SIMULATION BASED OPTIMIZATION: APPLICATIONS IN HEALTHCARE

Tarun Mohan Lal
Thomas Roh
Todd Huschka
Mayo Clinic
200 First St. SW
Rochester, MN 55905, USA

ABSTRACT

Increasing healthcare costs are driving the need for optimizing care delivery processes. Due to the complexity associated with healthcare processes, discrete event simulation is the most popularly used decision support tool in assessing trade-offs between multiple objectives of healthcare systems. However in situations where there is little or no structure to input constraints, it can be very difficult to evaluate all alternative configurations. Simulation based optimization is a technique used to efficiently find solutions to problems that have a large number of possible scenarios. In this method a simulation model is used to develop an approximate mathematical model that represents the surface of the results over a range of input values. This is then solved using linear programming or integer programming or other advanced optimization heuristics. In this paper, we discuss the methodology and applications of simulation based optimization, highlighting advantages, challenges and opportunities of using this method in healthcare.

1 INTRODUCTION

The growing costs of healthcare is a major concern for healthcare providers. As healthcare organizations move towards the goals of reducing costs, optimizing patient experience, and improving health of populations; operations research tool are becoming more important. These tools provide the ability to assess trade-offs between resource utilization, quality of service, and operating costs (Mohan Lal and Roh 2013). Operations research is a systems engineering methodology that applies advanced mathematical engineering analytics to enhance decision making within the context of solving complex problems like those in healthcare. When applied appropriately, it has the potential to estimate the consequences of alternatives and evaluate choices to see which choice would have the most beneficial impact before actual implementation. Discrete event simulation (DES) is one of the most commonly used operations research tool in healthcare. Its unique ability to account for high levels of complexity and variability that exist in the real world, along with animation capability make it easier to illustrate and gain buy-in from physicians and other clinical providers compared to other black-box mathematical models offered by operations research. However, DES also has some limitations. In scenarios where there are a large number of stochastic input decision variables and there is little information about the structure of output function using simulation modeling alone can be tedious and complicated. In such cases, optimization via simulation can help to maximize or minimize measures of the performance by evaluating the system using discrete event simulation (Banks 2004). This technique popularly known as simulation based optimization is fairly new and could be very valuable in analysis of healthcare systems.

Unfortunately, we have observed road blocks to the implementation of Operations Research solutions in real world situations. In particular the internal politics of individual hospital units can make
implementation very difficult when multiple units are involved. Ultimately, Operations Research techniques which would be considered novel have been rejected in favor of techniques which do not appear to be black-box type solutions. In our experience DES modeling has been relatively successful in bridging this problem by providing solutions which can be both easy to understand and pass face validity to many parties typically involved in these problems. Once face validity has been achieved and the resources (affected staff) are confident that the model accurately reflects their real-world situation it is possible to build an optimization framework over the simulation model in order to achieve to optimize the system. In this tutorial we briefly discuss the concept of simulation based optimization, describe the potential areas of application in healthcare and detailed examples of problems where simulation based optimization was successfully applied, to provide the basic education needed to encourage its usage in healthcare.

2 LITERATURE REVIEW

Discrete event simulation is one of the most widely used operations research method in healthcare and its usage has increased over the years indicated by literature survey articles by Jun (1999), Thorwarth (2009), Gunal (2010) each citing more than 100 articles that use discrete event simulation modeling in healthcare. Roberts (2011) recognizes that in healthcare it is often difficult to define a single performance characteristic. Especially in healthcare further investigation is often needed to understand how a change in the process leads to downstream impact. Hence simulation is considered an ideal technique to be used.

Literature also indicates that most researchers find value in using discrete event simulation to validate process improvement or re-engineering efforts as well as to provide support for operational decision making in relatively confined environments such as outpatient clinics, call centers, pharmacy, ICU, and Emergency Departments. Most simulation modeling literature describes the process of conducting extensive scenario analysis to find an optimal balance between staff, capital and facilities/equipment (Mustafee 2010, Caberara 2011). However, there are several potential disadvantages of resorting to iterative analysis that is required by simulation. Law and McComas (2000) pointed out that one of the problems with simulation is that historically it wasn’t considered an optimization technique. The inability of simulation to provide a single optimal solution, unlike other analytical modeling approaches, is not always appealing to customers or decision makers. The trial and error method is time consuming and decision makers particularly in healthcare; do often not appreciate tedious iterative review of output data. (Lowery 1996). In addition, Wilson (1981), Lane (2003), Brailsford (2009) point out the conundrum of implementing the simulation results in practice and highlight that model turnaround time and accuracy play a significant role in gaining customer buy–in, which suffers in the iterative approach.

