A BETTER APPROACH TO MODELING EMERGENCY CARE SERVICE

Sankar Sengupta
Meredith Deneweth
Robert Van Til

Oakland University
2200 North Squirrel Road
Rochester, MI 48309 USA

ABSTRACT

The objective of this paper is to develop and analyze models used in emergency care system. A different modeling approach is presented based on the idea of using a priority measure defined as "operating priority" that changes with time as a patient waits for service and each resource pays due consideration to skill hierarchy and selects the next patient for treatment. The new modeling approach breaks away from the conventional approach in which entity drives the request for resources. The resource is the driver in the new approach. This paper presents a comparison of performance between the conventional and the new approach in modeling emergency care. The paper also discusses potential enhancements of the proposed approach.

1   INTRODUCTION

Discrete event simulation models offer a very powerful analytic tool to understand system behavior of a health care system. Discrete event simulation qualifies as the appropriate tool since it meets all the major requirements of such a study. These requirements are as follows.

- Study of a complex system
- Presence of one or more source of randomness in the model
- Needs to study both time dependent as well as steady state behavior of the system

Simulation becomes an important tool for a more complex system with many elements, a large amount of interdependencies among the elements, and/or considerable variability (Babin and Greenwood 2011). In addition, the key role played by human beings in healthcare systems requires an approach which allows interaction between modeler and the user. Simulation meets this requirement (Brailsford 2007).

The role of different system modeling tools in study of health care systems is described in depth in Brailsford (2007). Healthcare models may be classified into three groups. The first group of models is frequently called "disease models" but they can also represent biological processes in healthy individuals. These models are often used for studying the clinical effectiveness or cost-effectiveness of some intervention. The second group of models represents operational or tactical models at the healthcare unit level. These models are usually concerned with modeling the flows of patient in a healthcare system and identifying and eliminating bottlenecks (Samaha, Armel, and Starks 2003). These models are used for capacity planning, resource allocation and process redesign. For example, estimating how many nurses and doctors are required to ensure that no patient spends more than four hours in the emergency room (ER) etc. (Harper and Shahani 2008). Professional labor is the biggest cost component of healthcare in a hospital, yet to answer any issue in the service delivery model is to add more people (Stansfield, Massey, and Manuel 2011). Finally, the third group of models may be described as strategic models or system-wide models

which often do not model individual patients at all. Unlike the first two groups where discrete-event model is often used, group three models always use System Dynamics models and answer more long-term and broad-based issues. For example issues such as, how might Electronic Health Record impact upon the health of the next generation of children or how many doctors will the US need in 2030 etc. are addressed by group three models.

The model presented in this paper conforms to a group two model. In this model of emergency care a hospital is treated as a factory in which a patient travels through the system as an entity. The patient demands different types of resources during its journey. The entity is disposed as it leaves the system. A set of performance measures, either cost or time based, are typically collected from a simulation model. This factory view of a hospital has some limitations while modeling emergency care. In the proposed approach, the resource and not the patient is in the drivers' seat.

2   LIMITATIONS OF THE CONVENTIONAL APPROACH

In the conventional factory view approach, entities join a queue with a priority and wait for resources. Resource is allocated based on priority. Due date, processing time, penalty for being late etc. are some of the elements in the list of priorities. In due date or processing time based priorities entities can remain in the system for a long time, a situation not permitted for patient needing emergency care. A tardy job may or may not have serious implications. We need to be aware of the following limitations.

- A patient with low priority may have a long waiting time before receiving treatment. This is not acceptable since most often severity of the patients' needs increases with time and patient must receive treatment within an acceptable time period. Hospitals are using promise to provide emergency care, within a specified time since arrival, as a differentiator in advertisement campaigns.
- The medical staff that include nurses, junior doctors (residents), senior doctors (attending), specialists, etc., has a skill hierarchy. For example senior personnel, while capable of doing any clinical task will not do so unless the other resources are busy, in order to remain available for higher skilled tasks. Junior personnel, on the other hand, are restricted to a subset of the list of possible clinical tasks. In conventional modeling approach, skill hierarchy is ignored and resources are treated as homogeneous leading to inefficient utilization of resources.

