Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

SIMULATING WAIT TIME IN HEALTHCARE: ACCOUNTING FOR TRANSITION PROCESS VARIABILITY USING SURVIVAL ANALYSES

Scott Levin

Johns Hopkins Medicine and Engineering
Billings Ste. 400, 600 N. Broadway
Baltimore, MD 21287, USA

Maxim Garifullin

GE Healthcare
540 W Northwest Hwy,
Barrington, IL 60010, USA

ABSTRACT

Wait or queuing time is a principal performance measure for many discrete-event simulation (DES) models in healthcare. However, variation in wait time is often caused by both occupied downstream servers (e.g., beds) and organizational and human transition processes. DES models that attribute wait solely to occupied servers, ignoring transition process variability, face challenges in adequate baseline validation. Embedding regression models for survival data in DES to estimate patient wait times is a method capable of integrating the effects of transition processes with queuing. Developing these models as a sub-component is further valuable in understanding the socio-technical system factors that drive prolonged waits. These general methods are exhibited in a DES for a large urban hospital with a primary output of wait time in the emergency department (ED) for transfer to an inpatient bed (boarding time). Simulated boarding time is compared before and after accounting for transition processes using survival analysis.

1 INTRODUCTION

DES models for healthcare facilities commonly focus on improving wait time, patient flow and management of capacity (Hamrock et al. 2014; Jacobsen et al. 2006). Although DES is adept at modeling the complex queuing structure for patients in healthcare environments, transition process variation driven by organizational and human factors is more difficult to capture mathematically. For example, analyses of patient location data used to construct DES models may find that patients are consistently waiting for servers (e.g., beds, imaging suites, clinicians) at time-points despite their availability. In the DES, queued patients would efficiently shift to open servers. However in clinical practice, transition process factors such as inefficient communication, lack of awareness of server availability, complex administrative guidelines, interruptions, and cumbersome documentation create further delays (Shi it al. 2015; Armony et al. 2010). These delays are not inherent to queuing nor well understood from time-stamped patient flow data alone. To fully capture the dynamics of healthcare facilities or any flow-based socio-technical system, transition process variability should be understood.

Not accounting for these processes can lead to results that severely under-estimate waiting. Moreover, DES wait time distributions may be difficult to validate against the observed healthcare system. To achieve sufficient validation, the model developer may be motivated to input additive time intervals to patients at transition points that are drawn from a distribution representing the difference between the current model and observed waits (Shi it al. 2015). A more in-depth approach, borrowed from lean methods, may motivate the model developer to map out the transition process and measure or elicit expert estimates of time distributions for each component; independent value of this investigation exists (Kang et al. 2014; Simon and Canacari 2012). There is debate concerning the appropriateness of these solutions to improve DES wait time validation, which depend upon the: (1) importance of wait time in relation to the overarching DES objectives, (2) proportion of the wait attributed to queuing versus transition processes, and (3) desired level of DES model detail and resources to accomplish.

Regardless of degree of detail, inputting additive wait times to account for transition process may fail to capture important system dynamics. In healthcare systems, patient characteristics (i.e., attributes), staffing characteristics, temporal factors, service lines, operational workload, and other factors may drive variation in transition process times. In this paper, we propose embedding regression models for survival data within the DES to account for these factors effects on delays in transition; preserving systems dynamics. This method is capable of estimating wait times by incorporating transition processes time with pure queuing. We will demonstrate these concepts by example, using a DES built for a large urban hospital. We will focus on the output wait time (i.e., boarding time) in the emergency department (ED) for admitted patients to transfer to an inpatient bed. The objectives are to:

1. Exhibit the original DES model and results without transition process time integrated.
2. Show development of regression models for waiting and mechanisms to embed within DES.
3. Highlight the final DES model and results accounting for variation in transition processes.

It is important consider that our purpose is to illustrate these general methods through an example. There are many ways to apply these same principles optimally specific to the DES being constructed.

