in Matlab](#).

Algorithm 1 BFGS Method

- 1: Given starting point $\mathbf{T}_0$, convergence tolerance $\varepsilon > 0$, inverse Hessian approximation H_0 , iteration index $s \leftarrow 0$;
- 2: **while** $\frac{\partial \log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}^{(s)}, \mathbf{q})}{\partial \mathbf{T}}^{(s)} > \varepsilon$ **do**;
- 3: Compute search direction

$$p_s = -H_s \times \frac{\partial \log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}^{(s)}, \mathbf{q})}{\partial \mathbf{T}^{(s)}};$$

- 4: Set $\mathbf{T}^{(s+1)} = \mathbf{T}^{(s)} + \alpha_s p_s$ (α_s obtained from a line search satisfying the Wolfe conditions);
- 5: Define $s_s = \mathbf{T}^{(s+1)} - \mathbf{T}^{(s)}$ and

$$u_s = \frac{\partial \log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}^{(s+1)}, \mathbf{q})}{\partial \mathbf{T}_{k+1}} - \frac{\partial \log L(\mathbf{W}, \mathbf{Y} | \mathbf{T}^{(s)}, \mathbf{q})}{\partial \mathbf{T}^{(s)}};$$

- 7: Compute the updated $H_{s+1} = (\mathbf{I} - \frac{s_s u_s^T}{u_s^T s_s}) H_s (\mathbf{I} - \frac{u_s s_s^T}{u_s^T s_s}) + \frac{s_s s_s^T}{u_s^T s_s}$;
 - 8: $s \leftarrow s + 1$;
-

4. Experiments

4.1. Data processing

We collected the transfer information of patients in the Banner University Medical Center Tucson - Main Campus and South Campus from 2012 to 2017. The data consists of patient medical record number, event type (admission, transfer, and discharge), unit ID, unit name, and the date and time the patients being transferred into and out of the units. The admission and discharge details as well as the demographic information are also included.

Since the LOS in ED is always a critical issue and it impacts the efficiency of healthcare and the survive of patients, we are interested in building a statistical model which can measure the LOS within ED and the transfer rates to other units inside a hospital. We filter the datasets to only contain the records of geriatric patients being admitted into ED directly, and then being transferred to other units for further surgical procedure or to GW before discharge. The information those patients under unqualified conditions, such as missing values, errors in record, etc, are illuminated.

The cleaned data set consists of 3225 records and contains information of 2849 patients, and the demographic information of the patients is presented in Table 1. More than half of the patients are females, and 75.18% of them are under 85 years old. Moreover, most of the patients are discharged to home for self-care.

Table 1: Demographic of patients

Category		#	Frequency
Number of patients		2849	100.00%
Gender	Female	1584	55.60%
	Male	1262	44.30%
	Unknown	3	0.10%
Age	Age $\leq$ 85	2142	75.18%
	Age $>$ 85	707	24.82%
Race	White or Caucasian	2430	85.29%
	American Indian or Alaska Native	63	2.21%
	Asian	27	0.95%
	Black or African American	82	2.88%
	Multiracial	23	0.81%
	Unknown	27	0.95%
	Other	191	6.70%
	Patient Refused	6	0.21%
Admit Type	Elective	6	0.21%
	Emergency	2623	92.07%
	Trauma Center	220	7.72%
Discharge Destination	Home and Self-care	2217	77.82%
	Other hospital or Facility	532	18.67%
	Readmission	4	0.14%
	Others	96	3.37%

Following the description of total LOS in hospital shown in Table 2, the patients were discharged averagely after 65.48 hours (approximately 3 days), while the maximum LOS in hospital is up to 926.62 hours (approximately 39 days), and the minimum LOS is 0.1 hours which refers to an immediate discharge after admission. The mean LOS is larger than the median value of LOS, indicating the overall LOS distribution is right skewed. Furthermore, for the distribution of the LOS, approximately 25% of patients were discharged after within 1 day, and out of these patients, approximately 75% left the hospital within three days of hospitalization.

Table 2: Overall LOS description

#	3225
max	926.62
min	0.10
mean	65.48
25th percentile	26.03
median	43.68
75th percentile	77.68

In the filtered data set, 4 transient units kept are ED, CDU, GW, and ICU. In Table 3, the number of visits in each unit, and the range (minimum/min and maximum/max values in hours) of the LOS in the units are presented. Since all of the patients are admitted into the ED, the visits in this unit is same with the number of patient records. The LOS in the ED ranges from 0.1 hour to 73.57 hours, which means that patients can be transferred to other units directly after admission or be provided with treatment in the ED.

The average stay in CDU is almost one day (23.88 hours). According to [16], the LOS for patients in the CDU could hopefully be decreased to under 18 hours. Besides developing and implementing inclusion criteria, analyzing the patients' characteristic and distribution in different LOS groups can also help. For the group with longer LOS in the CDU, the admission to the CDU can be restricted and the transfer to other units can also be considered to release the burden in the CDU.

