Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

THE CASE AGAINST UTILIZATION: DECEPTIVE PERFORMANCE MEASURES IN IN-PATIENT CARE CAPACITY MODELS

K Louis Luangkesorn
Theologos Bountourelis
Andrew Schaefer

Department of Industrial Engineering
University of Pittsburgh
1048 Benedum Hall
3700 O'Hara St.
Pittsburgh, PA 15261, USA

Spencer Nabors
Gilles Clermont

Department of Veterans Affairs Medical Center
University Drive C
Pittsburgh, PA 15240, USA

ABSTRACT

Health care capacity decisions are often based on average performance metrics such as utilization. However, such decisions can be misleading, as a large portion of the costs in service operations is due to the inability to provide service due to congestion. This paper will review sources of variation that affect in-patient care capacity and develop a series of models of patient flow in a health care facility. We demonstrate that even in settings where the patient population and services provided are fixed, models that do not account for natural variations in the arrival rate and correlation in patient lengths of stay in sequential units will show the same utilization, but underestimate congestion and the resulting costs. Therefore, we argue that utilization is an inappropriate measure for validating models and congestion metrics such as blocking and diversions should be used instead.

1 INTRODUCTION

As healthcare costs continue to dominate fiscal discussions on local and national stages, the optimization of hospital operational costs has become a pivotal subject in the pursuit of cost management and reductions. Capacity decisions in hospitals are usually based on a target occupancy level given historical demand (Green 2002, Tonor and Waldhorn 2010). As a result of efforts to reduce apparent excess capacity, the resulting capacity may lead to risk of an institution's not being able to provide appropriate care to patients.

Efforts to optimize healthcare resources should seek to balance the competing interests of providing service and operating costs. Other domains where capacity decisions are a necessary component of strategic planning have found that that decisions based on average demand and processing time alone may lead to lower upfront costs, but often lead to operational bottlenecks and shortage penalties because of variability in demand arrivals and workloads required by each arrival (Brennan and Schwartz 1985, Lee, Padmanabhan, and Whang 1997). In a healthcare system, this will manifest as periods of congestion, where patients meeting criteria for lower levels of care cannot be transferred due to lack of space, and diversions, where patients are sent to other health care facilities due to lack of space at the original facility. Using historical patient flow data from a tertiary care facility, we demonstrate the effects of different sources of variability on capacity utilization and congestion through simulations of a series of bed capacity models.

2 LITERATURE REVIEW

Most of the literature on capacity modeling in health care facilities does not consider variability. In this review, we look at some examples of modeling in health care facilities. Then we review literature on health care capacity modeling that specifically discusses sources of variability and their impact.

2.1 Capacity Modeling for Analysis and Managing In-patient Care/Critical Care Facilities

Modeling efforts in health care face challenges beyond those documented in other application domains. In a review of over 200 papers on simulation applications to healthcare problems Wilson (1981) found that only 16 reported that recommendations were acted upon. Brailsford et al. (2009) examined 342 articles from mainstream academic journal publications. They rated each based on the scale of implementation. 171 articles (50%) were suggested (theoretically proposed by authors), 153 (44.7%) were conceptualized (discussed with the clients), and only 18 (5.3%) were implemented (actually used in practice). Katsaliaki and Mustafee (2010) looked at 201 high-quality journal papers published from 1970 to 2007 on healthcare related simulation research. They found only 11 out of 201 (5.4%) reported on the implementation of results to stakeholder organizations. Van Lent et al. (2012) performed a similar review using papers on simulation in health care from 1997 to 2008. Out of 89 papers, 21 reported that the hospital accepted the results while 16 reported implementation or partial implementation of the recommendations. They identify a number of factors that tend to lead towards acceptance and implementation including data availability, validation and verification with historic data, and the use of generic models that present general principles.

Modeling efforts in health care face challenges beyond those documented in other application domains. Carter and Blake (2005) describe a sampling of projects they have participated in and discuss some special difficulties such as issues with data collection, dealing with patient confidentiality, unique characteristics of each facility, need for greater modeling detail, difficulty in tracking clinicians and points of decision making, and unplanned critical events. These factors can lead to difficulty in bringing modeling projects to completion, being relevant for decision makers, and developing generalizable insights. Some of the methods typically used by modelers such as limiting the scope of the model and abstracting some details can lead to the model missing details that lead to incorrect recommendations.

2.1.1 Modeling in Health Care Settings

The bulk of contributions in health care facility modeling can be partitioned into two main categories: (1) estimating and optimizing patient flow; and (2) reducing operating costs without compromising the level of patient care.

