Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

DATA-DRIVEN SIMULATION FOR HEALTHCARE FACILITY UTILIZATION MODELING AND EVALUATION

Xuxue Sun
Mingyang Li

Department of Industrial & Management Systems Engineering
University of South Florida
4202 E. Fowler Avenue
Tampa, FL 33620, USA

Chao Meng

Department of Marketing and International Business
Valdosta State University
1500 N. Patterson St.
Valdosta, GA 31698, USA

Nan Kong

Department of Biomedical Engineering
Purdue University
206 S Martin Jischke Dr
West Lafayette, IN 47907, USA

Hongdao Meng
Kathryn Hyer

School of Aging Studies
University of South Florida
13201 Bruce B. Downs Blvd
Tampa, FL 33612, USA

ABSTRACT

Utilization evaluation for healthcare facilities such as hospitals and nursing homes is crucial for providing high quality healthcare services in various communities. In this paper, a data-driven simulation framework integrating statistical modeling and agent-based simulation (ABS) is proposed to evaluate the utilization of various healthcare facilities. A Bayesian modeling approach is proposed to model the relationship between heterogeneous individuals' characteristics and time to readmission in the hospital and nursing home. An ABS model is developed to model the dynamically changing health conditions of individuals and readmission/discharge events. The individuals are modeled as agents in the ABS model, and their time to readmission and length of stay are driven by the developed Bayesian individualized models. An application based on Florida's Medicare and Medicaid claims data demonstrates that the proposed framework can effectively evaluate the healthcare facility utilization under various scenarios.

1 INTRODUCTION

To achieve high quality healthcare service delivery and cost-effective resource management among different healthcare service facilities is one of the important and essential objectives in the current healthcare research and practice. The increasing healthcare demand of the elderly due to the rapid population aging and high prevalence of diseases and disabilities, coupled with the costly and limited resources available in healthcare facilities, pose great challenges in current U.S. healthcare systems to ensure high quality of care. The utilization of healthcare facility is a key measure of healthcare service demand. The pursuit of match between healthcare demand and capacity requires a deep understanding of the relationship between the healthcare facility utilization and the various individual characteristics of aging population. Healthcare administrative claims data, originally generated for administrative and billing purpose, contains important information, such as healthcare time to readmission and length of stay (LoS), which can be leveraged to investigate the healthcare facility utilization. They provide valuable tracking

978-1-5386-3428-8/17/$31.00 ©2017 IEEE

information for admissions and discharges of elderly individuals with different health conditions and demographics. Since healthcare facility utilization may be affected by various individual characteristics and different types of healthcare facilities, such as acute care and long-term care facilities, may also have different influencing factors, it will be desirable to develop data-driven models by analyzing individual's time to readmission and LoS from historical administrative claims data. Compared to the low level clinical data, administrative claims data is less fragmented and contains integrated compatible readmission information among different healthcare facilities, while the detailed individual health conditions, such as physiological measurements, is unavailable in the high level claims data due to privacy issue. These unobserved factors may also affect individuals' healthcare utilization and need to be quantified explicitly. Thus, an efficient and effective statistical model needs to be developed to capture such unobserved heterogeneity and to quantify the influence of observed individual characteristics on different types of healthcare facilities utilization.

Several challenges are involved in healthcare readmission/LoS modeling and claims data analysis. The readmission data in practice exhibits right-skewness. This right-skewness makes the normality assumption in conventional statistical modeling approaches invalid. To address this skewness issue and the influences of different factors, many statistical modeling approaches have been investigated (Bernatz et al. 2015; Jasti 2008). However, these methods only consider heterogeneity induced by observed factors. They ignore the unobserved heterogeneity, which quantifies the effect of unobserved or unmeasurable factors such as the aforementioned physiological information with regarding to the high level claims data. Some approaches are developed recently to address the issue of unobserved heterogeneity (Lee et al. 2012; Kansagara et al. 2011). However, these existing studies mainly employ non-Bayesian estimation method, such as maximum likelihood estimation method. They can only assess the average healthcare utilization over a population but cannot provide an individualized model for every individual care recipient. Non-Bayesian methods also have issues in unknown parameter estimation when sample size is small. In addition, most of the previous studies only consider single type of healthcare facility, and fail to address the multiple types of competing healthcare facilities.

Although great efforts have been taken to develop statistical models to estimate individual patient's time to readmission and LoS in healthcare facilities, evaluating the performance of a complex healthcare system design that consists of multiple individuals is still challenging. Simulation is a powerful tool to study the behavior of complex systems in a dynamic environment. Agent-based simulation (ABS) is an emerging simulation paradigm to study individuals' decision making in various applications, such as transportation (Kim et al. 2017), homeland security (Yuan et al. 2015), and supply chain management (Meng et al., 2014). One of the major advantages of ABS is the capability of modeling individual objects (e.g., patients) and their interactions between each other and with external environment (e.g., healthcare insurance policy). In the healthcare system, patients' behaviors directly affect the utilization of different healthcare resources (e.g., facility and personnel). The allocation of healthcare resources conversely impact the individuals' decisions on healthcare service selection. In addition, the individuals' health condition and service requirements are dynamically changing over time. Therefore, ABS is considered as an effective approach to study complex systems such as healthcare systems.