2 ORIGINAL SIMULATION MODEL WITHOUT TRANSITION PROCESS VARIABILITY

The DES was developed for a 1000-bed urban, academic hospital to inform patient flow management strategies. The DES was built directly from patient flow data (i.e., arrival patterns, routing, lengths-of-stay) collected over 8-months from the inpatient admit-discharge-transfer (ADT), ED, operating room (OR), and direct access (DA) hospital information systems. These data were used to capture end-to-end patient flow across all inpatient units. ED and operating room (OR) pathways were explicitly modeled because they were the major sources of inpatient admissions. Patients were the only entities flowing in the DES. Patient flow routing and timing was probabilistically determined for each patient based on admission source (e.g., OR, ED, DA), care-level sequence (e.g., intensive care), and service line (e.g., hospitalist, pulmonary, gastroenterology) attributes. Inpatient beds were the only resources used to process patient entities; no staffing resources were included. The simulation was built within General Electric's (GE) Hospital of the Future™ simulation environment (GE Healthcare 2011).

2.1 Modeling Emergency Department Boarding Time

ED patients admitted to the hospital often experience significant waits to be transferred to an inpatient bed. This wait time from hospital admission decision to arrival to an inpatient unit is referred to as boarding time (American College of Emergency Physicians 2011). Excessive boarding is a common cause of ED crowding and important marker of overall hospital patient flow (Institute of Medicine 2006; Joint Commission 2012). Modeling the transition of patients from the ED to inpatient units, particularly the inpatient unit assignment decision, can be complex based on hospital size and clinical admission policies (Kang et al. 2014). Figure 1 depicts the simple process modeled in the ED. For hospital admitted patients, time from arrival to admission decision (i.e., treatment) was an average of 6.3 hours (median 5.1; IQR 3.2 – 8.0). Total boarding time was an average of 4.9 hours (median 4.0; IQR 2.8 – 6.4). In the DES, ED admissions are first assigned a pool of eligible downstream inpatient units (Figure 1). The priority for these units are constructed from transition probabilities observed in our historical patient flow data based on patients' attributes. Attributes determining inpatient unit included care-level (e.g., critical care, intermediate care, acute care) and clinical service (e.g., Medicine, Cardiology, Neurology). Patients transfer to the highest priority eligible unit based on the strength of priority (i.e., transition probability) and occupancy relative to other units. This logic was designed to imitate local optimum bed assignment decisions reflecting actual hospital processes.

Despite the simple transition depicted in Figure 1, the DES model of patient flow through the hospital (inpatient) was comprised of 94 distinct patient types based on the attributes described. These patient types have a unique combinations of arrival patterns, routing, and length-of-stay distribution. The arrival patterns account for hour-of-day, day-of-week, and seasonal effects. Routing is probabilistically determined based on patient type. Patients may follow one of 2,800+ unique pathways. An example pathway may be arrival from the ED to the ICU then to medicine unit A, then medicine unit B, then discharged. Each pathway step has a unique length-of-stay distribution again based on patient type.

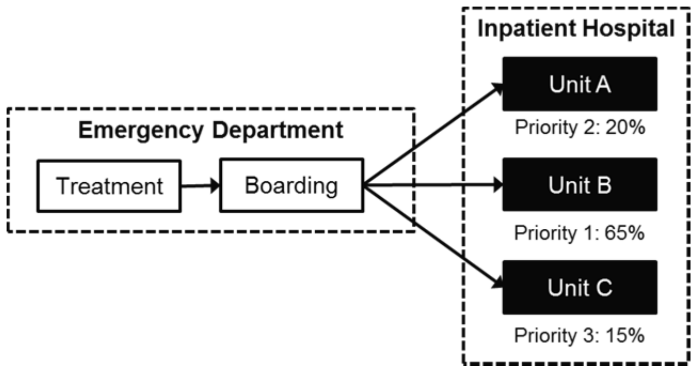

Figure 1: Hospital unit assignment for patients boarded in the emergency department.