The first category is driven by the need of many healthcare providers to streamline patient flow in a way that minimizes delays and maximizes patient satisfaction. Hence, a significant part of the literature adopts certain patient flow characteristics as an objective and as a starting point for their analysis. In this context, they examine necessary organizational and resource changes as well as additional interventions needed to optimize patient admission rates and delay times.

Lowery (1992) presented a simulation model of the surgical suite and critical care areas of a large hospital. The primary objective was to assess the impact of critical care bed configurations on performance measures such as bed utilization, the number of patients denied admission, bumped, or accommodated on alternative units. Harper and Shahani (2002b) demonstrated the importance of modeling the various types of patient flows when simulating bed occupancies and patient rejection rates. They showed that the explicit modeling of patient mix results in higher-fidelity models that are able to capture the bed occupancy fluctuations over time. Cochran and Bharti (2006) presented the methodology underpinning the compilation and validation of a large-scale simulation model that includes the ICU and telemetry units. The validation was done using hospital data for an existing facility. The model was subsequently used to solve a stochastic bed-balancing problem that leads to balanced bed utilization rates that minimize the blocking of beds across different units.

The second broad category is motivated by the pressing need to simultaneously reduce hospital operating costs without compromising the quality of provided health care. While some of the existing literature includes simulation models to inform resource allocation decisions in budget-constrained environments, models that consider the effect of bed and staffing allocations have received the most attention. The latter category of healthcare simulation models is especially relevant to the modeling, analysis, and management of ICUs.

Williams (1983) developed a simulation model to select the number of beds needed to meet the hospital standards of care. The model was calibrated with patient data collected over a period of 12 months. Williams used this model to examine the effect of the number of ICU beds and admissions, transfers, premature discharges, empty beds and operating costs. Vassilacopoulos (1985) used a simulation model to determine the number of beds required by inpatient units to satisfy several measures of operating efficiency such as high occupancy rates, immediate admission of emergency patients, and short patient waiting lists. Ridge et al. (1998) investigated the relationship between admission rules and rejection rates in a single ICU unit where the patients are classified as either emergency or non-emergency and rules incorporated for diverting patients when bed occupancy thresholds are reached. They highlighted the nonlinear relationship between the number of beds, the occupancy levels, and transfer rates.

2.1.2 Examination of Effects of Variability

Lowery (1993) constructed and validated a critical care simulation model by examining the occupancy rates and the number of refused admissions to the unit due to lack of available beds. Harper and Shahani (2002a) studied the complex relationships between bed allocation, bed occupancy rates, and the denied admission rate. Costa et al. (2003) illustrated the danger in using only average values to determine the number of critical care beds in the face of nonlinearity and system variability. By considering alternative bed configurations, it was shown that using only average values, the bed requirements can be underestimated.

Rauner et al. (2003) use generalized linear models (GLM) to model length of stay (LOS) in order to study the impact of various factors on hospital reimbursement rates. The GLM included disease code or surgical procedure, day and month of admission, admission type, and discharge type. This was used to test hypothesis regarding the significance of the day of week of admission, month of admission, and the type of admission and discharge. This work resulted in recommendations regarding hospital capacity planning and operating policies as well as improvements in the reimbursement policies to align hospital incentives with overall healthcare goal.

Shahani, Ridley, and Nielsen (2008) present a simulation model for a single critical care unit (CCU). Emphasis was given to the modeling of the patient mix using the CART method to obtain LOS distributions for statistically different patient categories. (Litvak et al. 2008) use both simulation as well as the Equivalent Random Method (ERM) to generate the same expectation and variance of overflow in a single multi-server unit. They allow for three types of ICU patients arriving according to separate Poisson processes and use these models to investigate regional ICU capacity in the Netherlands.

Queueing models have also been used to test assumptions about arrival distributions. Kim et al. (1999) look at an ICU unit where patients come from the general wards, emergency room, elective surgery and emergency surgery. They use an *M/M/c* queueing model for the ICU and use a simulation model to test various assumptions. In particular, using a uniform arrival process instead of Poisson leads to underestimating the number of waiting patients by 2/3. They further come to the conclusion that capacity issues are not due to the overall capacity of the unit, but to the timing of bed supply and demand. McManus et al. (2004) compare queueing models against simple averages. Using these models show exponential increase in rejection rates when utilization exceeds 80% compared to using simple averages. They comment that the stochastic nature of patient flow may lead health planners to underestimate resource needs in busy intensive care units. Griffiths et al. (2006) use a $M/H/c/\infty$ queuing model, representing high variability in LOS through a hyper-exponential distribution. They confirm that capacity planning in the intensive care unit (ICU) cannot simply be based on averages as this may generate an

underestimation of resource needs during busy periods. Kolker (2009) incorporated day-to-day variability in a model developed to look at the effects of managing the elective surgery schedule on ICU operations. Bekker and de Bruin (2010) look at the impact of time-stationarity on the number of required beds and fraction of refused admissions. They develop approximations of a $M_t/H/c/c$ queuing model. They find that this causes a considerable effect on the reported fraction of refused admissions as the weekday-weekend pattern of arrivals is taken into account.