In order to take advantage of the benefits of computer simulation modeling while trying to avoid some of the tedious nature of choosing the best solution or policy using a set of candidate parameter settings(inputs) we explore the nature of simulation based optimization in healthcare problems. Simulation based optimization is an emerging filed that integrates optimization into simulation analysis. Although this technique has been applied in other industries (Law 2000, Jung 2004, Schwartz 2006), simulation still not very popular in healthcare. Recent examples of application of this method in healthcare include study of Sundaramoorthi (2010), where this technique was used to plan nurse resource allocation to patients based on workload needs. Ahmed (2009) used simulation based optimization to design a decision support tool to determine the optimal number of doctors and other staff to maximize the number of patients seen. Most recently, Zhang (2012) applied this integrated approach to determine the staffing requirements of a long term care facility.

With the continually increasing ability of computers to run complex simulations in a reasonable time and the rise in optimization algorithms, along with the fact that most discrete event simulation vendor packages are now offering this capability built into their simulation tools; the ability to perform simulation optimization is becoming a more a viable method of analysis.
3 SIMULATION BASED OPTIMIZATION

3.1 Definition
Optimization is an operations research technique that seeks to maximize or minimize the performance measures by manipulating the input decision variables under certain restrictions defined by the constraints. Significant research has been done to solve or mathematically program optimization problems. However, one major limitation of optimization is that the factors/decision variables and outputs/responses are assumed to be known with certainty or discrete in nature. In the case of simulation modeling, the goals are similar when we are trying to evaluate alternative configurations and selecting the best system. The assumption is that the possible options are specified based on constraints or obligations. Situations when there is little or no structure to knowing the possibilities and the goal is to identify which of the possible many combinations of the input factors leads to the optimal performance, simple statistical analysis is not feasible or helpful. Even using experimental design only helps to identify the important factors out of multiple factors in the model and how they affect the outputs but cannot be used to determine the optimal combination of factor levels to maximize or minimize the response.

Simulation based optimization, also often known as simulation optimization, refers to the process of maximizing or minimizing the expected or long run average of key output performance measures from simulation modeling with respect to input variables/factors of the model as constraints. (Banks et. al. 2005). This technique is most valuable in situations where the analyst is trying to identify the set of input factors that lead to the optimal outputs. For example if \( a_1, a_2, a_3, \ldots a_m \) are the controllable input variables (decision variables) and \( Y(a_1, a_2, a_3, \ldots a_m) \) is the output random variable from the simulation model, and the goal is to maximize or minimize this output, we can define the optimization model as Max or Min \( \text{E}(Y(a_1, a_2, a_3, \ldots a_m)) \) subject to constraints defined by all combinations of \( a_1, a_2, a_3, \ldots a_m \) where \( m \) is large.

3.2 Simulation based Optimization Methods

3.2.1 Procedure
The process of conducting a simulation based optimization study is very similar to the frameworks used to build discrete event simulation models includes the following stages:
1. Define the problem and associated goals and metrics
2. Select the input variables
3. Define the constraints on decision variables
4. Determine the output or key performance measures
5. Collect the appropriate data from existing data sources or through time studies
6. Develop the discrete event simulation model
7. Validate and verify the model.
8. Optimization
9. Review the results

Once the simulation model is developed and validated (steps 1 through 7), the next step is to build the optimization model that can be solved using the optimization packages that are available within most simulation software’s today. Details of how these algorithms work and practical nature of their operation are explained in the section below.