In order to avoid the aforesaid limitations, Hay, Valentin, and Bijlsma (2006) proposes a new modeling approach with the following unique features.

- use of a priority measure defined as "operating priority" that changes with time as a patient waits for service
- each resource pays due consideration to skill hierarchy and selects the next patient for treatment

Both features together are expected to have favorable impact on system performance by lowering the waiting time and leading to better resource utilization.

The next section of the paper describes the salient features of the proposed approach.

2.1   The Salient Features of the New Approach

2.1.1 Priority Measure

The proposed approach uses a priority measure that is based on the severity of the need for service of a patient as the patient arrives. The priority for a patient, unlike in a manufacturing system where priority for a new job may not change with time, must change with time to reflect the reality. Based on clinical condition of the patient a priority is assigned on arrival. For example, a patient with acute respiratory problem will receive higher priority compared to a patient arriving with a minor ailment. The priority of a patient increases with time as a patient waits in the queue. The system status is checked after a fixed time

interval and patients' priority is updated. The plan has two unique components, which include updating interval and increment in priority based on the clinical condition of the patient.

In this paper both parameters have been selected in a subjective manner and require further analysis. The updating interval is expected to be small and increment in priority is expected to be dependent on the initial clinical condition of the patient. Table 1 and Table 2 show the increment plan used in this paper.

Table 1: Attending physician thresholds (from Hay, Valentin and Bijlsma (2006))

Skill	Threshold
Treat_Resuscitation	1
Treat_RapidAssessment	1
Treat_ConsultwithResident	1
Treat_StandardAssessment	10
Treat_continuitngCare	10
Treat SutureLaceration	15

Table 2: Resident physician thresholds (from Hay, Valentin and Bijlsma (2006))

Skill	Threshold
Treat_Resuscitation	1
Treat_StandardAssessment	1
Treat_continuitngCare	1
Treat SutureLaceration	1

2.1.2 Skill Hierarchy

A patient may be viewed as a job that requires one or more tasks to be performed on it concurrently. Each task represents a well defined medical procedure. In this case a resource unit is likely to have multiple skills thereby being capable of performing multiple tasks. For example, an attending physician is capable of performing a larger number of tasks compared to a resident physician. Each patient stakes a claim for a set of tasks to be performed either concurrently or in a specified order. The objective of this approach is to insure better resource utilization, making sure that a higher skilled resource is not utilized when a lower skilled resource is idle. This is achieved through assigning a threshold value of priority to each task performed by a type of resource. A low level task will have a larger threshold value for a higher priced resource compared to that of a lower priced resource. For example, an attending physician will have higher threshold value for a "treat suture laceration" task compared to that for a resident physician. The value of current priority of a task required by a patient must equal or exceed the threshold value before the resource owning that skill can be requested. Table 1 and Table 2 provided by Hay, Valentin and Bijlsma (2006) offer a better understanding to the readers of skill hierarchy and the concept of threshold value. The threshold skills and level amounts differentiate one resource over another. The authors did not use data from these tables and included them as an example for structure.

3   MODEL DESCRIPTIONS

Arena simulation software (Kelton, Sadowski, and Sturrock 2007) model is used for this study. Patients arrive to the emergency department based on a random inter-arrival pattern. On arrival each patient is assigned a severity measure from a discrete random distribution, based on patient's medical condition and urgency of care required as shown in Table 3. The patient then receives a numerical score as priority measure which is dependent on its severity measure. The patient waits in a queue and the priority measure is updated after a fixed time interval until the patient receives service. The increase in priority measure is also dependent on patient's severity. An updating interval of 5 minutes is used in this model.