2.2 Summary

Reports examining regional hospital bed capacity tend to use average performance based standards for planning purposes. Traditionally, researchers have used occupancy rates or utilization as this data is readily available in healthcare systems, but this does not account for variation or the fact that not all beds provide the same capability (Bazzoli et al. 2006).

While there are reports in the literature that discuss the dangers of using average based metrics, stationary arrival rates, and assuming homogeneous patient populations when modeling length of stay, many documented models that are used in practice make those assumptions. Reports examining regional hospital bed capacity tend to use average based standards for planning purposes, with expectations of a certain percentage of beds over the average (Tonor and Waldhorn 2010), but in reality there is no objective standard for measuring bed capacity constraints.

Our earlier work (Bountourelis et al. 2011) demonstrated the need to look at medically indicated length of stay (MLOS) as opposed to actual LOS as well as the need to look at in-patient care throughout the hospital instead of ICU in isolation, which are lacking in the overwhelming majority of patient flow models in the literature. This paper will develop an set of models of patient flow that examines the effects of non-stationary arrival rates and heterogeneous population LOS in a facility that accounts for multiple levels of care on both average based and service based performance measures. This will demonstrate that utilization does not reflect congestion and the resulting costs of blocking and diversions.

3. MODELING CASE STUDY

In this modeling exercise, we will develop a range of models with increasing fidelity. We will start with a mean value analysis that ignores variation, than progressively increase the fidelity of the model by incorporating non-stationarity of demand and correlated length of stays in the various units.

The hospital model is of a facility with two patient sources and four levels of care (Figure 1). Patients can arrive from *internal* sources, such as clinics or surgery. *External* sources would include the emergency department, other care institutions, or direct admissions from home.

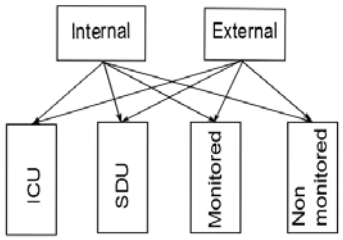

Figure 1: Model diagram of patient sources and units

A patient admission is defined as a sequence of stays in units corresponding to four levels of care: intensive care, step down unit (a lower level of care than intensive care), monitored (telemetry), and non-

monitored (general medical or surgical wards without telemetry). Patients can enter at any point, and as they progress in their treatment, can in principle move from any unit to any other unit until they reach a conclusion of their treatment. In addition, the patient can have as part of the sequence of stays a period in an outside location.

3.1 Data

For arrival and patient length of stay data we use patient arrival, patient movement, and transfer request data collected over a 18 months from July 2010 through December 2011 from a local tertiary care facility (Bountourelis et al. 2011). For each admission the data includes the source of the patient, day and time of initial admission, the sequence of units the patient was in, and the medically required length of stay in each unit.

Length of stay data is based on a combination of data on patient movement as well as requests to transfer patients. Following Bountourelis et al. (2011) we use the time between the request to transfer in and request to transfer out of a unit as the MLOS (Figure 2). Using the MLOS for each patient that is based on physician determination of the suitability of a patient to move provides length of stay input data that is independent of the state of the system. This then allows us to use the period between the time a request to transfer is received to the time that a patient actually moves to the next unit as a performance measure of blocking.

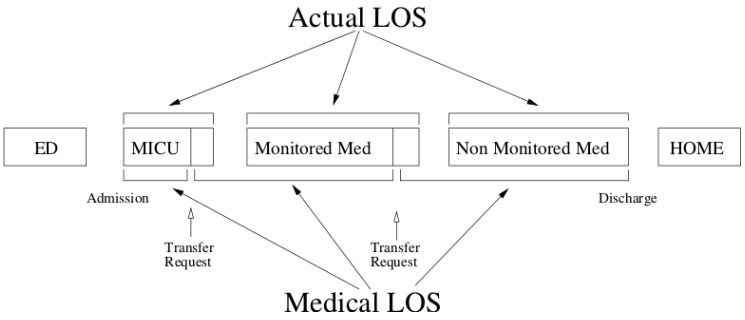

Figure 2: Patient length of stay sequence using medially indicated length of stay (MLOS)

3.2 Model Development

We develop a notional hospital based on mean value analysis of arrivals and MLOS requirements. For each ward in the hospital, the average arrivals per week and average medical length of stay was calculated (see Table 1). The average MLOS times the number of stays generated in a week for each unit yields the average bed hours of care generated in a week, or the weekly demand of bed hours. From this, by dividing by the hours per week we determine the average beds required and also the number of beds that would be equivalent to 20% reserve capacity.