In order to model healthcare systems in the ABS, the individuals' characteristics and their decisions on healthcare facility selection must be properly defined. To do so, the agents, which are to model the individuals, must be driven by valid statistical models developed based on real data. Therefore, it is critical to integrate ABS and statistical models to produce realistic outputs. In this paper, we propose a data-driven simulation approach that integrates Bayesian analytical modeling and ABS to evaluate the utilization of healthcare facility. We consider two types of healthcare facilities, namely the acute care facility of hospitals and long-term care facility of nursing homes. The proposed statistical models can jointly estimate the facility specific individual observed and unobserved heterogeneity, and capture both within-individual dependency and between-individual independency. The derived healthcare demand can be used as a service metric and performance measure (e.g. length of stay) for the utilization evaluation. For each individual, multiple individual characteristics, such as ethnic group, age, gender, availability of caregiver, and health condition, are considered to determine the time to readmission and LoS in the hospital and nursing home. An ABS model is then developed to model the individuals' readmission and LoS in the hospital and nursing home. The event of individual's readmission and discharge in the simulation model are driven by the Bayesian individualized models. Healthcare facility utilization is defined as the simulation output.

The remainder of the paper is organized as follows. Section 2 discusses the integration of Bayesian modeling approach and agent-based simulation. Section 3 provides a real application to demonstrate the effectiveness of the proposed data-driven simulation approach. Section 4 concludes the paper with suggestions for future work.

2 DATA DRIVEN SIMULATION APPROACH

The proposed data driven simulation approach integrates the Bayesian modeling approach and ABS for healthcare systems with a heterogeneous population. The Bayesian modeling approach is to estimate each individual's time to readmission to different types of healthcare facilities. The ABS is to simulate each individual's readmission and discharge events to estimate the utilizations of different types of healthcare facilities.

2.1 Bayesian Statistical Modeling

Consider a heterogeneous population of N elderly individuals, and individual i can be readmitted to one of M types of healthcare facilities, namely, type m facility, $i=1,\ldots,N$, $m=1,\ldots,M$. Denote the j^{th} time to readmission to type m facility of an elderly individual i as T_{ij}, $j=1,\ldots,c_{mi}$, where c_{mi} is the total number of readmissions to facility with type m of i^{th} individual. To account for the uncertainty and variability of individualized multi-type facility readmission, an advanced statistical modeling approach needs to be developed. Several issues involved in the statistical modeling and analysis of time to readmission data should be addressed: (i) the data right-skewness which invalidates the conventional normality assumption; (ii) the consideration of both within-individual dependency and between-individual independency; (iii) the individualized model of considering both individual facility-specific observed and unobserved heterogeneity; and (iv) the competing risk of individuals requiring multiple types of healthcare facilities.

To simultaneously address the aforementioned issues, a data-driven individualized multi-type healthcare facility readmission model is proposed as

$$r_{im}(t) = r^b{}_m(t)\exp\left(\boldsymbol{\beta}_m{}^T\mathbf{x} + \gamma_{im}\right),\ i = 1,\ldots,N, m = 1,\ldots,M \tag{1}$$

where $r^b{}_m(t)$ is baseline readmission rate to type m healthcare facility of an individual in the absence of facility-specific individual observed and unobserved heterogeneity. $\boldsymbol{\beta}_m$ and $\boldsymbol{x}$ are vectors of facility-specific covariate coefficient and covariates, which represent the facility-specific individual observed heterogeneity. γ_{im} is a random factor and is used to quantify the facility specific individual unobserved heterogeneity. Weibull hazard function can be assumed for $r^b{}_m(t)$ due to its flexibility in modeling right-skewness data, i.e., $r^b{}_m(t) = \lambda^r_m k^r_m t^{k^r_m - 1}$ where λ^r_m and k^r_m are facility specific rate parameter and shape parameter of Weibull distribution respectively. Superscript r denotes "readmission". The overall probability of no readmission to any facilities for individual i is then given by

$$S_i(t) = \exp\left[-\sum_m \lambda^r_m t^{k^r_m} exp\left(\beta_m{}^T\mathbf{x} + \gamma_{im}\right)\right], m = 1,\ldots,M \tag{2}$$