Table 3: Statistical description of LOS in each unit

Unit Name	#	max	min	mean	25th percentile	median	75th percentile
ED	3225	73.57	0.10	8.53	5.39	7.37	10.27
GW	1746	836.65	0.38	79.61	36.08	61.93	98.82
CDU	1301	119.03	0.57	23.88	15.42	20.33	26.57
ICU	129	914.43	1.98	102.60	25.43	47.75	111.25

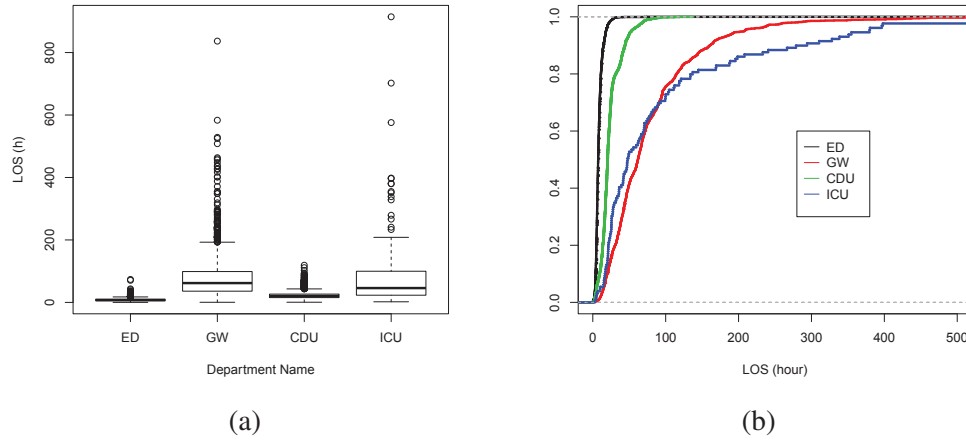


Figure 3: The boxplot and CDF of LOS in each unit

The Fig. 3 shows the box-plot and the CDF distribution of the LOS in each unit, in which a visual comparison of range, average, percentile of the LOS among

units are presented. In Fig. 3(a), the GW has the most outliers. But according to Kolmogorov-Smirnov (KS) tests, the LOS data does not follow normal distribution, then we still keep the outliers. The time spent in ED is smaller and has less variation than other units, which can also be illustrated in the CDF plots in Fig. 3(b).

By fitting the LOS data by Coxian PH distributions, each unit can have various number of LOS groups, ranging from 2 to 7 from existing literature. In this paper, we first use classic Coxian PH distributions to fit the dwell in each unit (refer to [23] for the methods), then assign the visits in each unit into the corresponding optimal group numbers.

4.2. Intra-hospital patient transfer

In Section 4.1, the LOS is analyzed only in the perspective of a single unit without considering the intra-transfers from or to the unit. In this section, we introduce the types of transfers in our data set, and conduct analysis on the LOS in each transfer route. Among all records of the patient flow information, 51.38% of them become ED-GW patients, 40.34% of them are ED-CDU patients, 4.00% visit ICU, and the rest 4.25% are discharged directly after being treated in ED.

Since all of the patients in our collected data start their stay in a hospital from ED, the LOS distributions of patients in ED before being transferred to other units are analyzed in Table 4. The patients discharged from ED directly have the largest mean and the smallest minimum of LOS, and that is due to the complex reasons of leaving the hospital ranging from death, admission no show, inter-hospital transfer, to discharges to home.

Table 4: LOS distribution in ED before transfer to the next unit

To Unit	#	max	min	mean	25th percentile	median	75th percentile
Ward	1657	43.30	1.50	9.00	5.90	7.87	10.73
CDU	1301	73.57	1.00	7.98	5.07	6.88	9.62
ICU	129	19.57	1.17	5.62	3.37	4.52	7.03
Discharge	138	34.23	0.10	10.60	4.26	9.91	16.47

The patients transferred to CDU also have large statistical values of LOS in the ED. The patients who are transferred to CDU from ED are not well enough for direct discharge but also not in a severe condition to be admitted to the ward. In Fig. 4, the plots of CDF of the time spent in each unit during intra-transfer routes are compared against the one of the overall LOS in that unit.

In the literature, the transfers from ED to CDU, GW, ICU, and from ICU to GW are widely studied, so are the challenges during these transfers. In [32], a recent study shows that patients in a CDU had a 25% shorter overall length of stay,

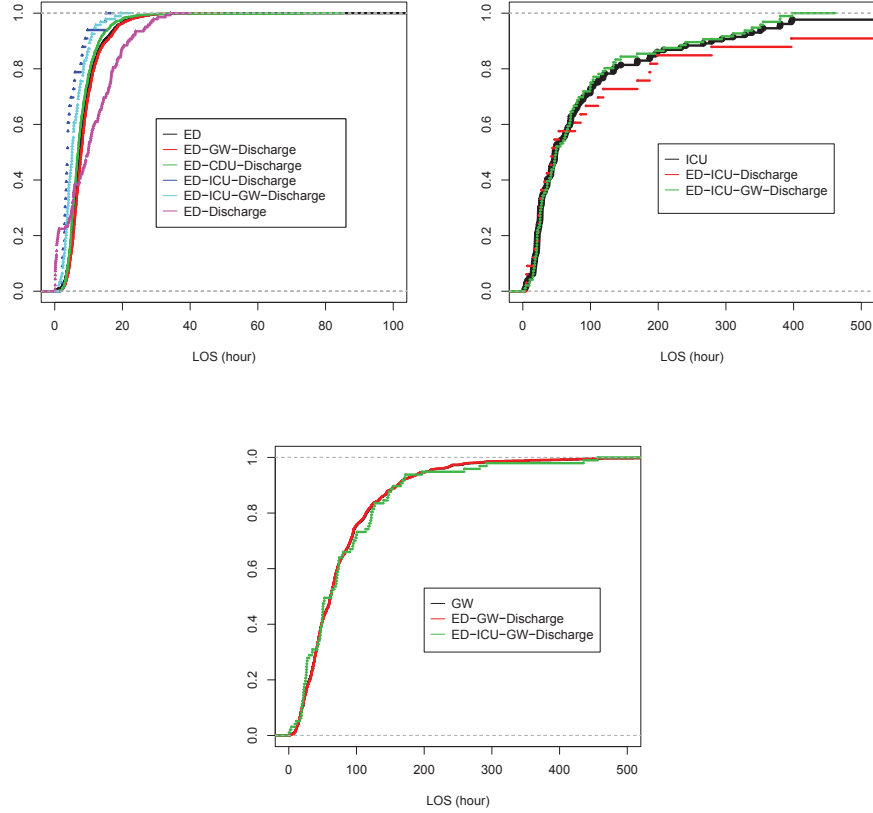


Figure 4: LOS in each [intra-transfer](#) compared with the total LOS in each unit