Table 1: Level of care statistics

	ICU	SDU	Monitored	Non-monitored
Mean MLOS	108.3	116.7	106.8	105.6
Standard Deviation MLOS	130.8	159.7	126.9	119.5
Bed hours/week	4035	1257	4637	10440
Mean required beds	24.0	7.5	27.6	62.1
Model configuration (20% extra capacity)	29	9	34	75

3.3 Simulation Model Description

To determine the arrival rate, the sum of the hourly arrival rates over the week was taken. Then the average interarrival time was calculated and the resulting value used as the mean of the exponential distribution for interarrival times (Table 2).

Table 2: Weekly internal and external arrivals

	Internal	External
Arrivals per week	32.6	92.8
Interarrival time (hours)	5.15	1.81
Fraction ICU	0.24	0.32
Fraction SDU	0.23	0.05
Fraction Monitored	0.30	0.49
Fraction Non-monitored	0.51	0.98

For all simulation models, when an admission arrival is generated, the simulation samples from the historical data of patients with the given patient source. Each patient sampled is defined by the historical sequence of stays in the hospital along with historical medical length of stay.

Blocking occur if a patient MLOS has been reached within the simulation so is ready to be transferred, but there is no bed available in the destination unit. When this occurs, the patient remains in place occupying the current bed, and the time is credited to the MLOS in the destination unit. When a bed is opened, the patient will transfer into the destination unit. The time in between the time the transfer was authorized and the actual transfer is charged as blocking.

Diversions occur when at the time of admission or if a patient needs to move to a higher level of care, there is no bed available. If not, check at higher levels of care. If no bed is available then the patient is declared to be diverted and leaves the system.

We implement four models using the SimPy simulation library (Vignaux, Muller, and Helmbold 2012) that differ according to the representation of the arrivals and length of stay as described in Table 3: (i) a base model with stationary (S) arrivals and independent (I) MLOS (SI), (ii) stationary arrivals and empirical (E) MLOS (SE), (iii) non-stationary (N) arrival rates and independent MLOS (NI), and (iv) non-stationary arrival rates and empirical MLOS (NE). The implementation details are described in the following sections.

Table 3: Characteristics of models for comparison

Model	**SI**	**SE**	**NI**	**NE**
Demand	Stationary	Stationary	Non-stationary	Non-stationary
LOS type	Independent	Empirical	Independent	Empirical

3.3.1 Stationary vs. Non-Stationary Arrivals

We employ two approaches for modeling patient arrivals corresponding to (i) a stationary, and (ii) non-stationary Poisson distribution. Non-stationary arrivals correspond to a more accurate representation of the arrival stream, however it can be more complex to represent in models. The question to be addressed is if ignoring non-stationarity effects the findings of the model.

For stationary arrivals, the patients are generated hourly using a Poisson distribution with a fixed rate resulting from determining the average hourly arrival rate from each source over a week.

To model non-stationary arrivals, we calculated the average number of arrivals for each hour of the week over the 18 month data collection period. The external and internal arrivals are shown in Figure 3. The calculated rates reflect the variability of patient arrivals during the day and the fact that external patients tend to also arrive on weekdays as opposed to weekends.

In both cases, arrivals from each source are generated for each hour of the simulation. After the number of arrivals are generated, exact time of the arrival is distributed within the hour using a Uniform(0, 1) distribution. Patient arrivals are generated every hour. For each hour, the number of arrivals are generating using a Poisson distribution, then the time of the arrivals is randomly chosen within the hour using a uniform distribution. This procedure is followed for both internal and external patients. This method generates patients where the rate of arrivals changes over time. In particular, it reflects the fact that patients tend to arrive on weekdays as opposed to weekends, and patients that come from other areas of the hospital tend to arrive during standard working hours, while those from outside the hospital arrive seven days a week.