3.2.2 Optimal System Seeking Methods
As discussed previously, after the simulation model is developed it can be represented as an approximate mathematical model/equation (which is often very difficult to do) that represents the surface of the results

1263
Mohan Lal, Roh, and Huschka

over the range of input values. This equation is a regression model, where independent variables are the simulation input parameters and the dependent variable is the response of interest. Several such equations need to be developed when there is more than one output of interest. Then several simulations are carried out to determine parameters of the regression equation and the regression model can be refined further over a larger range of input variables. After which, the equation can be solved to find the optimal solution. Several approaches to optimize simulations based on response surfaces, meta models and neural networks are prominent in literature. (Laguna 2002, April 2003, and van Beers 2003) The methods to optimize a simulation mostly are trying search the space for possible input-factor combinations and require varying levels of information. This is an active area of research and the approaches to solve the simulation optimization algorithms are growing. Fu (2002) categorizes these methods into 4 main approaches namely stochastic approximation which is gradient based, sequential or response surface methodology, random search and sample path optimization. However, it’s important to note that the solution is as accurate as the equation and the time invested in developing the right equation depends on the level of accuracy needed. Derivatives of the input parameters provide the vector that estimates the change in the output values with small changes to the inputs.

Most of simulation software vendors like Arena, Simul8, MedModel, Any Logic have integrated optimization packages that develop the sequence of equations to represent system configurations, each configuration representing a set of inputs/scenarios so that the most optimal system design can be obtained to meet the objective. These solutions are near optimal and do not guarantee the most optimal solution. Also, the optimization packages require specification of a number of options, parameters and tolerances, which influences the results. (Law and Kelton 2000). Depending on the complexity of the problem at hand it might make sense for the user to analytically solve the problem or utilize the software capabilities.

4 DETAILED CASE STUDY

4.1 Problem Description

In 2012, the emergency department (ED) at St. Mary’s Hospital in the Mayo Clinic was planning to renovate the current space and change the versatility of the new rooms. Prior to the ongoing renovations, patients were assigned to certain room types based on age and acuity level. The new rooms were planned to be “universal” in that they can accommodate patients of all types. The only exception were trauma patients that would still have their own rooms because of the amount of equipment and level of urgency associated with their care. Under the new conditions and with yearly increasing demand, ED leadership needed to plan in advance any new physician hires for 2015 (these decisions are made well in advance because of the length of the hiring process and future budget planning deadlines.)

Due to the need for forecasting future events and a model that captures the complexity of an emergency department system, discrete event simulation was chosen as the modeling technique. Underlying the goal of determining additional hires, two questions needed to be answered.

- What is impact on load when moving from parallel queues to a shared queue and
- what is the optimal schedule for physicians over a weekly schedule?

4.2 Data Analysis

Existing data from 2011 was analyzed to determine the features of the simulation. The data included patient age, Emergency Severity Index, arrival time, abandonment time, bed entry time, and bed exit time. Patients arrival rates were modeled as Poisson based on time of day and day of week. The bed length of stay distributions were modeled separately between acuity and whether the patient was pediatric. Because the bed length of stay combines several processes and delays, a combination of distributions were combined to produce a close fitting theoretical distribution.
The Emergency Severity Index (ESI) is an acuity level system. ESI first subdivides patients into “urgent” those that need immediate medical care and “non-urgent” those that can wait. Level 1-2 refers to “urgent” patients where the ESI 1 patients are trauma and ESI 2 are non-trauma. “Non-urgent” patients are separated by the estimated number of resources that a patient may require. ESI 3 patients require 2 or more resources, ESI 4 patients require only one resource, and ESI 5 patients require no resources. Only patients with ESI 2-5 were of interest because ESI 1 patients were assumed to utilize special resources that are not part of the scope of this project.

The emergency department was studied and a patient flow model was developed. Since consultant schedules were the ones of interest, nursing and other auxiliary staff were not considered. Patients arrive into the system and are placed into queues based on acuity and pediatric status. Those with the lowest acuity are given highest priority to be placed into a bed. Upon entering patients are also assigned a probability of leaving based on last patient’s waiting time with matching ESI. Patients that decided to leave immediately balk the queues. Since consultant workflow and process times were not available in the data. If patients are less than 18, they may enter the Pediatric only pod if it is open. If not, pediatrics patients were divided into the acuity queues that they shared with adult patients. Consultants were tied to a maximum of 8 beds that they could service at one time. Patients would then seize one of those beds in one of the consultant pods and hold that resource based on the fitted bed length of stay distributions. Figure 1 shows how a patients pathways were constructed in the simulation model.