resulting in estimated cost savings of \$1 billion/year across the US. The ED-CDU transfers play an increasingly significant role in decreasing overall LOS, reducing number of admissions to wards, and enhancing quality and safety of treatment.

For the ED unit, as one of most common unit where the patients in GW are transferred from, the waiting time of a patient from the time of requesting a bed in GW to the time of leaving the ED is a significant measure of health care efficiency [33]. An adequate modeling of the transfer process can help avoid delays in an ED-GW transfers which are due to the allocation issue.

The ED-ICU patients have relative short LOS because of the severity in the situation, and doctors in ED have to make decision on the further treatment as soon as possible. The ED and ICU are two main admission sources of inpatients admission to GW. Even though there exists cases where patients go through multiple

transfers, such as from GW to ICU, and then back to GW, we only consider the initial visit to GW.

Although the ICU-GW transfers are not as urgent as ED-GW since patients have already received satisfactory treatments in ICU [33], there still exist challenges in the handing over patients from the ICU to the GW as communication and collaborative failure [34]. One of the potential solutions is to figure out the characteristics of patients involved in the ICU-GW transfer.

Above all, in this paper, the transfer routes we consider include, ED-GW-Discharge, ED-CDU-Discharge, ED-ICU-Discharge, ED-Discharge, ED-ICU-GW-Discharge. The unit transfers include ED-GW, ED-CDU, ED-ICU, ED-Discharge, ICU-GW and ICU-Discharge. One needs to notice that the sojourn time within one unit follows a Coxian PH distribution, while the overall LOS before discharge follows a general PH distribution.

4.3. Model construction

As stated in Section 4.2, four units are chosen as ED, GW, CDU, and ICU in our model. Each unit can have a variety number of LOS groups (Section 4.1). According to the top two optimal fitting results, we assign 4 or 5 groups to unit ED, 2 or 3 groups to unit GW and unit CDU, and 2 groups to unit ICU. We construct the aggregated PH distribution based on various combinations of LOS groups and use the MLE method to estimate the parameters. In Table 5, the number of LOS groups in each unit and the log likelihood value as well as Akaike Information Criterion (AIC) are presented. The set of LOS group numbers having the maximal log-likelihood and minimal AIC are chosen as $m_1 = 4, m_2 = 3, m_3 = 4, m_4 = 2$.

Table 5: The log-likelihood and AIC values

m_1	m_2	m_3	m_4	log-likelihood	AIC
4	3	3	2	-30089.63	60253.26
4	2	3	2	-30126.70	60321.41
4	3	4	2	-29761.80	59601.60
4	2	4	2	-29799.21	59670.41
5	3	3	2	-30275.71	60635.42
5	2	3	2	-30329.99	60737.98
5	3	4	2	-29947.88	59983.76
5	2	4	2	-29985.29	60052.57

The transfer routes mentioned in Section 4.2 are shown in Figure 5. Patients admitted to the ED can either go to one of the GW, CDU, and ICU or discharge. The patients can also be transferred to the next LOS group in the same unit if they

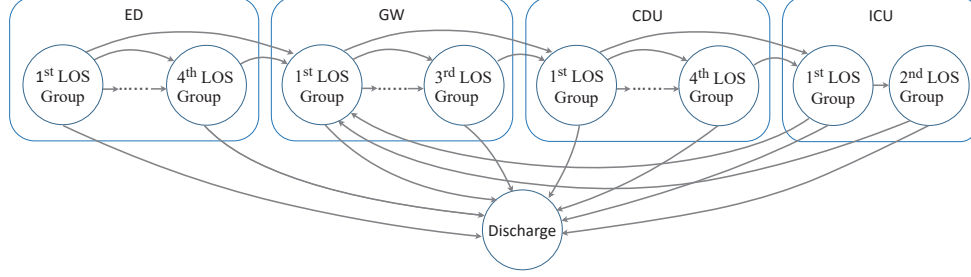


Figure 5: 4-unit hospital patient flow

stay longer. Besides, no matter which LOS group the patients belong to, they can also be discharged from hospital directly.