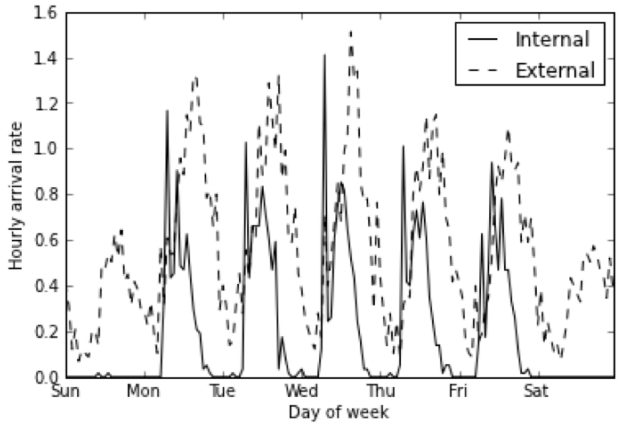

Figure 3: Arrival rate by source and hour of week

3.3.2 Independent vs. Empirical LOS Sequences

The MLOS was similarly modeled in two ways. First was to use an independent fitted distribution, the second was to use the empirical LOS sequence of a single patient from admission to discharge.

To fit a distribution to the MLOS for each unit, we used a lognormal distribution. We noted in our literature review that this was commonly used for modeling patient LOS. We evaluated this assumption using QQ-plots and the Chi-squared test for each of the units. For example, with the ICU there were 1453 data points. We see that the resulting QQ-plot (Figure 4) follows the 45^{o} line closely until the last 9 points (the tail 0.6%). Further evaluation using the Chi-squared test with 11 bins, sizing the bins to ensure that all bins had a reasonable number of points, we obtained a p-value of 0.32, which does not reject a hypothesis that the ICU MLOS is correctly modeled with a lognormal distribution.

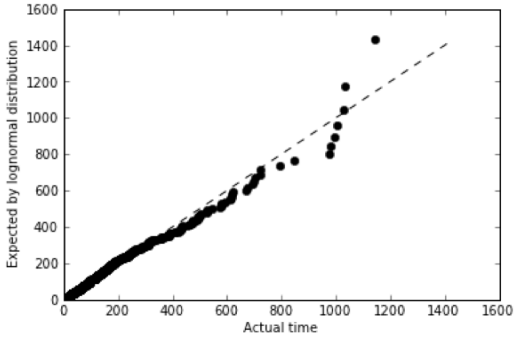

Figure 4: QQ Plot of Actual vs. Expected ICU LOS using Lognormal distribution

The second method is to use the empirical length of stay sequence. One of the chief benefits of using an empirical sequence of LOS is that this accounts for patterns of LOS across units for a given patient. This accounts for correlation as well as the fact that different treatment protocols will result in different LOS patterns.

For every admission the simulation will use the historically experienced, or empirical, MLOS for each unit for a randomly selected patient from the same source. This allows for the simulation to incorporate any correlation that may exist between the MLOS in different units. Model SE will be used to examine the impact of considering empirical LOS to the base SI model, and model NE will be used to investigate the joint effects of non-stationary arrivals and empirical LOS.

4. SIMULATION STUDY

The performance measure in this study is the total of blocking costs (Table 4) and diversion costs (Table 5). Blocking is calculated by determining the bed day cost for the medically indicated unit and the unit the patient is forced to stay in due to lack of space, and charging the prorated difference. Diversion costs are calculated based on a per incident basis, differentiated between primary care and critical care patients. We use values that are proportional to those found in our partner facility. (Bountourelis et al. 2011).

Table 4: Daily bed costs.

ICU	SDU	Monitored	Non-monitored
$1200	$650	$500	$400

Table 5: Diversion cost per incident in model.

Critical Care	$21,000
Primary Care	$6,600

The simulation was pre-loaded with patients followed by a warmup period. The model was preloaded with a number of patients equal to the mean number of patients as calculated in the mean value analysis. For each patient, a random stay in the sequence of length of stay was drawn as the current step in the sequence. For that stay, a random fraction of the historical MLOS for that patient was drawn as the remaining time for the MLOS for that patient. After an eight week warmup period, the simulation was run for one year.

To determine the required number of replications, 25 replications were made of each of the four models. The total number of replications, N, was then determined so that for each of the four models the expected 95% confidence interval for total annual cost was within 10% of the average total annual cost as determined by the initial replications. Through this procedure we determined $N = 232$.

4.1 Unit Utilization and Blocking

While the standard metric for hospital bed use is occupancy (which corresponds to utilization), the statistic that directly relates to diversion and blocking is the probability a unit is full. The utilization and percent of time the unit was full for the four models is given in Table 6.

In Table 6 we see that for all four models, the observed utilization is the same across models. Note that if utilization was used as the performance measure for model validation, all models would have been validated. When looking at the probability the unit was full, this is different. This does show a difference across the models, in particular with the ICU. The stationary/independent model (SI) shows a lower period where the ICU is full than the other models. In particular, the models where arrivals are non-stationary (NI and NE) show the period the ICU is full to be a third larger than for model SI.