The probability density function of facility specific time to readmission can be expressed as

$$f_{im}(t) = r_{im}(t)S_i(t),\ i = 1,\ldots,N, m = 1,\ldots,M \tag{3}$$

Denote $\mathbf{D}$ as the set of all available data, i.e., $\mathbf{D}=\{t_{ij}, \boldsymbol{x}_i, i=1,\ldots,N; m=1,\ldots,M; j=1,\ldots, c_{mi}\}$, and $\boldsymbol{\Theta}$ as a collection of all unknown parameters, i.e., $\boldsymbol{\Theta}=\{\{\lambda^r_m\},\{k^r_m\},\{\boldsymbol{\beta}_m\}, m=1,\ldots,M\}$. Conventional estimation methods, such as maximum likelihood estimation method, fail to carry out the estimation of all γ_{im}'s, as they will be integrated out in the marginal likelihood function during estimation. To achieve the joint estimation of both the unknown parameters and all γ_{im}'s, the proposed approach carry out the estimation under Bayesian framework. The joint posterior of $\mathbf{\Theta}$ and all γ_{im}'s can be derived as

$$\pi\left(\mathbf{\Theta} \mid \mathbf{D}, \{\gamma_{im}\}\right) = L\left(\mathbf{\Theta} \mid \mathbf{D}, \{\gamma_{im}\}\right)\pi\left(\mathbf{\Theta}\right)\prod_{i}\prod_{m}\pi\left(\gamma_{im}\right) \tag{4}$$

where $\pi(\mathbf{\Theta})$ is the joint prior density function of unknown parameters, and $L(\mathbf{\Theta} \mid \mathbf{D}, \{\gamma_{im}\})$ is the joint likelihood function, defined as $L(\mathbf{\Theta} \mid \mathbf{D}, \{\gamma_{im}\}) = \Pi_i [\, \Pi_m \Pi_j\, r_{im}(t_{ij}) \cdot \Pi_{ai}\, S_i(t)]$, $i=1,\ldots,N$, $m=1,\ldots,M$, and $j=1,\ldots, c_{mi}$, $a_i=1,\ldots, \sum_m c_{mi}$.

$$L\left(\mathbf{\Theta} \mid \mathbf{D}, \{\gamma_{im}\}\right) = \prod_{i=1}^{N}\left[\prod_{m=1}^{M}\prod_{j=1}^{c_{mi}} r_{im}\left(t_{ij}\right) \cdot \prod_{1}^{\sum_m c_{mi}} S_i(t)\right] \tag{5}$$

Markov Chain Monte Carlo (MCMC) method (Roberts et al. 1997) can be employed to obtain the posteriors of all unknown parameters and all γ_{im}'s. The sampling algorithm can be summarized as

Step 0: Initialize $\mathbf{\Theta}$ and $\{\gamma_{im}\}$ as $\mathbf{\Theta}^{(0)} = \{\{\lambda_m^{r(0)}\}, \{k_m^{r(0)}\}, \{\boldsymbol{\beta}_m^{(0)}\}, \{\gamma_{im}^{(0)}\}$. For $\tau=1,\ldots, \tau_{\max}$, repeat Step 1-4:

Step 1: For $i=1,\ldots,N$, $m=1,\ldots,M$, draw $\gamma_{im}^{(\tau)}$ from $\pi(\gamma_{im} \mid \mathbf{D}_i, \{\lambda_m^{r(\tau-1)}\}, \{k_m^{r(\tau-1)}\}, \{\boldsymbol{\beta}_m^{(\tau-1)}\})$

Step 2: For $m=1,\ldots,M$, draw $\lambda_m^{r(\tau)}$ from $\pi(\lambda_m^{r(\tau)} \mid \mathbf{D}, \{k_m^{r(\tau-1)}\}, \{\boldsymbol{\beta}_m^{(\tau-1)}\}, \{\gamma_{im}^{(\tau)}\})$

Step 3: For $m=1,\ldots,M$, draw $k_m^{r(\tau)}$ from $\pi(k_m^{r(\tau)} \mid \mathbf{D}, \{\lambda_m^{r(\tau)}\}, \{\boldsymbol{\beta}_m^{(\tau-1)}\}, \{\gamma_{im}^{(\tau)}\})$

Step 4: For $m=1,\ldots,M$, draw $\boldsymbol{\beta}_m^{(\tau)}$ from $\pi(\boldsymbol{\beta}_m^{(\tau)} \mid \mathbf{D}, \{\lambda_m^{r(\tau)}\}, \{k_m^{r(\tau)}\}, \{\gamma_{im}^{(\tau)}\})$

where $\tau_{\max}$ is the maximum number of iterations in the sampling. It is noticed that the total readmission data of all individuals contribute to the estimation of facility specific unknown parameters, while time to readmission data to any facilities of individual i contribute to the estimation of γ_{im}.

Based on the estimated parameters, individual cumulative risk can be analyzed. The cumulative probability of readmission to a specific healthcare facility m of individual i over time can be represented as

$$F_{im}\left(t\right) = \int^{t} f_{im}\left(v\right)dv \tag{6}$$

where $F_{im}(t)$ is essentially a representation of cumulative incidence function (CIF). The upper limits of $F_{im}(t)$ quantify the eventual probability that readmission to healthcare facility m will happen on individual i. The upper limits of $F_{im}(t)$ approximate to proportion value $c_{mi}/(\sum_m c_{mi})$ where c_{mi} is the number of readmission to facility with type m for individual i, and $\sum_m c_{mi}$ is the total readmission times of individual i.