For the sub-set of patients admitted to general medical units (52% of total ED admissions), we investigated the relationship between boarding time and aggregate medicine unit occupancy (Figure 2). The median total boarding time was consistently around 4 hours when occupancy was less than 85%. Above 85%, a proportional relationship between occupancy and boarding time emerges; the most substantial increase not being occurring until above 95% occupied. Figure 2 provides evidence that a major contributor to total boarding time was transition process-related.

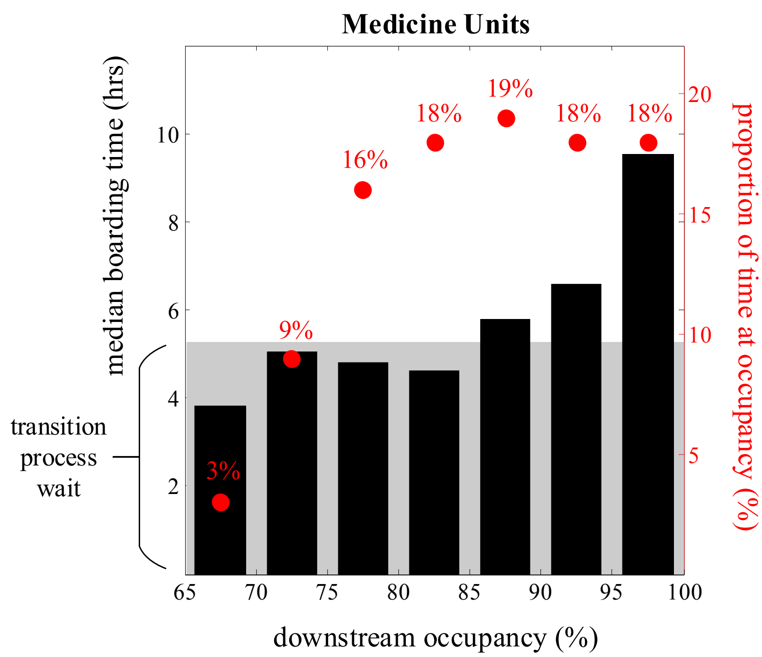

Figure 2: Relationship between boarding time and downstream medicine units occupancy.

Ignoring transition process not surprisingly results in severe under-estimates of boarding time. These original model results compared to observed may be seen in Figure 3. Assuming an efficient queuing system meant almost no waiting and created further validation challenges for ED operational outputs such as census levels and timing of hospital admissions critical to the model. This motivated us to model patient and system factors that may be associated with transition processes using survival analyses.

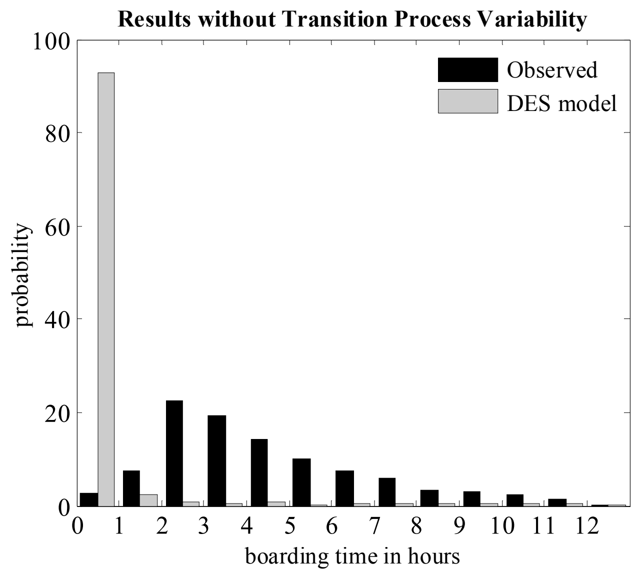

Figure 3: Boarding time distribution comparison for observed versus DES model results.