Table 6: Unit performance measures

Model	**SI**	**SE**	**NI**	**NE**
Demand	**Stationary**	**Stationary**	**Non-stationary**	**Non-stationary**
LOS type	**Independent**	**Empirical**	**Independent**	**Empirical**
Utilization (percent)				
ICU	63	63	63	63
SDU	77	76	77	77
Monitored	80	81	80	80
Non-monitored	86	83	86	86
Probability unit full (percent)				
ICU	0.63	0.67	0.84	0.80
SDU	33.7	30.1	33.6	29.9
Monitored	18.8	19.7	20.7	21.1
Non-monitored	17.0	11.7	20.1	14.1

4.2 Costs Due to Diversion and Blocking

The difference in the percent of time the unit is full translates into costs from diversion or blocking when arrivals observe full units. Using notional costs with this model leads to results as shown in Table **7**.

Table 7: Annual Diversion and Blocking

Model	**SI**	**SE**	**NI**	**NE**
Demand	**Stationary**	**Stationary**	**Non-stationary**	**Non-stationary**
LOS type	**Independent**	**Empirical**	**Independent**	**Empirical**
Annual diversions				
Critical Diversion	9.3	10.0	14.2	13.2
Primary Diversion	0.98	1.0	2.8	2.2
Annual blocking costs				
Blocking Cost	42,700	113,000	60,500	114,000
Diversion Cost	202,000	216,000	315,000	291,000
Total Cost	**244,000**	**330,000**	**375,000**	**405,000**

Models SE, NI, and NE report significantly greater costs than model SI. While adding the effects of non-stationary arrivals has a greater effect than using empirical length of stay sequences, combining both yields higher reported costs. Compared to highest fidelity model (NE), the standard model which assumes stationary arrivals and independently derived length of stay (SI) underreports costs by 40%.

5. DISCUSSION

While models of health care capacity are typically based on average occupancy, the important metrics are measures of congestion, blocking and diversions. This has analogs in other fields such as using the stock-out percentage in supply chains or loss functions in queuing service applications such as those applied to call centers. In supply chains and call centers, it is well understood that the appropriate performance metric of the system is the ability of the system to provide a service when required and this performance is then balanced against the costs to provide that capacity.

Using measures of congestion, the fact that in-hospital care arrivals and length of stay are not stationary or independent across units makes it necessary to use models that can explicitly represent these aspects of the system. The effect of non-stationarity is particularly pronounced.

6. CONCLUSION

As modelers, we have a desire to use models that are as simple as possible. Banks and Chwif (2010) admonish, "Keep the model simple, but not too simple. Make the model complex, but not too complex." While most health care capacity analysis is based on average performance based statistics such as utilization or occupancy, these statistics do not provide indication about the facility ability to provide services to patients. As observed in other areas where the performance measure of interest is service, the performance measures of interest should be based on the ability of the system to provide that service, in this case patient blocking and diversions. This work benefited greatly from our partner facility beginning to collect transfer request data in electronic form shortly before the beginning of this project. Previously, like many other facilities, this information was not retained after the transfer was completed.

For system performance, it is important to identify significant sources of variation and include them in the model. In particular, the variation in arrivals over the course of a week and the fact that patient length of stays are not independent can lead to significant changes in the observed congestion.

We are currently applying the concepts presented here in a more detailed model of our partner facility with a goal of informing capacity decisions. Through these more detailed models, we will also further examine the effects of non-stationary and correlated LOS on congestion.

ACKNOWLEDGEMENTS

This work was supported by the Department of Veterans Affairs (VA), Veterans Health Administration, VA Pittsburgh Healthcare System (VAPHS) Veterans Engineering Resource Center (VERC). The authors would like to thank the staff of the VAPHS and the VERC for their assistance and feedback including Robert Monte and Matt Jenkins. This work has also benefited from the comments of anonymous reviewers. The contents of this document do not represent the views of the Department of Veterans Affairs or the United States Government.

REFERENCES

Banks, Jerry, and L Chwif. 2010. "Warnings About Simulation." *Journal of Simulation* 5 (November): 279–291. doi:10.1057/jos.2010.24.

Bazzoli, Gloria J., Linda R. Brewster, Jessica H. May, and Sylvia Kuo. 2006. "The Transition from Excess Capacity to Strained Capacity in U.S. Hospitals." *The Milbank Quarterly* 84 (2): pp. 273–304.

Bekker, R., and A.M. de Bruin. 2010. "Time-dependent Analysis for Refused Admissions in Clinical Wards." *Annals of Operations Research* 178 (1): 45–65. doi:10.1007/s10479-009-0570-z.