2.2 Agent-Based Simulation

An ABS model is developed to model the readmission and discharge events of each individual i who has potential healthcare service needs for M types of facilities. In the simulation model, we assume that all the elderly individuals are in the community dwelling status before being readmitted to any healthcare facility. Individuals in this status do not need any healthcare services. However, an individual may need a certain type of healthcare service (e.g., nursing home service) over time, and can be readmitted to the corresponding type of healthcare facility to receive service. An individual's probability of readmission to a certain type of healthcare facility is time-dependent and driven by the individualized readmission model (Eq. 1), meaning that an agent's healthcare service demand is driven by the statistical models. The status of an individual is then changed to the corresponding type of the facility. Then the individual stays in that status for a certain time, and this LoS in that facility is determined by the following equation (Sun et al. 2017).

$$d_i(t) = d^b\left(t\right)\exp\left(\boldsymbol{\alpha}^T\mathbf{y} + \delta_i\right), i = 1,\ldots,N \tag{7}$$

where δ_i is a random variable with distribution function $G(\cdot)$ to characterize each individual's LoS. We have $d^b\left(t\right) = \lambda^l k^l t^{k^l-1}$, where λ^l is the rate parameter and k^l is the shape parameter of the Weibull distribution. The superscript l represents "LoS". After being discharged from the healthcare facility, an individual enters back community dwelling status or exits the simulation when he/she reaches a randomly generated death age. Figure 1 shows the state chart for agents.

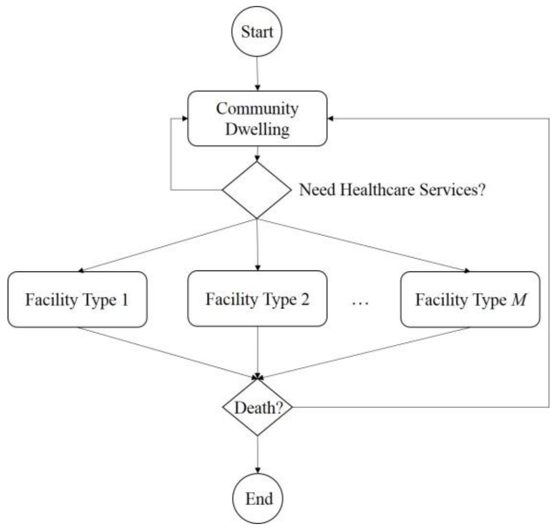


Figure 1: State chart for agents of elderly individuals.

Simulation outputs include the number of individuals in each type of healthcare facility throughout the entire simulation. Animations include the states of agents and plots showing the number of individuals in various healthcare facilities in real-time.

3 APPLICATION

A real case study is presented to demonstrate the effectiveness of the proposed approach. The data used is a subset of Florida claims data, including 217 elder people. Two facility types, namely, hospital and nursing home are considered. The individual characteristics available in the data set include binary variables, such as gender indicator, as well as categorical variables, such as Assisted daily living (ADL) total score, and continuous variables, such as age. Classical variable selection methods, such as random forest based method, can be used to select the significant covariates. Five covariates are selected in the proposed model, x_1 for indicator for white ethnic group, x_2 for age, x_3 and x_4 representing female indicator and caregiver indicator respectively, and x_5 is score for total help in activities of daily living. Ten facility specific covariates coefficients are involved in the model. Totally 14 unknown parameters and 217 individual γ_{i1}'s, γ_{i2}'s can be jointly estimated.

Table 1: Estimation results of model parameters.

Parameters	Posterior mode	2.5%	25%	50%	75%	97.5%
k_1	1.36538	1.19000	1.29700	1.35800	1.42700	1.58200
k_2	1.10523	0.94020	1.04600	1.10300	1.16100	1.28400
λ_1	0.00102	0.00013	0.00053	0.00104	0.00199	0.00681
λ_2	0.00102	0.00007	0.00042	0.00106	0.00255	0.01302
β_{11}	-0.46868	-0.83680	-0.59130	-0.46760	-0.34450	-0.10750
β_{12}	-0.37511	-0.89031	-0.54450	-0.37440	-0.20400	0.13600
β_{22}	0.02201	-0.00812	0.01171	0.02188	0.03220	0.05241
β_{52}	0.74749	0.22840	0.56680	0.74540	0.92540	1.28500

The estimation is carried out under Bayesian estimation framework. As no prior knowledge is available, non-informative priors are assumed for parameters. To be specific, normal priors are assigned for all facility specific covariates coefficients, and inverse gamma distributions are assumed for shape parameters k_1 and k_2. The point estimate and interval estimate of some estimated unknown parameters are presented and summarized in Table. 1 for illustration purpose. The facility specific shape parameter k is larger than 1, which indicates that both the individualized instantaneous probability of readmission to hospital and nursing home will increase over time. λ_1 and λ_2 can be interpreted as the baseline average hospital readmission and nursing home readmission respectively, in the absence of effect from covariates and individualized unobserved heterogeneity. It is observed that β_{11} has strong negative effect on hospital readmission, which implies that the "caucasian" ethnic group is less likely to be readmitted to hospital, compared to other ethnic groups. Although β_{12} is not significantly different from zero at 95% credible interval, it still contains rich information. β_{12} has negative effect at 50% credible, and shows strong concentration on negative values. This implies that the "white" ethnic group is less likely to be readmitted to nursing home. β_{22} neither has a significant effect at 95% credible interval, however, it is significantly positive at 50% credible interval. When the individual get older, the probability of readmission to nursing home becomes higher. β_{52} is significantly positive at 95% credible interval, which infers that the individual who needs more help in activities of daily living is more likely to be readmitted to nursing home and has shorter time to readmission to nursing home, from previous facilities (either hospital or nursing home).