3 REGRESSION MODEL DEVELOPMENT AND INTEGRATION INTO SIMULATION

3.1 Regression models for survival data

Survival analysis is a set of statistical methods that aim to understand time duration. Examining time-to-death for study populations or time-to-failure for mechanical systems (also called reliability analysis) are typical applications. Regression models may be fit to survival data to determine covariates (independent variables) relationship to time-to-event (dependent variable) (Hosmer et al. 2008). Survival regression models come in both parametric and semi-parametric forms. Parametric models place a known functional form on the event time distribution (e.g., Exponential, Weibull, Gompertz, Log-Normal, and Gamma). Advantages include ease of estimation and flexibility to perform additional analyses (e.g., random effects) (Cook 2008). The semi-parametric approach, Cox proportional hazard model, may be best suited for DES applications because there are no assumptions regarding the underlying event time distribution. Given the hazard function $\boldsymbol{\lambda}$ determining the event rate at time $\boldsymbol{t}$ conditional on surviving until time $\boldsymbol{t}$, the Cox proportional hazard model is specified as (Cox 1972):

$$\lambda(t|x) = \lambda_o(t) exp(B_1 x_1 + \cdots + B_k x_k)$$

Covariates $\boldsymbol{x}$ enter the model linearly (parametric component) and are assumed to have a proportional effect on the hazard function $\boldsymbol{\lambda(t \backslash x)}$ over time (Hosmer et al. 2008). This assumption may be tested using regression diagnostics (Grambsch and Therneau 1994). If violated, there are several means of addressing non-proportional hazards out of the scope of this article (Grambsch and Therneau 1994; ,Schemper 1992; Cox 1972). The Cox regression approach was used for our hospital DES.

3.2 Cox Regression Models for Emergency Department Boarding

We hypothesized that queuing measures, patient care-level and time-of-day would drive transition process time. This hypothesis was based upon information gained from semi-structured interviews of ED and inpatient clinicians in concert with basic preliminary examination of boarding time data. These covariates were extracted from clinical information systems for each patient at time of admission decision (i.e., start of boarding time). A Cox proportional hazard regression model was built from these data and is displayed in Table 1. Clinical department occupancy was the proportion of beds occupied for the specific department (e.g., Cardiology, Medicine, Neurology) managing the admitted patient. Unit occupancy is the proportion of beds occupied within the specific unit the patient transferred to (i.e., highest priority unit). Clinical department occupancy and unit occupancy was on average 81.9% (95% Percentile 67.5% to 97.5%) and 82.5% (95% Percentile 53.3% to 100.0%). Correlation between these occupancy measures was not prohibitive (ρ= 0.52). Number of ED admissions in the last 4 hours captured the busyness of admissions personnel. This queuing measure was an average of 2.88 (95% Percentile 0 to 8). Patient care-level assessed whether the patient was admitted to a general acute care (89.5%), intermediate (4.25%), or critical care unit (6.27%). Time-of-day was also relevant because of changes in hospital staffing levels and ancillary service availability by hour. It's important to note that each of these same covariates (Table 1) were measurable in the simulated hospital environment.

Table 1: Cox proportional hazard regression model for emergency department boarding.

Covariate	*B*	Standard Error	Hazard Ratio	P-Value
Queuing Measures				
Clinical department occupancy	-4.623	0.213	0.010	<0.001
Unit occupancy	-0.478	0.111	0.620	<0.001
Number of ED admission requests in last 4 hrs	0.041	0.006	1.042	<0.001
Patient Care-Level [Ref: general acute care]				
Critical care	0.543	0.047	1.722	<0.001
Intermediate care (e.g., step down unit)	0.561	0.057	1.752	<0.001
Time-of-Day [Ref: morning 6:00 – 11:59]				
Afternoon 12:00 – 17:59	-0.300	0.039	0.741	<0.001
Evening 18:00 – 23:59	-0.336	0.035	0.715	<0.001
Night 0:00 – 5:59	-0.327	0.034	0.721	<0.001