Bountourelis, Theologos, K. Louis Luangkesorn, Spencer Nabors, Gilles Clermont, Andrew J. Schaefer, and Lisa Maillart. 2011. "Development and Validation of a Large Scale ICU Simulation Model with Blocking." In *Proceedings of the 2011 Winter Simulation Conference*. Phoenix, Arizona, United States.

Brailsford, S.C., P.R. Harper, B. Patel, and M. Pitt. 2009. "An Analysis of the Academic Literature on Simulation and Modelling in Health Care." *Journal of Simulation* 3: 130–140.

Brennan, Michael J., and Eduardo S. Schwartz. 1985. "Evaluating Natural Resource Investments." *The Journal of Business* 58 (2) (April 1): 135–157.

Carter, M., and J. Blake. 2005. "Using Simulation in an Acute-care Hospital: Easier Said Than Done." *Operations Research & Management Science* 70, Part 2: 191–215.

Cochran, J.K., and A. Bharti. 2006. "Stochastic Bed Balancing of an Obstetrics Hospital." *Health Care Manage Sci* 9: 31–45.

Costa, A. X., S. A. Ridley, A. K. Shahani, P. R. Harper, V. De Senna, and M. S. Nielsen. 2003. "Mathematical Modelling and Simulation for Planning Critical Care Capacity." *Anaesthesia* 58: 320–327.

Green, Linda V. 2002. "How Many Hospital Beds?" *Inquiry* 39 (4): 400–412.

Griffiths, J. D., N. Price-Lloyd, M. Smithies, and J. Williams. 2006. "A Queueing Model of Activities in an Intensive Care Unit." *IMA Journal of Management Mathematics* 17: 277–288.

Harper, P. R., and A. K. Shahani. 2002a. "Modelling for the Planning and Management of Bed Capacities in Hospitals." *The Journal of the Operational Research Society* 53, No. 1: 11–18.

Harper, P.R., and A.K. Shahani. 2002b. "Modelling for the Planning and Management of Bed Capacities in Hospitals." *Journal of the Operational Research Society* 53: 11–18.

Katsaliaki, K, and N Mustafee. 2010. "Applications of Simulation Within the Healthcare Context." *Journal of the Operational Research Society* 62 (8) (October 13): 1431–1451. doi:10.1057/jors.2010.20.

Kim, S. C., I. Horowitz, K. K. Young, and T. A. Buckley. 1999. "Analysis of Capacity Management of the Intensive Care Unit in a Hospital." *European Journal of Operational Research* 115: 36–46.

Kolker, Alexander. 2009. "Process Modeling of ICU Patient Flow: Effect of Daily Load Leveling of Elective Surgeries on ICU Diversion." *Journal of Medical Systems* 33 (1): 27–40. doi:10.1007/s10916-008-9161-9.

Lee, Hau L., V. Padmanabhan, and Seungjin Whang. 1997. "Information Distortion in a Supply Chain: The Bullwhip Effect." *Management Science* 43 (4) (April 1): 546–558.

van Lent, Wineke AM, Peter VanBerkel, and Wim H van Harten. 2012. "A Review on the Relation Between Simulation and Improvement in Hospitals." *BMC Medical Informatics and Decision Making* 12 (1) (March 14): 18. doi:10.1186/1472-6947-12-18.

Litvak, N., M. van Rijsbergen, R.J. Boucherie, and M. van Houdenhoven. 2008. "Managing the Overflow of Intensive Care Patients." *European Journal of Operational Research* 185: 998–1010.

Lowery, J. C. 1993. "Multi-hospital Validation of Critical Care Simulation Model." *Proceedings of the 1993 Winter Simulation Conference*. WSC '93: 1207–1215.

Lowery, J.C. 1992. "Simulation of a Hospital Surgical Suite and Critical Care Area." In *Proceedings of the 1992 Winter Simulation Conference*, 1071–78. WSC '92.

McManus, Michael L., Michael C. Long, Abbot Cooper, and Eugene Litvak. 2004. "Queuing Theory Accurately Models the Need for Critical Care Resources." *Anesthesiology* 100 (5) (May): 1271–1276.

Rauner, Marion Sabine, Achim Zeiles, Michaela-Maria Schaffhauser-Linzatti, and Kurt Hornik. 2003. "Modelling the Effects of the Austrian Inpatient Reimbursement System on Length-of-stay Distributions." *OR Spectrum* 25 (2): 183–206. doi:10.1007/s00291-003-0120-z.

Ridge, J.C., S.K. Jones, M.S. Nielsen, and A.K. Shahani. 1998. "Capacity Planning for Intensive Care Units." *European Journal of Operational Research* 105: 346–355.