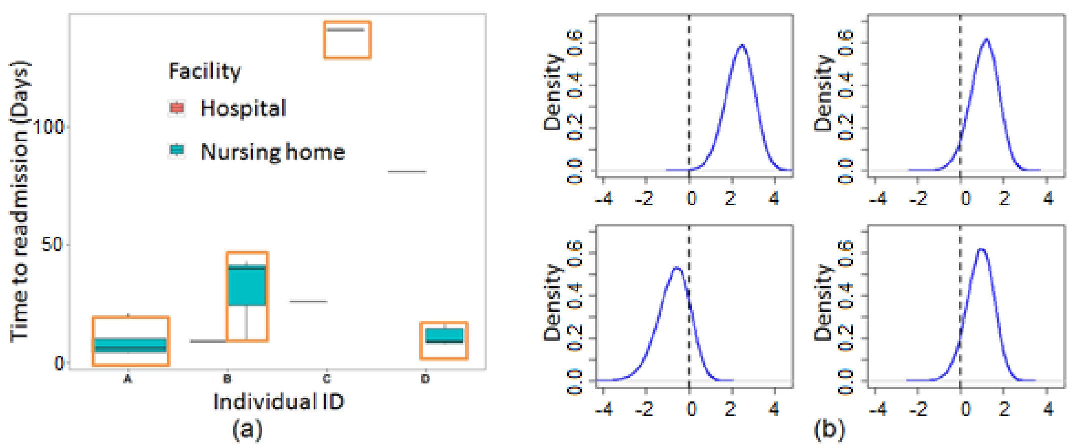


Figure 2: (a) Readmission time of four individuals, (b) Density plot of individualized nursing home specific unobserved heterogeneity, top left to right for individual A and B, bottom left to right for individual C and D respectively.

The positive value of facility specific individual unobserved heterogeneity indicates that individual is more likely to be readmitted to a specific facility, and thus has a shorter time to readmission, and vice versus. As shown in Figure 2 (a), individual C has relative longer time to readmission to nursing home, compared to individual A, B and D. Individual A has relative shortest time to readmission to nursing home. In the density plots of γ_{i2}'s for the four individuals, as illustrated in Figure 2 (b), γ_{C2} concentrate on negative values, and it indicates that individual C is less likely to be readmitted to nursing home, and thus has longer time to readmission to nursing home. All of the γ_{i2}'s for other individuals show a strong positive value concentration, and γ_{A2} has largest positive posterior mode among all individuals. The results inferred from γ_{i2}'s are consistent with the real case, that is, individuals A, B, D are more likely to be readmitted to nursing home and has relative shorter time to readmission to nursing home than individual C, and individual A is relatively least likely to be readmitted to nursing home due to the effect from strongest positive concentration and largest posterior mode of γ_{i2}'s.

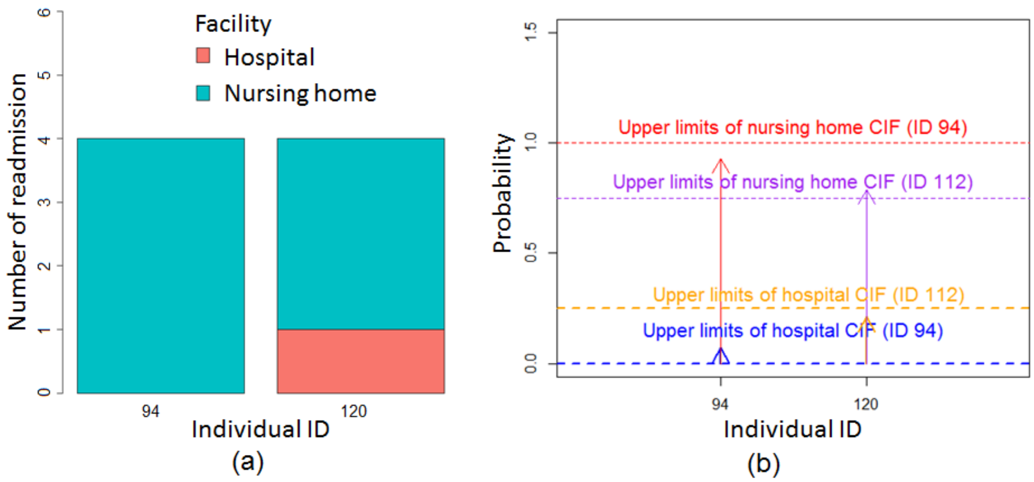


Figure 3: Upper limits validation of hospital and nursing home cumulative incidence function.