The initial distribution is a vector with only non-zero value for the first element since all of the transfer routes start from the first group in the ED. That is

$$\pi = (\pi_1^1, 0, \dots, 0, 0)_{1 \times (\sum_{k=1}^4 m_k)} = (1, 0, \dots, 0, 0)_{1 \times (\sum_{k=1}^4 m_k)},$$

and the transition matrix is constructed as

$$\begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \mathbf{T}_{13} & \mathbf{T}_{14} \\ \mathbf{0} & \mathbf{T}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{42} & \mathbf{0} & \mathbf{T}_{44} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{T}_{11} &= \begin{pmatrix} -\lambda_{11}^1 & \lambda_{11}^{1,2} & 0 & 0 \\ 0 & -\lambda_{11}^2 & -\lambda_{11}^{2,3} & 0 \\ 0 & 0 & -\lambda_{11}^3 & -\lambda_{11}^{3,4} \\ 0 & 0 & 0 & -\lambda_{11}^4 \end{pmatrix}, \mathbf{T}_{12} = \begin{pmatrix} \lambda_{12}^{1,1} & 0 & 0 \\ \lambda_{12}^{2,1} & 0 & 0 \\ \lambda_{12}^{3,1} & 0 & 0 \\ \lambda_{12}^{4,1} & 0 & 0 \end{pmatrix}, \mathbf{T}_{13} = \begin{pmatrix} \lambda_{13}^{1,1} & 0 & 0 & 0 \\ \lambda_{13}^{2,1} & 0 & 0 & 0 \\ \lambda_{13}^{3,1} & 0 & 0 & 0 \\ \lambda_{13}^{4,1} & 0 & 0 & 0 \end{pmatrix}, \\ \mathbf{T}_{14} &= \begin{pmatrix} \lambda_{14}^{1,1} & 0 \\ \lambda_{14}^{2,1} & 0 \\ \lambda_{14}^{3,1} & 0 \\ \lambda_{14}^{4,1} & 0 \end{pmatrix}, \mathbf{T}_{22} = \begin{pmatrix} -\lambda_{22}^1 & \lambda_{22}^{1,2} & 0 \\ 0 & -\lambda_{22}^2 & -\lambda_{22}^{2,3} \\ 0 & 0 & -\lambda_{22}^3 \end{pmatrix}, \mathbf{T}_{33} = \begin{pmatrix} -\lambda_{33}^1 & \lambda_{33}^{1,2} & 0 & 0 \\ 0 & -\lambda_{33}^2 & -\lambda_{33}^{2,3} & 0 \\ 0 & 0 & -\lambda_{33}^3 & -\lambda_{33}^{3,4} \\ 0 & 0 & 0 & -\lambda_{33}^4 \end{pmatrix}, \\ \mathbf{T}_{42} &= \begin{pmatrix} \lambda_{42}^{1,1} & 0 & 0 \\ \lambda_{42}^{2,1} & 0 & 0 \end{pmatrix}, \mathbf{T}_{44} = \begin{pmatrix} -\lambda_{44}^1 & \lambda_{44}^{1,2} \\ 0 & -\lambda_{44}^2 \end{pmatrix}. \end{aligned}$$

Table 6: The Estimated Transfer Rates

Transfer Rates	ED				GW				CDU				ICU	
	Group1	Group2	Group3	Group4	Group1	Group2	Group3	Group4	Group1	Group2	Group3	Group4	Group1	Group2
ED	Group1	-0.6740	0.6662		0.0011				0.0004				0.0004	
	Group2		-0.7581	0.7544	0.0021				0.0012				0.0004	
	Group3			-0.7458	0.1409			0.4509	0.1404				0.0102	
	Group4				0.7139			-1.2100	0.4416				0.0203	
GW	Group1				-0.0268	0.0264								
	Group2					-0.0394	0.0068							
	Group3						-0.0106							
CDU	Group1								-0.1682	0.1665				
	Group2									-0.1665	0.1654			
	Group3										-0.1807	0.1025		
	Group4											-0.0852		
ICU	Group1				0.0400								-0.0556	0.0150
	Group2				0.0054								-0.0055	

4.4. Results analysis

The parameter estimation results are presented in Table 6. The $m_k \times m_k, k = 1, \dots, 4$ matrices on the diagonal positions are the transition rates among LOS groups within the same unit, while off diagonal parameters are the transition rates among different units. For example, the transition rate from the first LOS group to the second LOS group in ED is 0.6662, and the transition rate from the first LOS group from ED to GW is 0.0011.

According to eq. (2), the LOS range in each groups with consideration of the involved transfers can be obtained. In Table 7, the minimum/min and maximum/max values in hours of LOS in each unit with impacts of transfers are shown.

Table 7: Aggregated PH distribution fitted in each unit

	ED		GW		CDU		ICU	
	min	max	min	max	min	max	min	max
Group1	0.10	1.42	0.38	4.28	0.57	2.43	1.98	99.55
Group2	1.50	2.07	4.32	122.20	2.45	3.03	99.75	914.43
Group3	2.10	6.57	122.25	836.65	3.10	19.40	-	-
Group4	6.58	73.57	-	-	19.42	119.03	-	-

Moreover, we can summarize that 98.84% of Group1 patients and 99.51% of Group2 patients staying in the ED will come to the next LOS group for longer stay in the ED. The last LOS group in the ED have a probability of 0.59 to be admitted into the GW and a probability of 0.36 to visit the CDU. The results coincide with the truth that 99.58% of GW patients and 99.46% of CDU patients previously stay in the ED for more than 2.07 hours. We can conclude that if a patient has been cared for more than about 2 hours in the ED, there is a high chance for him to be transferred to other units for further treatment.

Similarly, 98.51% of Group1 patients in GW and over 98% of Group1 and Group2 patients in CDU will stay longer. The transition probability to discharge are 0.8247 from the second group in the GW and 0.4328 from the third group in the CDU respectively. That is to say, for those patients who have already been cared for 5 days in GW or over 1 day in CDU are probably leaving the hospital very soon. Both the first and the second groups in the ICU have a relative high probability of being transferred to the GW as 0.7194 and 0.9818 respectively, which means that the transfer route in the hospital would hard to come to an end soon once entering the ICU.