Shahani, A.K., S.A. Ridley, and M.S. Nielsen. 2008. "Modelling Patient Flows as an Aid to Decision Making for Critical Care Capacities and Organisation." *Anesthesia* 63 (10): 1074 – 80.

Tonor, Eric, and Richard E. Waldhorn. 2010. *The Next Challenge in Healthcare Preparedness - Catastrophic Health Events*. Center for Biosecurity of UPMC. http://www.upmc-biosecurity.org/website/resources/publications/2010/2010-01-29-prepreport.html.

Vassilacopoulos, G. 1985. "A Simulation Model for Bed Allocation to Hospital Inpatient Departments." *Simulation* 45, No. 5: 233–241.

Vignaux, Tony, Klaus Muller, and Bob Helmbold. 2012. "SimPy Manual V2.3." http://simpy.sourceforge.net/.

Williams, S.V. 1983. "How Many Critical Care Beds Are Enough?" *Critical Care Medicine* 11 (6): 412–416.

Wilson, J. C. Tunnicliffe. 1981. "Implementation of Computer Simulation Projects in Health Care." *The Journal of the Operational Research Society* 32 (9): 825–832. doi:10.2307/2581399.

AUTHOR BIOGRAPHIES

LOUIS LUANGKESORN is a Research Assistant Professor at the Department of Industrial Engineering at the University of Pittsburgh. He received a B.S. in General Engineering and a B.A. in Political Science from the University of Illinois-Urbana, an M.A. in Science, Technology and Public Policy from The George Washington University, and a Ph.D. in Industrial Engineering and the Management Sciences from Northwestern University. His research interests include logistics and resource management in health care and emergency response settings, as well as the use of simulation and modeling for policy analysis and evaluation. His email address is lol11@pitt.edu.

SPENCER G. NABORS is an attending physician at the Department of Critical Care Medicine at the University of Pittsburgh Medical Center and the Department of Veterans Affairs (VA) Pittsburgh Healthcare System. He received his BA in Bio-Chemistry and Philosophy from New York University, College of Arts and Sciences, his MA in Bio-Clinical ethics from New York University, Graduate School of Arts and Sciences, his MPH in Health Policy & Management from Columbia University, Mailman School of Public Health and his MD from SUNY Downstate, College of Medicine. His research interests include Process Modeling, Simulation and it's applications in various clinical practices, Management Sciences and Critical Care Ultrasound technologies. His work is supported by a training grant from the National Institutes of Health (HL07820). His email is naborssg@upmc.edu.

THEOLOGOS BOUNTOURELIS is a Research Assistant Professor at the Department of Industrial Engineering at the University of Pittsburgh. He received a B.S. in Mathematics from the Aristotle University of Thessaloniki, Greece, and a PhD in Operations Research from the Georgia Institute of Technology. His research interests, are in the area of Markov Decision processes, Machine Learning theory, Simulation and its applications in various technological contexts including HealthCare applications. His email address is thb28@pitt.edu.

ANDREW SCHAEFER is a Professor of Industrial Engineering and Wellington C. Carl Fellow at the University of Pittsburgh. He has courtesy appointments in Bioengineering, Medicine, and Clinical and Translational Science. He received his PhD in Industrial and Systems Engineering from Georgia Tech in 2000. His research interests include the application of stochastic optimization methods to health care problems, as well as stochastic optimization techniques, in particular stochastic integer programming. He is interested in patient-oriented decision making in contexts such as end-stage liver disease, HIV/AIDS, sepsis, and diabetes. He is also interested in health care systems, including operating rooms and intensive care units. He is an Associate Editor for INFORMS Journal on Computing and IIE Transactions. His email is schaefer@pitt.edu.

GILLES CLERMONT is an Associate Professor of Critical Care Medicine at the University of Pittsburgh. He earned his undergraduate and medical degree from McGill University in Montréal, Quebec, Canada. From there he went to the University of Montréal, where he earned his Master of Science in Physics. After six years of private practice, Dr. Clermont returned to the University of Montréal to serve as Chief Resident at the Hôpital Notre-Dame. He then became a Research Fellow in Critical Care Medicine at the University of Pittsburgh, School of Medicine. Dr. Clermont is on the editorial boards for several peer-reviewed journals, such as Drug Discovery Today, Journal of Statistical Physics, and Critical Care, amongst others. He is also the Vice-President of the Society for Complexity in Acute Illness (SCAI). His current research interests are Complexity in Critical Illness, Epidemiology of Critical Illness, and Cost Effectiveness Analysis. His email address is clermontg@upmc.edu.