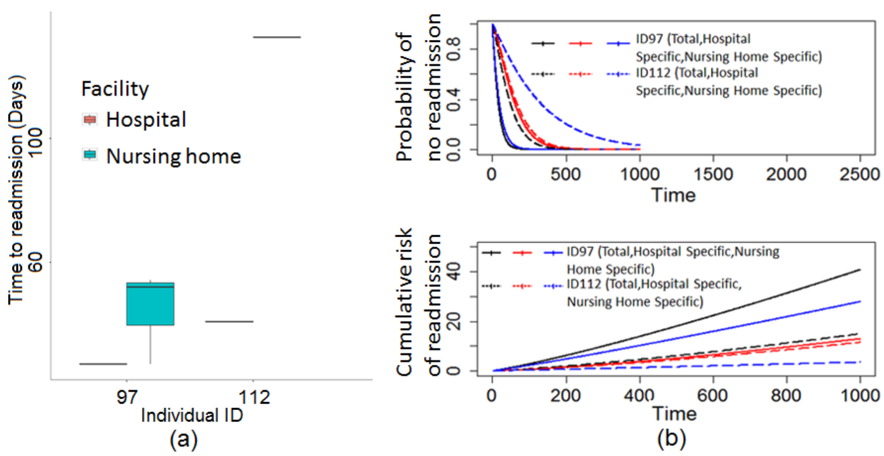


Figure 4: individualized risk analysis.

Based on the estimated unknown parameters and facility specific individual unobserved heterogeneity, individualized competing risk analysis can be carried out. The facility specific individual cumulative incidence function can be obtained and the upper limits of hospital CIF and nursing home CIF can be calculated. The upper limits theory in context of readmission can be interpreted as upper limits of specific healthcare facility CIF should equal the proportion of number of readmissions to corresponding healthcare facility over total readmissions number. As shown in Figure 3, the upper limits of nursing home CIF for both two individuals approximate to the proportion of number of readmission to nursing home over total number of readmissions respectively. Upper limits of hospital CIF can also be validated. The individualized cumulative risk curve and probability of no readmission over time can be obtained. As shown in Figure 4 (b), the total, hospital specific and nursing home specific cumulative risks of individual ID#97 are all larger than the ones of individual ID#112, which are consistent with real data in Figure 4 (a).

The corresponding ABS model is then developed in AnyLogic. Based on the data used in our case study, each agent in the simulation model has three possible states (see Figure 5), which are Com_dwelling, Nursing_home, and Hospital, representing the agent being in the community (no healthcare service needed), nursing home, and hospital, respectively. The data driven simulation model is flexible enough to incorporate other additional facility types and individual attributes if the corresponding data is available.

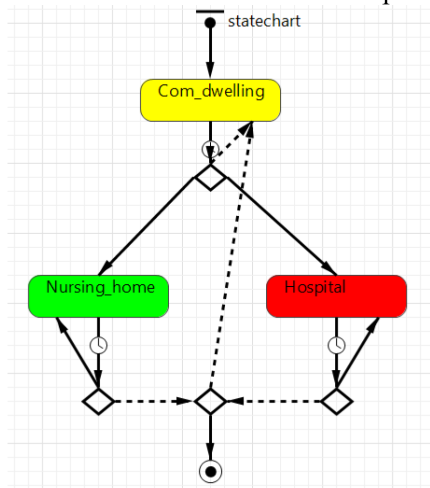


Figure 5: State chart for agents.

In the experiment, we conducted 2 scenario based analyses. In Scenario 1, 3 geographic regions (Regions 1-3) with different population age distribution are modeled to study the impact of population age on healthcare facility utilization. In Scenario 2, another 3 geographic regions (Regions 4-6) with different ethnic groups are modeled to study the impact of ethnicity on healthcare facility utilization. Table 2 shows the coefficients of the individual characteristics in the 6 designed regions. The readmission probability of individuals is defined by Eq. 1 and the data in Table 1. The LoS of individuals are defined by Eq. 7. Table 3 shows the variables of individual characteristics as simulation inputs. y_1, y_2 and y_3 are dummy variables representing ethnic groups of "White", "African American", and "Hispanic", and the baseline ethnic group is "other", while y_4, y_5, and y_6 represent age, gender, and Charlson comorbidity index, respectively

Table 2: Coefficients of individual characteristics for LoS in simulation.

Coefficients	Nursing home	Hospital
α_1	0.31793	0.31793
α_2	0.24265	0.24265
α_3	-0.06581	-0.06581
α_4	0.00215	0.00215
α_5	0.07670	0.07670
α_6	-0.03613	-0.03613
k^l	1.13879	1.93879
λ^l	0.00958	0.00958
δ_i	uniform(-1,2)	uniform(-1,2)