Table 3: Patient mix and priority updating plan

Patient-Mix	Clinical Priority	Updating Interval	Increment in Priority/Minute
70% severity=1	20 for severity= 1	3 minutes	5 for severity=1
20% severity=2	10 for severity =2	3 minutes	2 for severity=2
10% severity=3	5 for severity =3	3 minutes	1 for severity=3

The nurses, resident physicians and attending physicians constitute the three types of resources used in this model. As described earlier in this paper, updating the priority measure periodically and paying attention to resource hierarchy, are the distinguishing features of this model. Each resource searches the patient queue in a way so that a resource will not access a patient unless the patient's priority equals or exceeds the threshold for the resource and no resource in the lower levels of the hierarchy is idle. Table 4 shows the threshold values in the simulation with the nurse (Resource Type 1) having the lowest threshold value, being able to service patients with lower needs. The attending physician (Resource Type 3) requires a much larger threshold and services only the more severe patients.

Table 4: Threshold value for different resources

Resource Type	Threshold Value
1	5
2	10
3	15

Features of Arena such as, "wait and signal" and "copy" have been used to model the search of patient queue by a resource. Table 5 lists all the inputs used in the Arena model. Table 6 lists the outputs received from the model.

Table 5: List of inputs

1.	Inter-arrival time distribution
2.	Levels of severity of a patient
3.	Severity and clinical priority relationship
4.	Priority updating interval
5.	Priority updating scheme
6.	Categories of resources used
7.	Threshold of priority for each category of resource
8.	Number of units of each resource
9.	Patient service time

Table 6: List of outputs

1.	Average waiting time in system
2.	Average time in system
3.	Utilization of resources

The simulation model is analyzed as a non-terminating system. The data from a long simulation run is used to identify the transient period. The method of batch-means is used to estimate the performance measures. The following measures of performance are of interest in this study.
- average waiting time
- average time spent in the system

- utilization of resources

The basic model used three types of resources and included one unit in each of the three categories such as nurse, resident and attending physician.

Two different scenarios have been compared in the first study. In the first scenario, we have used skill hierarchy for the resources and a threshold value for each resource. The second scenario treated the three categories of resources as homogeneous and modeled them as three units of identical resource. In the first scenario as shown in Figure 1, termed as the proposed approach, the resource is the driver as each unit of resource searched the patient queue and selected a patient, when successful, for service.
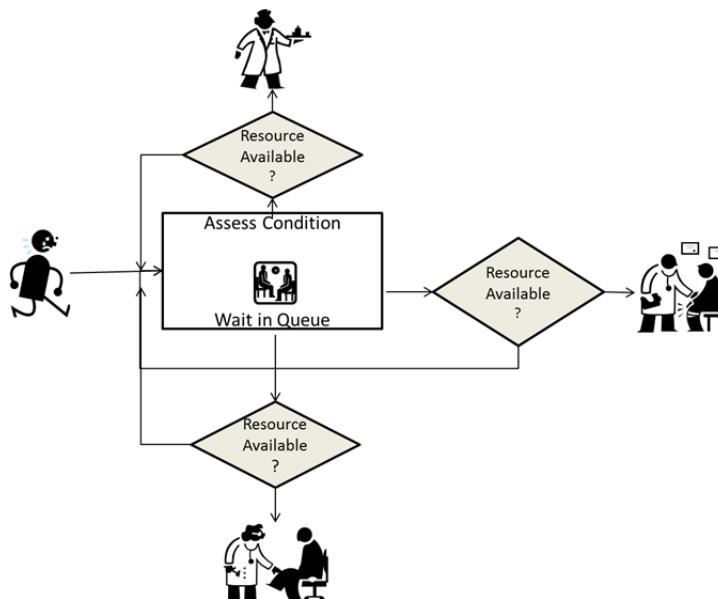