Figures 6, 7, 8, and 9 show a comparison of the age, gender, admission type and discharge destination, respectively, of patients who are involved in a unit transfer among four LOS groups in ED. For ED-CDU patients, trauma-center admission

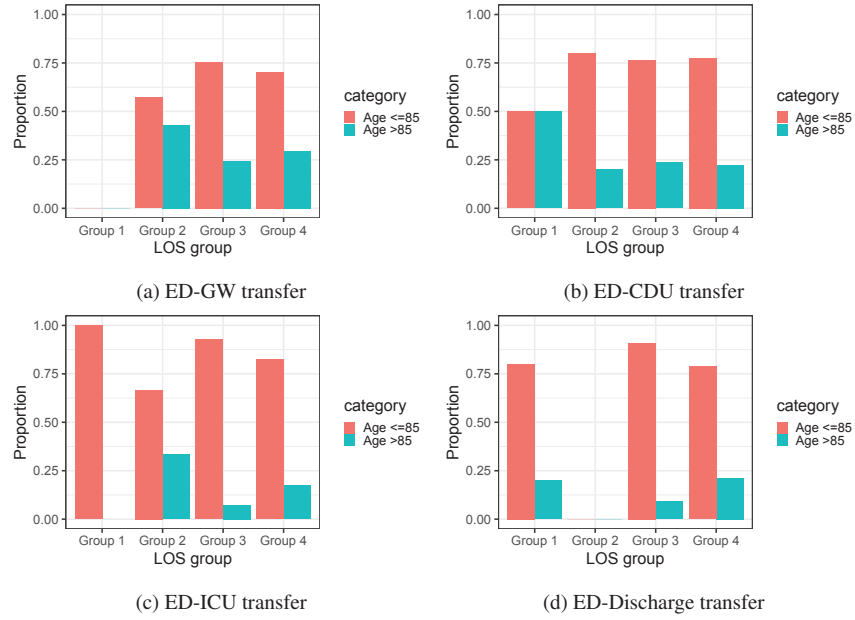


Figure 6: The age in LOS groups in each transfer from ED

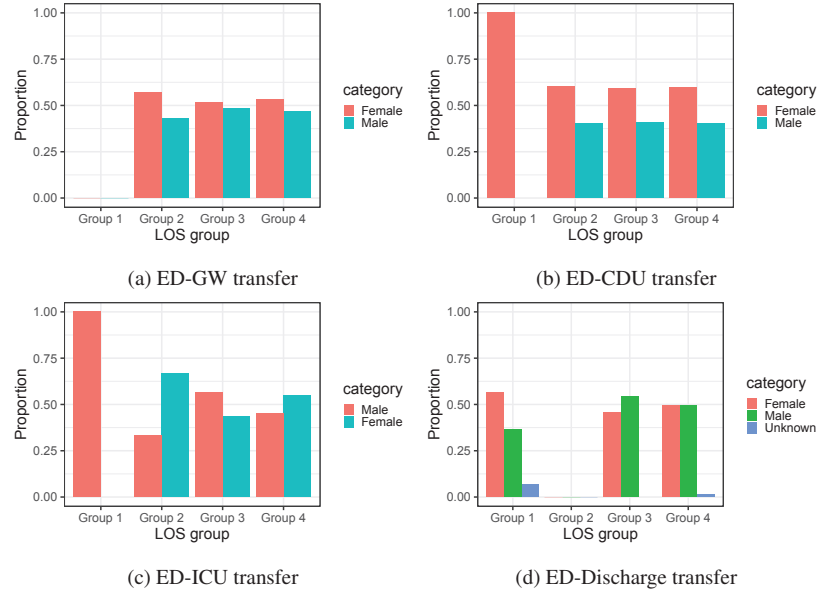


Figure 7: The gender in LOS groups in each transfer from ED

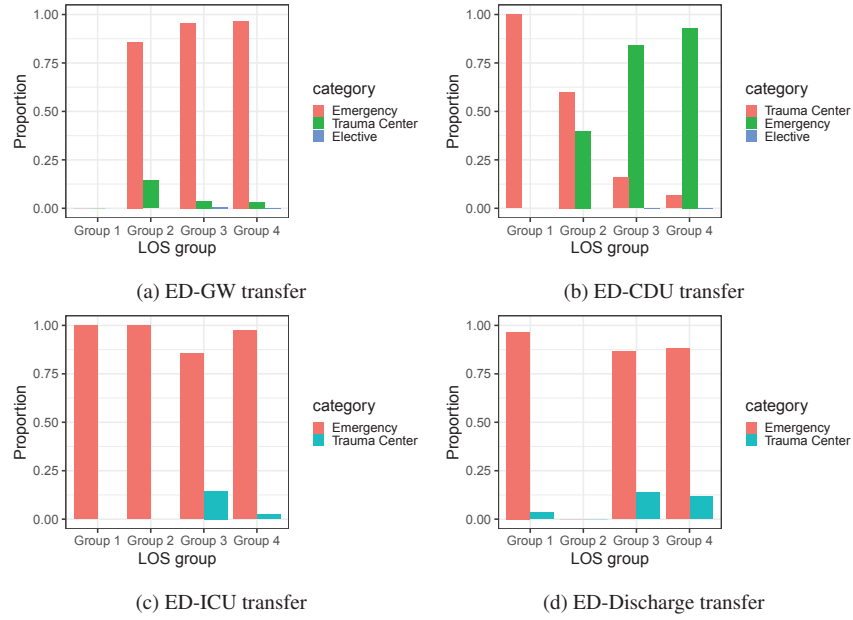


Figure 8: The admission type in LOS groups in each transfer from ED

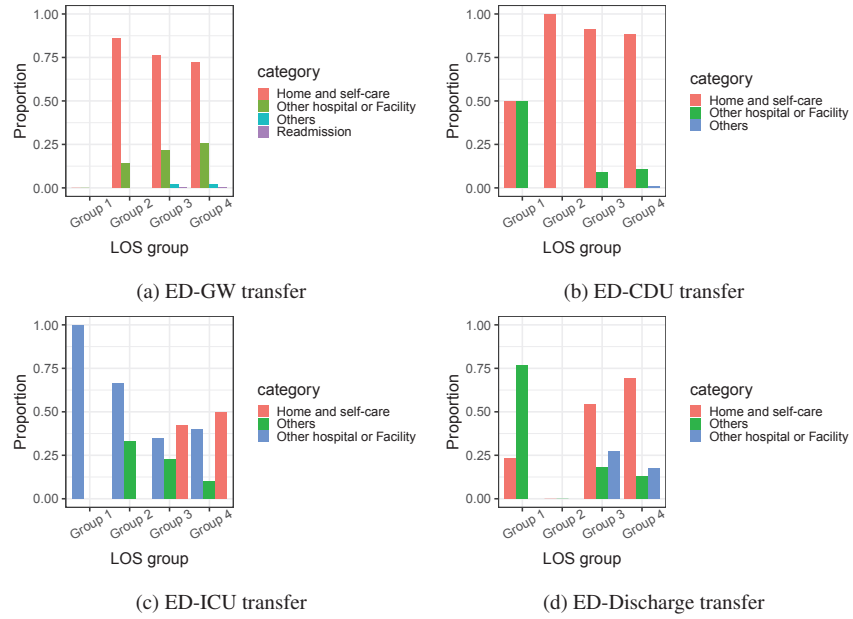


Figure 9: The discharge destinations in LOS groups in each transfer from ED

may associate with a shorter stay in ED and a relatively lower rate of discharging to home. Trauma-center, as a unit which treats the most high-risk of injuries, comes with less direct discharge to home following the treatment in both [35] and [36]. In this case, a further study in ED-CDU transfer can be conducted to verify the function of the CDU in improving the healthcare efficiency and outcomes.