Hazard ratios for each covariate (Table 1) are interpreted as shifting the baseline probability density function for wait time. Covariates with hazard ratios greater than one create an inverse relationship in wait time duration. For example, patients admitted to critical care units (HR 1.722) have higher likelihood of rapid transfer (decreased wait) compared to patients admitted to acute care beds (reference). Conversely, covariates with hazard ratios less than one indicate a lower likelihood of early transfer. For example, as clinical department occupancy increases (HR 0.010) so does wait time distribution.

Ultimately the Cox regression model (Table 1) is used to estimate a hazard ratio for each patient given their set of covariates $\boldsymbol{x}$ at admission decision time. The full distribution of hazard ratios for all patients may be seen in Figure 4 (Top) below. The relationship between a patient's hazard ratio and their survivorship function (SF) and probability density functions (PDF) are also seen in Figure 4 (Bottom). Compared to baseline, low hazard ratios alter the PDF for higher probability of pro-longed waits (red line) and high hazard ratios have the reverse effect (blue line).

3.3 Embedding Regression Model into the Discrete Event Simulation

The Cox regression model was implemented in the DES to generate a boarding time for each patient. Specific simulation logic for each patient is as follows (Figure 5):

1. Capture queuing, patient care-level, and timing covariates at the admission decision time.

2. Input covariates to Cox regression model (Table 1) to generate a patient-specific hazard ratio.
3. Map the hazard ratio to an individual SF.
4. Draw a boarding time from the patient-specific SF to assign and execute in the DES.

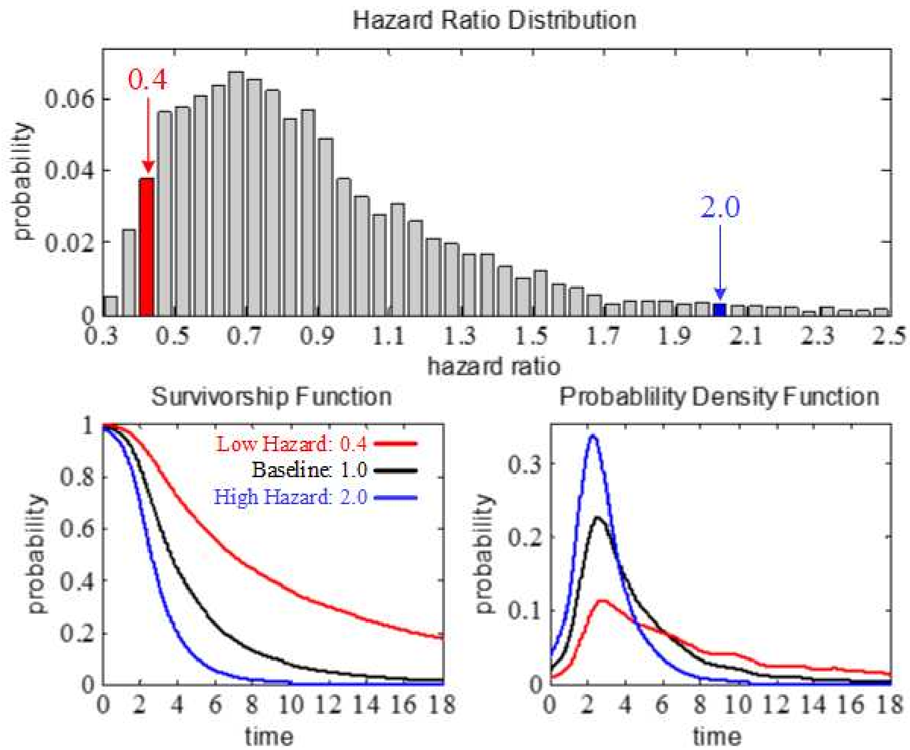

Figure 4: Hazard ratio distribution and translation to patient-specific probability density function.