Figure 1: Proposed approach

In the second scenario as shown in Figure 2, termed as the Previous approach, the patient is the driver and demanded access of resource and seized an idle unit. It is expected that compared to the second scenario, the first scenario will lead to better resource utilization at the expense of higher time in system. This expectation is confirmed by simulation results.
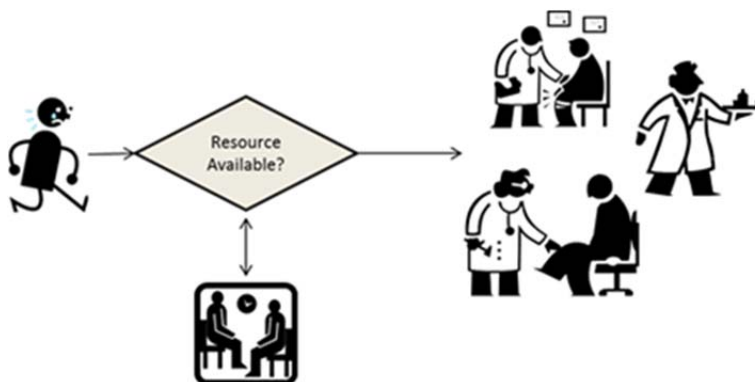

Figure 2: Previous approach

The second type of analysis on this model was to look at the allocation of multiple resources and the average time a patient spends in the queue. The first study used one of each resource (attending, resident and nurse). Simulation trials with various combinations of resources were run using the model to measuring the number of each resource, utilization and total time the patient waited in the waiting room (queue) before receiving appropriately assessed medical attention. For example, the first trial involved a total of 5 resources; 3 nurses, 1 resident and 1 attending. The second trial involved a total of 5 resources as well, but with 2 nurses, 2 residents and 1 attending. Each combination trial ran non-terminating for 90 days or 129,600 minutes. When a patient was assigned to a resource, the resource became unavailable to another patient for a time given by a triangular distribution of 15, 30 and 45 minutes. During this time, the next patient in the same severity level must wait in the queue or be assigned to another available resource meeting his level of criteria. Fourteen combinations of the three types of resources were simulated to include 2-4 nurses, 1-4 residents and 1-4 attending physicians. The wait time was recorded and confirmed to have under a 5% half width to be of verified statistical significance. A lower average wait time with additional total resources is expected to be confirmed by the simulation results.

4   RESULTS

Table 5 displays the utilization of each type of resource comparing scenario I and scenario II. The simulation shows the lower level resources (nurse and resident) are more utilized in scenario I than in scenario II. The average time in queue is longer in scenario I (the proposed approach), than in scenario II (the previous approach).

Table 5: First analysis simulation results

Resource \ Utilization	Scenario I (Proposed Approach)	Scenario II (Previous Approach)
Type 1 (Nurse)	74.2%	64%
Type 2 (Resident)	64.4%	64%
Type 3 (Attending)	46.8%	64%
Average Time in Queue	36.4 minutes	34.6 minutes

Figure 3 exhibits the results of several simulation runs with multiple resource combinations. The number of each resource per trial is shown by differentiating color in a bar graph. The line graph displays the corresponding average time in the waiting room from the given trial. The trials started with a total of 5 resources and went up to a total of 12 resources.

It must be emphasized that parameters used in this study are based on estimates by persons familiar with the problem. The authors did not estimate any parameter based on analysis of historical data. The authors plan to use real world data to drive this model as part of a future project with a local hospital.

4.1   Analysis of Results

The results show the proposed approach leads to a better distribution of resource utilization. A higher skilled resource such as an attending physician is available to perform other more skilled medical tasks that are not covered by other resources. The study also shows a statistically significant difference in average time in queue between the two scenarios. The first scenario is expected to lead to larger time in queue because of threshold based allocation of resource and adherence to the rules of hierarchical resource structure.