For ED-ICU patients, the first LOS group in ED only consists of those who are less than 85 years old and most of them are males. Staying in ED for less than 1.5 hours and being sent to ICU very soon may be due to the condition of patient getting worse and results in inter-hospital transfer at last. For those stay in ED for a short period and discharge directly, uncontrolled worsening and higher rate of mortality are more likely to happen.

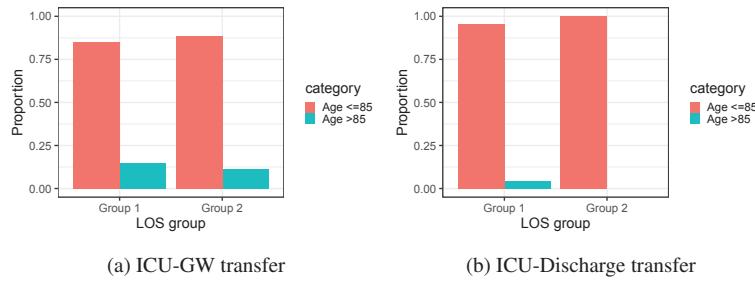


Figure 10: The age in LOS groups in each transfer from ICU

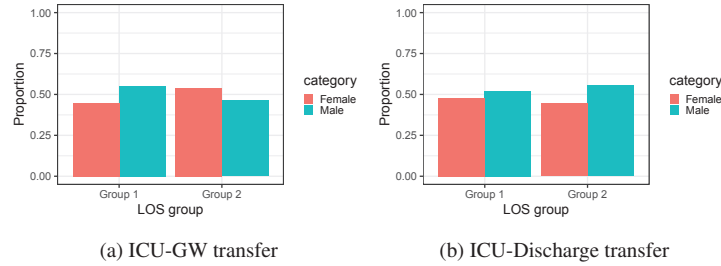


Figure 11: The gender in LOS groups in each transfer from ICU

Similarly, Figures 10, 11, 12 and 13 present comparisons on both LOS groups in ICU with consideration to the ICU-GW and ICU-Discharge transfers. Based on our data, patients who follows ED-ICU-GW-Discharge route and stay in ICU over 5 days (120 hours) are probably transferred to other hospital or healthcare facility for further treatment and recovery care. Meanwhile, patients who leave the ICU from the longer stay group have a high rate of mortality or readmission.