Computing individual SFs per patient during a DES model run is computationally inefficient and may not be feasible. This was the case for our hospital DES due to the number of ED admissions (~20,000 patients) per model run, simulated run time (365 days), and multiple replications (30 runs) required to produce robust results. To address, we cached and stored a set of SFs for hazard ratios in 0.1 increments over the full range. These functions were stored as a series of points (survival time, survival probability) in the DES database. This effectively mapped a 0.1 length bin of hazard ratios to a unique SF. Upon DES initiation, the set of SFs were retrieved from the database and stored in memory as a hash map. During the simulation run, a random number between 0 and 1 was generated for each patient at admission decision time. The random number was translated to a boarding time by iterating through the hazard ratio matched SF (database row) until the cumulative probability of the SF was greater than the random number. The corresponding time was executed in the DES as the patient's boarding time. Figure 5 outlines integration of the Cox regression model, most notably the patient to SF mapping procedure (3). This mapping process is likely required to maintain computational efficiency (i.e., minimize simulation run time) and provides a framework executable in many DES environments.

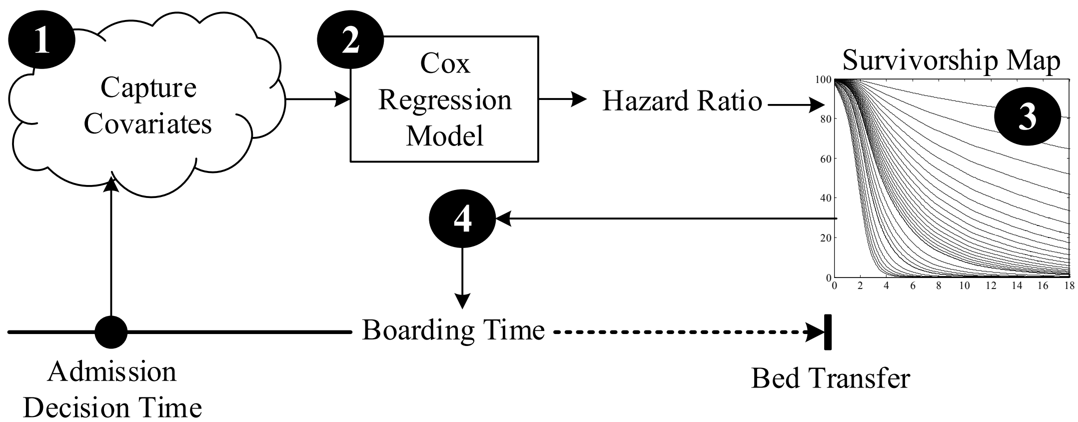

Figure 5: Embedding Cox regression model with the DES to compute boarding time.

4 FINAL SIMULATION RESULTS WITH TRANSITION PROCESS VARIABILITY

Incorporating Cox regression-based estimates of boarding time in the hospital DES enabled successful validation. DES boarding time output closely matched the observed distribution from our dataset (Figure 6) and ED-to-inpatient bed transfer times were accurately reflected. Further, DES hour-by-hour output census levels for the ED and all inpatient units matched actual census figures within 5%. This was an acceptable level of accuracy determined by hospital stakeholders and DES developers.

The DES was used to test multiple scenarios to improve patient flow. One realistic scenario was shifting 13 unused medicine beds from the older section of the hospital to the new area where they may be actively used. This would add 13 adult medicine beds to the 187 already in operation, representing a 7% increase in capacity. These additional beds were to be scattered to individual units (no more than 1 additional bed per unit) and were hypothesized by hospital stakeholders to have a large impact on boarding time. However, because of transition processes, there were minimal effects on boarding time. At DES model baseline, boarding time was an average of 4.9 hours (median 3.9; IQR 2.7 – 6.1) closely matching historical data (Figure 6). Results of adding 7% capacity was an average boarding time of 4.6 hours (median 3.5; IQR 2.5 – 5.5) which was much less of a change than anticipated. Hospital stakeholders agreed that effort directed at improving the efficiency of the transition process may have more impact than increasing capacity.