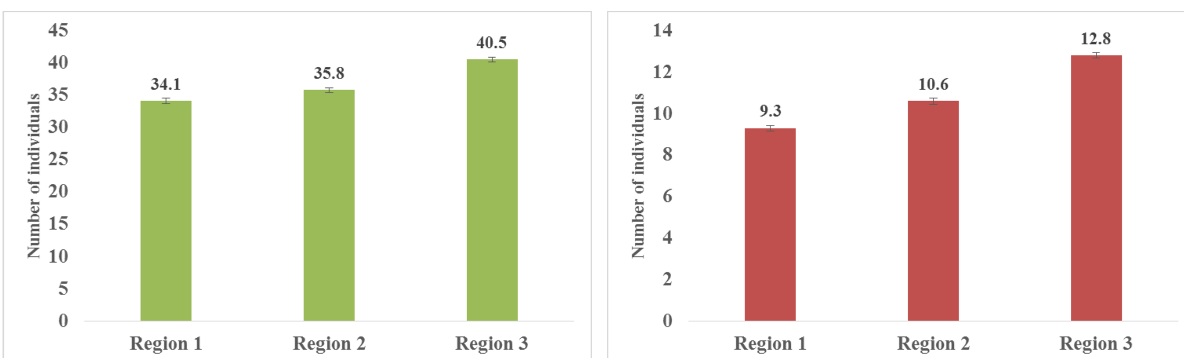
Table 3: Variables of individual characteristics in simulation.

Variables	**Region 1**	**Region 2**	**Region 3**
x_1	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
x_2, y_4	uniform_discr(45,55)	uniform_discr(55,65)	uniform_discr(65,75)
x_3	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)

x_4	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
x_5	uniform_discr(0,5)	uniform_discr(0,5)	uniform_discr(0,5)
$[y_1, y_2, y_3]$	uniform_discr(0,1) and $y_1+y_2+y_3=1$	uniform_discr(0,1) and $y_1+y_2+y_3=1$	uniform_discr(0,1) and $y_1+y_2+y_3=1$
y_5	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
y_6	uniform_discr(0,5)	uniform_discr(0,5)	uniform_discr(0,5)
Variables	**Region 4**	**Region 5**	**Region 6**
x_1	1	0	0
x_2, y_4	uniform_discr(65,75)	uniform_discr(65,75)	uniform_discr(65,75)
x_3	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
x_4	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
x_5	uniform_discr(0,5)	uniform_discr(0,5)	uniform_discr(0,5)
$[y_1, y_2, y_3]$	[1, 0, 0]	[0, 1, 0]	[0, 0, 1]
y_5	uniform_discr(0,1)	uniform_discr(0,1)	uniform_discr(0,1)
y_6	uniform_discr(0,5)	uniform_discr(0,5)	uniform_discr(0,5)

Each region is designed to have 150 individuals (agents) that may need healthcare services from nursing home and hospital. The steady state of simulation was reached in about 3 years for each scenario. 5 years were simulated and 1865 samples were collected. All the simulation results for 6 regions are statistically significant.

Figure 6 show the facility utilization results for Scenario 1. We can see that as the average age of the population increases, the utilization of both nursing homes and hospitals increases. However, when the individuals are relatively young (e.g., in Regions 1 and 2), their utilization of nursing home service shows a slower increase compared to that from older individuals. These results indicate that the healthcare facilities in the region with an older population tend to have a higher utilization.

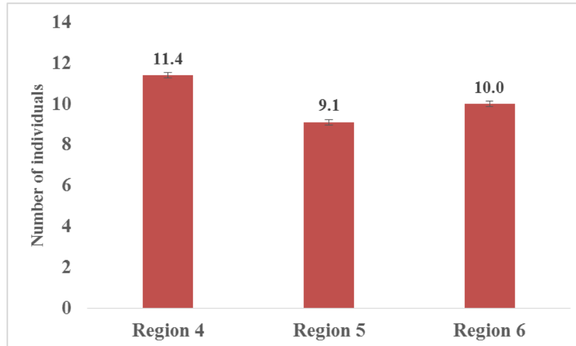


(a) Number of individuals in nursing home. (b) Number of individuals in hospital.

Figure 6: Number of individuals using healthcare services in Regions 1-3.

Figure 7 show the facility utilization results for Scenario 2. We see that white people (Region 4) tend to stay in the nursing home relative to African American and Hispanic people. When comparing African American people (Region 5) and Hispanic people (Region 6), African American people tend to use nursing homes, while Hispanic people tend to go to hospitals. These results reveal that healthcare facility utilization is affected by the ethnicity.

(a) Number of individuals in nursing home.



(b) Number of individuals in hospital.

Figure 7: Number of individuals using healthcare services in Regions 4-6.

4 CONCLUSIONS

In this paper, a data-driven simulation approach is proposed to model and to evaluate healthcare facility utilization for both acute care facility and long-term care facility. In the proposed approach, a Bayesian statistical modeling approach is proposed to analyze complex data structures of individuals' time to readmission and LoS observations by incorporating heterogeneous observed (e.g., gender) and unobserved individual characteristics based on real administrative claims data. An ABS model is further developed to study the readmission and discharge events of individuals based on the individualized Bayesian statistical models. In the developed ABS model, the agents mimic the individuals who have potential needs for healthcare services. The agents' readmissions and LoS are driven by the developed Bayesian statistical models. In the conducted experiments, 2 scenarios are studied to demonstrate the impact of population age and ethnicity on healthcare facility utilization. The statistical models are developed based on Florida's Medicare and Medicaid claims data. . The simulation results reveal that 1) the healthcare facilities in the region with an older population tend to have a higher utilization, and 2) the healthcare facility utilization is affected by the ethnicity. Our future research works are to study additional observed factors for statistical modeling and incorporate the impact of healthcare facility capacity and payment policies, such as reimbursement policies, on individuals' healthcare service selection to produce more realistic results.