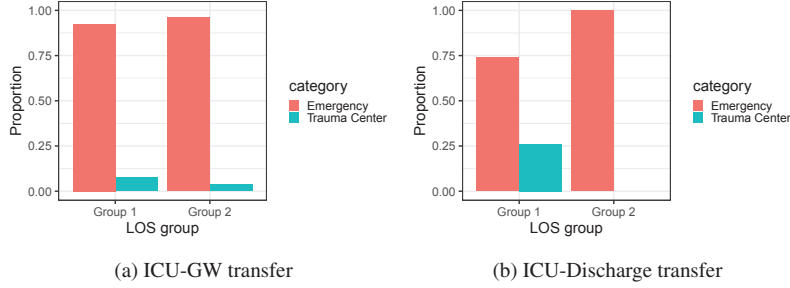


Figure 12: The admission type in LOS groups in each transfer from ICU

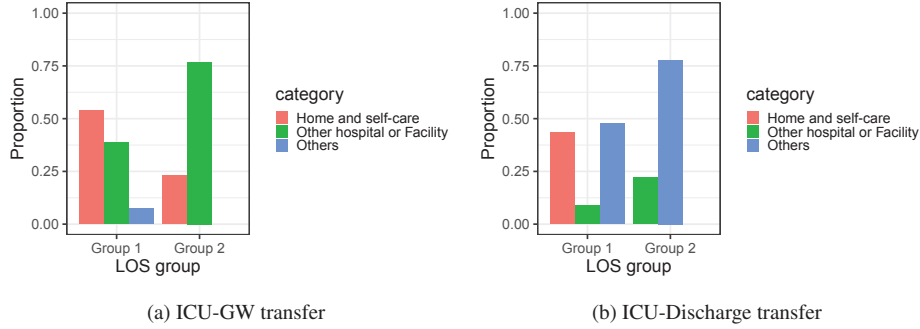


Figure 13: The discharge destinations in LOS groups in each transfer from ICU

5. Conclusions

In this paper, the overall LOS information in ED, GW, CDU, ICU, and the LOS of patients involved in each intra-transfer are investigated. An aggregated Markov chain is constructed based on the patient flow information, and each state in the aggregated Markov chain corresponds to a unit in the hospital. Patients in each unit can be divided into multiple groups according to their LOS in the unit, and there exist transition rates from shorter LOS groups to longer LOS groups.

Even though there exist cases where patients go through multiple transfers, such as from GW to ICU, and then back to GW, we only consider the initial visit to GW. Although the ICU-GW transfers are not as urgent as ED-GW since patients have already received satisfactory treatments in ICU [33], there still exist challenges in the handing over patients from the ICU to the GW as communication and collaborative failure [34]. One of the potential solutions is to figure out the characteristics of patients involved in the ICU-GW transfer.

By fitting aggregated PH distribution to the sequences of time patients spent in each unit, the transition rate between two LOS groups in each unit and the transfer

rates among units are estimated. Further analysis such as the LOS distribution (in the current unit) of patients who discharged from each group in the previous unit are conducted respectively. Moreover, covariates such as age, gender, admission type, and discharge destination are analyzed in terms of patients in each unit along with patients in each transfer route. The analysis shows that the characteristics of patients, their LOS in a unit, and the following transfer behaviors are associated, which gives us an overall view of which factors to notice when a new patient comes to the hospital. For example, patients stay in the ED for more than 2 hours would probably be transferred to next unit for further treatment. For patients who are transferred from ED to CDU, trauma-center admission is associated with a shorter stay in ED and a lower home discharge rate.

The covariate analysis in this paper are mainly from a interpretation perspective. Actually, the impacts of covariates can be incorporated into the aggregated PH distribution, and the corresponding algorithm should be developed. We focus on the intra-hospital transfer in this paper, and in reality, the transfers among different healthcare facilities happen all the time. For future research, the aggregated PH distributions can be extended to multiple layers where the inter-hospital transfers will be the upper level state in the aggregated Markov chain. The algorithms for estimating the multi-level aggregated PH distributions are worth further study to build a complete transfer system.

Acknowledgement. We would appreciate the University of Arizona Center for Biomedical Informatics & Biostatistics Department of Biomedical Informatics Services for providing the data. This material is based upon work supported by National Science Foundation Grants #1634282, #1635379 and #1740858.

References

- [1] Kulshrestha, A., Singh, J. (2016). Inter-hospital and intra-hospital patient transfer: Recent concepts. *Indian Journal of Anaesthesia*, 60(7), 451-457.
- [2] Warren, J., Fromm, R. E., Orr, R. A., Rotello, L. C. and Horst, H. M. (2004). Guidelines for the inter-and intrahospital transport of critically ill patients. *Critical Care Medicine*, 32(1), 256-262.
- [3] American College of Emergency Physicians (1990). Principles of appropriate patient transfer. *Annals of Emergency Medicine*, 19(3), 337-338.
- [4] Ulrich, R. S. and Zhu, X. (2007). Medical complications of intra-hospital

patient transports: Implications for architectural design and research. *HERD: Health Environments Research & Design Journal*, 1(1), 31-43.

- [5] Blay, N., Duffield, C. M. and Gallagher, R. (2012). Patient transfers in Australia: implications for nursing workload and patient outcomes. *Journal of Nursing Management*, 20(3), 302-310.
- [6] Chalfin, D. B., Trzeciak, S., Likourezos, A., Baumann, B. M. and Dellinger, R. P.(2007). Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical Care Medicine*, 35(6), 1477-1483.
- [7] Singer, A. J., Thode Jr. H.C., Viccellio, P. and Pines, J. M. (2011). The association between length of emergency department boarding and mortality. *Academic Emergency Medicine*, 18(12), 1324-1329.
- [8] Escobar, G. J., Greene, J. D., Gardner, M. N., Marelich, G. P., Quick, B. and Kipnis, P. (2011). Intra-hospital transfers to a higher level of care: contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *Journal of Hospital Medicine*, 6(2), 74-80.
- [9] Dunn M. J., Gwinnutt C. L., Gray A. J. (2007). Critical care in the emergency department: patient transfer. *Emergency Medicine Journal*, 24(1), 40-44.
- [10] Alamanou, D.G. and Brokalaki, H. (2014). Intrahospital transport policies: The contribution of the nurse. *Health Science Journal*, 8(2), 166-178.
- [11] Hendrich, A. L., Lee, N. (2005). Intra-unit patient transports: time, motion, and cost impact on hospital efficiency. *Nursing Economics*, 23(4), 157-164.
- [12] Baek, H., Cho, M., Kim, S., Hwang, H., Song, M. and Yoo, S., (2018). Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. *PloS One*, 13(4), e0195901.
- [13] Andrulis D. P., Kellermann A., Hintz E. A., Hackman B. B., Weslowski V. B. (1991). Emergency departments and crowding in United States teaching hospitals. *Annals of Emergency Medicine*, 20(9), 980-986.
- [14] Herring, A., Wilper, A., Himmelstein, D. U., Woolhandler, S., Espinola, J. A., Brown, D. F., Camargo, Jr. C. A. (2009). Increasing length of stay among adult visits to US emergency departments, 2001-2005. *Academic Emergency Medicine*, 16(7), 609-616.