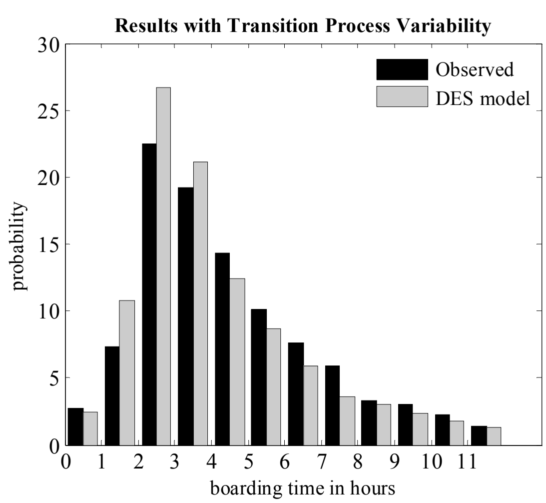

Figure 6: Boarding time distribution comparison for observed versus DES model results.

5 DISCUSSION

A method to embed regression models for survival data in DES to account for variability in transition process time has been demonstrated. In socio-technical environments, such as healthcare, mathematically modeling the effects of both queuing and transition process variation may be required to accurately understand waiting and overall system dynamics. Regression models for survival data (parametric or semi-parametric) are capable of assessing the hypothesized drivers of wait time which may include queuing measures, patient characteristics, staffing levels, timing factors, clinical service lines, operational workload and other factors. For example in our simple model, occupancy levels within the clinical department was determined more influential than occupancy for the individual destination unit (Table 1). Further, elevated care level (intensive or intermediate care) and morning admission decisions were associated with decreased waiting. Using regression for survival data to quantify relationships like these has value independent of the DES in understanding operations.

Despite the simplicity of the regression model illustrated, the approach to execution within a DES is flexible to accommodate more complex models in an identical manner. However, the regression covariates are limited to measures available in the DES. Thus, factors that influence wait must be explicitly part of the DES for the method to be useful. This has potential to introduce additional complexity that DES developers should consider in the context of the overarching modeling objectives (i.e., marginal utility).

It is also important to note that using survival analyses within DES was previously demonstrated by the authors (Levin et al. 2008; Levin et al. 2011). These publications were designed for presentation to a clinical audience (cardiology) with limited description of the DES and no explanation of the logic for regression model integration. The previous DES models were for different health systems with different objectives. However, in constructing the larger-scale hospital DES described in this paper, similar challenges in accounting for transition process time arose. This provided evidence of a problem more general to DES modeling in healthcare. This motivated the production of this paper as a contribution to an audience well versed in DES methodology.

6 CONCLUSION

Wait time is often a key performance measure targeted in DES models in healthcare. Waiting in these environments is caused not only by occupied downstream resources, but also variability in human and organizational processes associated with transition. To address, we describe a novel method to integrate regression models for survival data in DES capable of quantifying the drivers of transition processes to estimate wait time. These methods are illustrated by example within a large hospital DES.

REFERENCES

American College of Emergency Physicians (ACEP). 2011. “Definition of a Boarded Patient: Policy Statement” Accessed March 24, 2015. http://www.acep.org/Content.aspx?id=75791

Armony, M., I. Shlomo, A. Mandelbaum, Y. Marmor, Y. Tseytlin, G. Yom-Tov. 2010. “On Patient Flow in Hospitals: A Data-Based Queuing-Science Perspective”. Working paper, New York University, New York.