The results displayed in Figure 3 show the resource allocation trials where the more resources added, the lower the wait time for the patient. The total number of resources did not strictly line up with the corresponding time in queue For example, the first set of trials had 3 nurses, 1 resident and 1 attending. The second trial had the same total of 5 resources but had a much less average waiting time allocating 2 nurses, 2 residents and 1 attending. Therefore the allocation of type of resource was shown to make a difference rather than total resources. The trial with 3 nurses, 1 resident and 1 attending demonstrated having

more lower paid staff did not pay off. Higher level staff is needed to take care of more critical patients in a timely manner.
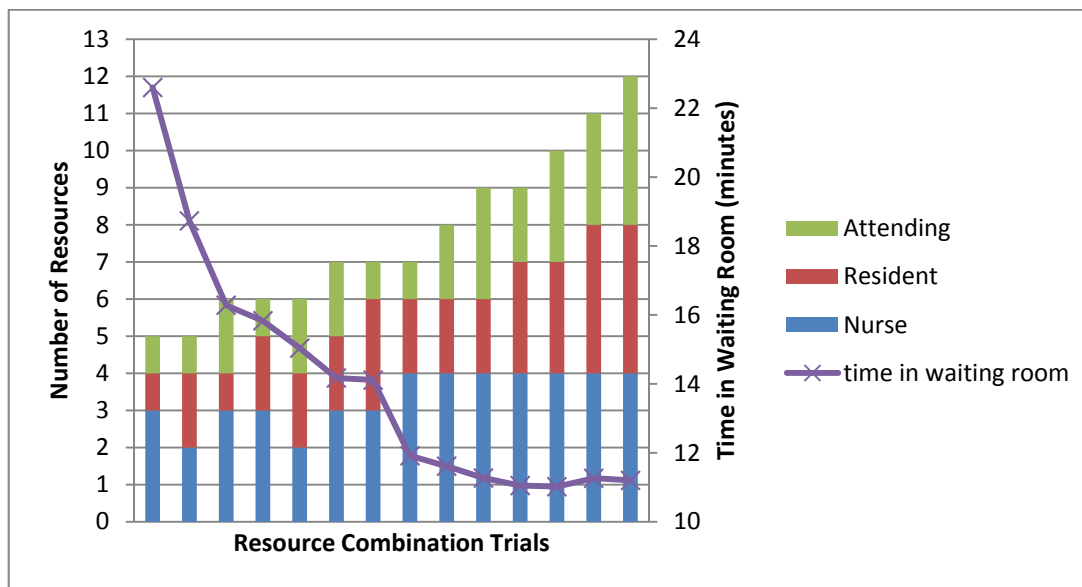

Figure 3: Second analysis, numbers of resources by trial and time in waiting room

Figure 3 also shows the smallest average wait time did not come from the trial with the most resources. The lowest average wait time of 11.02 minutes was found with the total number of resources of 10 allocated by 4 nurses, 3 residents and 3 attending physicians. When additional residents and or attending were added to the model, the average waiting time of the patient did not decrease. The second lowest average wait time of 11.05 minutes (.03 minutes or 1.8 seconds less than first place) had a total number of 9 resources with 4 nurses, 3 residents and 2 attending physicians.

4.2   Statistical Analysis

Each measure was estimated with 5% or better relative precision. The difference in average time in queue between the two operating procedures is statistically significant at 5% level of significance.

5   DISCUSSION

The proposed modeling approach offers an operating procedure that leads to better utilization of resources. This approach guarantees that a higher priced resource is not put into service until all lower priced resources are utilized. This makes good business sense as well as making sure that persons with specialized skills are available to handle more difficult tasks. The major inputs are as follows.

- different categories of severity that should be included in the model
- priority scale to be used upon arrival of a patient
- priority updating scheme
- threshold value for each resource

Each input is subjective in nature and demands that the modeler has the required medical knowledge.

This study has taken a simplistic view of patient's needs whereby all the needs of a patient can be taken care of by any one of the resources. In reality a patient may be viewed as a job that requires service from a number of different resources either concurrently or sequentially where some of the resources pro-

vide unique services. The authors plan to expand this model as a variant of a job shop to derive more insights into system behavior.

The additional resources lowered the time a patient spent in the waiting room (queue). However, does the cost (measured by utilization) of an additional physician or nurse reflect in the time saved in the waiting room for a patient? For example, the difference between the first and second shortest average wait time was .03 minutes and one attending physician. Is the cost of one physician worth .03 minutes of an average patient's life?