ACKNOWLEDGMENTS

This work was supported in part by University of South Florida Research & Innovation Internal Awards Program under Grant No. 0114783.

REFERENCES

Bernatz, J.T., J.L. Tueting, and P.A., Anderson. 2015. "Thirty-day Readmission Rrates in Orthopedics: a Systematic Review and Meta-analysis". *PloS one* 10(4):e0123593.

Jasti, H., E.M.Mortensen, D.S. Obrosky, W.N. Kapoor, and M.J. Fine. 2008. "Causes and Risk Factors for Rehospitalization of Patients Hospitalized with Community-acquired Pneumonia". *Clinical Infectious Diseases* 46(4):550-556.

Kansagara, D., H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani. 2011. "Risk Prediction Models for Hospital Readmission: a Systematic Review". *Jama* 306(15): 1688-1698.

Kim, S., Y. J. Son, Y. C. Chiu, M. A. B. Jeffers, and C. D. Yang. 2016. "Impact of Road Environment on Drivers' Behaviors in Dilemma Zone: Application of Agent-based Simulation". *Accident Analysis & Prevention* 96:329-340.

Lee, E.W. 2012. "Selecting the Best Prediction Model for Readmission". *Journal of Preventive Medicine and Public Health* 45(4):259-266.

Meng, C, Y. J. Son, and B. Hu. 2014. "Capacity Reservation for a Decentralized Supply Chain under Resource Competition: a Game Theoretic Approach." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2036-2047. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc..

Roberts, G. O., and S. K. Sahu. 1997. "Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59(2):291-317.

X. Sun, H. Meng, K, K. Hyer and M. Li. 2016. "Bayesian Data Analytics for Personalized Hospital Length-of-Stay Modeling and Analysis." In *Proceedings of the Industrial and Systems Engineering Research Conference*, Anaheim, CA., May 21-24.

Yuan, Y., Z. Wang, M. Li, Y. J. Son, and J. Liu. 2015. "DDDAS-based Information-aggregation for Crowd Dynamics Modeling with UAVs and UGVs". *Frontiers in Robotics and AI* 2: 8.

AUTHOR BIOGRAPHIES

XUXUE SUN is current PhD student (2016 spring) in the Department of Industrial & Management Systems Engineering at the University of South Florida. Her research interest is Bayesian data analytics, statistical modeling in healthcare system and reliability engineering. Before joining USF, she had research experiences in biomedical signal processing from institutes at Chinese Academy of Science and RWTH Aachen university, and coding experience from IBM. Her e-mail address is xuxuesun@mail.usf.edu.

CHAO MENG is a lecturer of logistics and supply chain management at Valdosta State University. He received his Ph.D. in Systems and Industrial Engineering and M.S. in Industrial Engineering from the University of Arizona in 2016. He also received his M.S. in Logistics Engineering and B.S. in Logistics Engineering from Southwest Jiaotong University in 2010 and 2008. His research interests include supply chain coordination, logistics network optimization, and multi-paradigm simulation. He is a member of INFORMS, IISE, POMS, and SIGSIM. His email address is cmeng@valdosta.edu.

NAN KONG is an Associate Professor of Biomedical Engineering at Purdue University. He received his PhD in Industrial Engineering from the University of Pittsburgh in 2006. His research expertise involves applying OR/MS techniques, and quantitative analytics in general, to healthcare systems operations and medical decision making problems. His other research expertise includes stochastic integer programming. He is a member of INFORMS, IISE, and MDM. His email address is nkong@purdue.edu.

HONGDAO MENG is Associate Professor in School of Aging Studies at the University of South Florida. He has expertise in geriatrics, health economics, and long-term care services research. He has acquired considerable expertise in assessment of physical and mental health conditions, healthcare utilization and expenditures in LTC systems. His email address is meng@usf.edu.

KATHRYN HYER is Professor and Director from Florida Policy Exchange Center in School of Aging Studies at the University of South Florida. Her research interest is to study the impact of State and federal regulations on assisted living and nursing home residents' quality of care and quality of life, especially at the end of life. Her research includes program evaluation with a special emphasis on evaluation of geriatric educational interventions for health professionals and students. Her email is khyer@usf.edu.

MINGYANG LI is an assistant professor in the Department of Industrial & Management Systems Engineering at the University of South Florida. He received his Ph.D. in Systems & Industrial Engineering

and his M.S. in Statistics from the University of Arizona in 2015 and 2013, respectively. He also received his M.S. in Mechanical & Industrial Engineering from the University of Iowa in 2010 and his B.S. in Control Science & Engineering from Huazhong University of Science and Technology in 2008.His research interests include reliability and quality assurance, Bayesian data analytics, and system informatics. He is a member of INFORMS, IISE, and ASQ. His e-mail address is mingyangli@usf.edu.