Cook, A. 2008. “Chapter 6 Parametric Survival Models”. Introduction to Survival Analysis Course Notes. Accessed March 24, 2015. https://courses.nus.edu.sg/course/stacar/internet/st3242/handouts/ch6l2.pdf

Cox, D. 1972. “Regression Models and Life-Tables”. *Journal of the Royal Statistical Society*. 34(2):187-220.

Grambsch, P., T. Therneau. 1994. “Proportional Hazard Tests In Diagnostics Based On Weighted Residuals” *Biometika*. 81:515-526.

General Electric Healthcare. 2011. “Hospital of the Future Snapshots of Success” Accessed March 24, 2015. http://partners.gehealthcare.com/Hospital_Of_the_Future_Snapshots-CS-0112.pdf

Kang, H., H. Nembhard, C. Rafferty, C. DeFlitch. 2014. "Patient Flow In The Emergency Department: A Classification And Analysis Of Admission Process Policies" *Annals of Emergency Medicine*. 64(4):335-342.

Hamrock, E., K. Paige, J. Parks, J. Scheulen, S. Levin. 2012. "Discrete Event Simulation for Healthcare Organizations: A Tool for Decision Making," *Journal of Healthcare Management*. 58(2):110-124.

Hosmer, D., S. Lemeshow, S. May. 2008. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. 2nd ed. Hoboken, NJ: John Wiley and Sons, Inc.

Institute of Medicine. 2006. "Hospital-Based Emergency Care: At the Breaking Point" Washington, DC: National Academies Press.

Levin, S., R. Dittus, D. Aronsky, M. Weinger, J. Han, J. Boord, D. France. 2008. "Optimizing cardiology capacity to reduce emergency department boarding: A systems engineering approach" *American Heart Journal*. 156(6):1202-1209.

Levin, S., R. Dittus, D. Aronsky, M. Weinger, D. France. 2011. "Evaluating the Effects of Increasing Surgical Volume on Emergency Department Patient Access" *BMJ Quality and Safety*. 20:146-152.

Jacobsen, S., S. Hall, J. Swisher. 2006. "Discrete-Event Simulation of Health Care Systems" In *Patient Flow: Reducing Delay in Healthcare Delivery*, edited by R. Hall, 211-252. United States: Springer.

Joint Commission. 2012. "Patient Flow Through the Emergency Eepartment" R3 Report. Accessed March 24, 2015. http://www.jointcommission.org/assets/1/18/R3_Report_Issue_4.pdf

Schemper, M. 1992. "Cox Analysis of Survival Data with Non-Proportional Hazard Functions" *The Statistician*. 41:455-465.

Shi, P., M. Chou, J. Dai, D. Ding, J. Sim. 2015. "Models and Insights for Hospital Inpatient Operations: Time Dependent ED Boarding Time" *Management Science* 24:13-14.

Simon, R., E. Canacari. 2012. "A Practical Guide to Applying Lean Tools and Management Principles to Health Care Improvement Projects" *Association of Perioperative Registered Nurses Journal*. 95(1):85-100.

AUTHOR BIOGRAPHIES

SCOTT LEVIN is an Associate Professor in the Department of Emergency Medicine and Civil Engineering at the Johns Hopkins University (JHU), Baltimore. He is the Director of the JHU Systems Institute http://systems.jhu.edu/ and member of the Department of Operations Integration within the Johns Hopkins Health System. He holds a Ph.D. in biomedical engineering from Vanderbilt University, Nashville, TN. His research focuses on the application and development of systems engineering tools to improve the value of healthcare and medical decision-making. His e-mail address is slevin33@jhmi.edu.

MAXIM GARIFULLIN is a Simulation Architect in GE Healthcare Partners. He holds an M. Sc. in computer science from St. Petersburg Polytechnic University. Maxim focuses on applications and development of simulation decision support tools that help healthcare providers plan, design new, and redesign existing facilities, that are more efficient in the delivering healthcare. His email address is Maxim.Garifullin@ge.com.