- [15] Lo, S. M., Choi, K. T., Wong, E. M., Lee, L. L., Yeung, R. S., Chan, J. T., Chair, S. Y. (2014). Effectiveness of Emergency Medicine Wards in reducing length of stay and overcrowding in emergency departments. *International Emergency Nursing*, 22(2), 116-120.
- [16] Crosby, C. Development and implementation of admissions criteria for a clinical decision unit, Touro University, Nevada, In partial fulfillment of requirements for the Doctor of Nursing Practice, Date of Submission: June 20, 2018
- [17] Cooke M. W., Higgins J., Kidd P. (2003). Use of emergency observation and assessment wards: a systematic literature review. *Emergency Medicine Journal*, 20(2), 138-142.
- [18] Hadden, D. S., Dearden, C. H., Rocke, L. G. (1996). Short stay observation patients: general wards are inappropriate. *Journal of Accident & Emergency Medicine*, 13(3), 163-165.
- [19] Groenland, C. N., Termorshuizen, F., Rietdijk, W. J., van den Brule, J., Dongelmans, D. A., de Jonge, E., ... & Jewbali, L. S. (2019). Emergency Department to ICU Time Is Associated With Hospital Mortality: A Registry Analysis of 14,788 Patients From Six University Hospitals in The Netherlands. *Critical Care Medicine*, 47(11), 1564-1571.
- [20] Moitra, V. K., Guerra, C., Linde-Zwirble, W. T., & Wunsch, H. (2016). Relationship between ICU length of stay and long-term mortality for elderly ICU survivors. *Critical Care Medicine*, 44(4), 655-662.
- [21] Fackrell, M. W. (2003). Characterization of matrix-exponential distributions. PhD Thesis, School of Applied Mathematics, University of Adelaide, South Australia.
- [22] Marshall, A. H., McClean, S. I. (2004). Using Coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. *Health Care Management Science*, 7(4), 285-289.
- [23] Gu, W., Fan, N., Liao, H. (2019). Evaluating readmission rates and discharge planning by analyzing the length-of-stay of patients. *Annals of Operations Research*, 276(1-2), 89-108.
- [24] Turgeman, L., May, J., Ketterer, A., Sciulli, R., Vargas, D. (2015). Identification of readmission risk factors by analyzing the hospital-related state transitions of Congestive Heart Failure (CHF) patients. *IIE Transactions on Healthcare Systems Engineering*, 5(4), 255-267.

- [25] Colquhoun, D. and Hawkes, A. G. (1981). On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B*, 211, 205-235.
- [26] Colquhoun, D. and Hawkes, A. G. (1982). On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B*, 300, 1-59.
- [27] Fredkin D. R., Rice J. A. (1986). On aggregated Markov processes. *Journal of Applied Probability*, 23(1), 208-214.
- [28] Qin, F., Auerbach, A., Sachs, F. (1997). Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. Lond. B*, 264(1380), 375-383.
- [29] Karimi, S., Liao, H., Fan, N. (2021). Flexible Methods for Reliability Estimation Using Aggregate Failure-time Data. *IIE Transactions*, 53(1), 101-115.
- [30] Xie, H., Chaussalet, T. J., Millard, P. H. (2005). A continuous time Markov model for the length of stay of elderly people in institutional long-term care. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 51-61.
- [31] Hamdani, Fatima E., et al. (2017). Diagnostic and modeling of elderly flow in a French healthcare institution. *Computers & Industrial Engineering*, 112, 675-689.
- [32] Ross, M. A., Hockenberry, J. M., Mutter, R., Barrett, M., Wheatley, M., & Pitts, S. R. (2013) Protocol-driven emergency department observation units offer savings, shorter stays, and reduced admissions. *Health Affairs*, 32(12), 2149-2156.
- [33] Shi, P., Dai, J., Ding, D., Ang, S. K., Chou, M., Jin, X., Sim, J. (2014). Patient flow from emergency department to inpatient wards: Empirical observations from a Singaporean hospital. <http://dx.doi.org/10.2139/ssrn.2517050>.
- [34] Bunkenborg, G., Bitsch Hansen, T., Holge-Hazelton, B. (2017). Handing over patients from the ICU to the general ward: A focused ethnographical study of nurses' communication practice. *Journal of Advanced Nursing*, 73(12), 3090-3101.
- [35] Bala, M., Kashuk, J. L., Willner, D., Kaluzhni, D., Bdolah-Abram, T., Almqy, G. (2014). Looking beyond discharge: clinical variables at trauma admission predict long term survival in the older severely injured patient. *World Journal of Emergency Surgery*, 9(1), 10.

- [36] Strosberg, D. S., Housley, B. C., Vazquez, D., Rushing, A., Steinberg, S., Jones, C. (2017). Discharge destination and readmission rates in older trauma patients. *Journal of Surgical Research*, 207, 27-32.