A balance of all three levels of resources represented in this paper are needed to create a lower wait time for a patient entering the emergency room. Using 3 low level nurses and only 1 resident and 1 attending led to a longer average time for a patient to wait to seek medical attention. The analysis also showed that larger amounts of higher skilled workers did not significantly benefit the patient's wait time. Figure 3 shows that when the number of the resident resource went from 3-4, the average total wait time for the patient did not improve.

The proposed simulation model offers a decision making tool to arrive at operating conditions that meets specific goals. For example, low waiting time prior to receiving service at an ER unit is used as an advertisement tool by different hospitals. A local hospital claims to guarantee an upper bound of 29 minutes for waiting time in the ER Unit. The website for many hospitals have live current waiting room times posted for customers to view.

In the future, the proposed model can be combined with an optimization procedure to arrive at optimal values of a set of control variables. The future enhancements to this model are as follows.

- Modeling each patient's needs as a set of medical tasks or procedures and treating the problem as a job shop
- Solving the problem as a multi-criteria optimization problem

6   LIMITATIONS AND CONCLUSION

The goal of this paper was to assign resources to appropriate level patients based on initial condition at arrival. If the level of patient was not present, the medical professional was free to do other work. The paper was not based on empirical study. It only shows the structure for the approach. The medical condition of the patients was used to allocate resources and not to diagnose or repair condition. Processing time of the patients did not depend on severity of medical condition of each patient.

Bed assignments are often a constraint in problems involving a medical facility. It was not considered in this study.

The proposed modeling approach offers an operating procedure that leads to better utilization of resources.

REFERENCES

Babin, P., and A. Greenwood, A. 2011. "Discretely Evaluating Complex Systems." *Industrial Engineer Magazine* 43(2):34-38.
Brailsford, C. S. 2007. "Tutorial: Advances and Challenges in Healthcare Simulation Modeling." In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 1436-1448. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Harper, P. R., and A. K. Shahani. 2008. "Modeling for the Planning and Management of Bed Capacities in Hospitals." *Journal of the Operational Research Society* 53:11-18.
Hay, M. A., E. C. Valentin, and R. A. Bijlsma. 2006. "Model Emergency Care in Hospitals: A Paradox-The Patient Should Not Drive the Process." In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. R. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 439-445. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kelton, W. D, R. P. Sadowski, and D. T. Sturrock. 2007. *Simulation with Arena.* 3rd ed. New York: McGraw-Hill, Inc.

Samaha, S., W. S. Armel, and D. W. Starks. 2003. " The Use of Simulation to Reduce the Length of Stay in an Emergency Department." In *Proceedings of 2003 Winter Simulation Conference*, edited by S. E. Chick, P. J. Sanchez, D. M. Ferrin, and D. J. Morrice, 1907-1911. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Stansfield, T., R. Massey, and J. Manuel. 2011. "Life Support for Hospital Staff." *Industrial Engineer Magazine* 43(2):28-33.

AUTHOR BIOGRAPHIES

SANKAR SENGUPTA is an Associate Professor in the department of Industrial and Systems Engineering at Oakland University. His research interests include: modeling and performance analysis of manufacturing and service systems including healthcare system, supply chain modeling and risk analysis, reliability modeling and novel applications of simulation. He has years of industrial experience and is actively involved with local industry. He has published in reputed journals and has made technical presentation both in national and international conferences. He can be reached at sengupta@oakland.edu.

MEREDITH DENEWETH is a Ph.D. student at Oakland University in Systems Engineering. She received a B.S. in Industrial Engineering from Purdue University and a M.S. in Vehicle Design and Manufacturing from Central Michigan University. She has thirteen years of automotive engineering experience specializing in the supply chain of vehicle model launches. Her research interests are in supply chain simulation and modeling. Her email is medenewe@oakland.edu.

ROBERT P. VAN TIL is a Professor in the Department of Industrial and Systems Engineering at Oakland University, Rochester, MI. His current research interests include modeling and analysis of manufacturing system, development of simulation based predictive tools and application of lean principles to manufacturing and service systems. His e-mail address is vantil@oakland.edu.