Figure 1: Simulation patient flow map.
4.3 Simulation Model

A discrete event simulation was built in Arena 14.0 (Kelton et. al 2002), and validated that included the aforementioned features. The simulation bed capacity was then changed to infinite. Pediatric consultant and pod schedules were not being considered for change. Therefore, those capacity constraints were included in the modeling. By allowing the patients to seize an infinite amount of beds at any time, we are able to measure the offered load of the patients on the system (Marmor 2009). The results from the simulation are shown in Figure 1 and show the median number of patients that are in the system during each hour. Tuesday, Wednesday, Thursday, Friday, and Sunday were grouped together after exhibiting very similar patterns.

4.4 Mixed Integer Program

With the simulated output of load and also knowledge of the constraints of scheduling systems, it is possible to create a mixed integer program to solve for optimal physician scheduling. Since load was measured in beds, one major assumption is a ratio of staff-to-beds. Therefore, the simulated load was divided by 8 to be able to pair load and capacity. The constraints placed in the model to create realistic solutions were as follows:

- Physicians will only work 8 hour schedules
- At least one physician must be on duty at any time
Shifts only can start at the beginning of each hour
- The ratio of physicians-to-beds is 8:1

\[ s_h = \text{shifts covering hour } h \]
\[ x_h^+ = \text{positive error at hour } h \]
\[ x_h^- = \text{negative error at hour } h \]
\[ H = \text{hours in a week} \]
\[ \gamma_h = \text{patient load at hour } h \]
\[ c = \text{patient load per consultant} \]
\[ \min \sum_{h=1}^{H} x_h^+ + x_h^- \]

\[ s.t. \]
\[ cs_h + x_h^+ - x_h^- = \gamma_h \]
\[ x_h^+, x_h^- \geq 0 \]
\[ \forall h \in H \]
\[ s_h \in \mathbb{Z}^+ \]

### 4.5 Results

A total of 168 8-hour schedules are possible. The objective function seeks to minimize the absolute error between the offered load and the scheduled capacity for the physician led teams. The optimization program was built in Excel and Open-Solver was used to solve since a large number of variables were involved. Figure 3 shows the results for the optimal shift scheduling. In Figure 4, each block shows an individual 8 hour physician shift. The leftmost column indicates the pod that the physician will be working in.
Figure 3: Compares the physician capacity over the course of a week to the patient load on the system. Load is converted from bed usage by the physicians-to-beds ratio.

Figure 4: Sample breakdown of Monday capacity showing consultant-to-pod schedules.

The solution shown above is optimized to a 50% service level in that the offered load was computed as the median of the simulation output. In other words, patients will immediately be placed into a bed 50% of the time. To create schedules for higher or lower service levels, different quantiles of the load can be computed from the simulation output. By adding another constraint, the total number of individual schedules in a week can be set to a desired minimum, exact, or maximum number of schedules.
Table 1: Shows a sample schedule of the exact time and number of physician shifts.

<table>
<thead>
<tr>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>End</td>
<td>Start</td>
<td>End</td>
<td>Start</td>
<td>End</td>
<td>Start</td>
</tr>
<tr>
<td>6:00</td>
<td>14:00</td>
<td>6:00</td>
<td>14:00</td>
<td>6:00</td>
<td>14:00</td>
<td>0:00</td>
</tr>
<tr>
<td>9:00</td>
<td>17:00</td>
<td>9:00</td>
<td>17:00</td>
<td>9:00</td>
<td>17:00</td>
<td>8:00</td>
</tr>
<tr>
<td>12:00</td>
<td>20:00</td>
<td>12:00</td>
<td>20:00</td>
<td>12:00</td>
<td>20:00</td>
<td>9:00</td>
</tr>
<tr>
<td>14:00</td>
<td>22:00</td>
<td>14:00</td>
<td>22:00</td>
<td>14:00</td>
<td>22:00</td>
<td>12:00</td>
</tr>
<tr>
<td>17:00</td>
<td>1:00</td>
<td>17:00</td>
<td>1:00</td>
<td>17:00</td>
<td>1:00</td>
<td>16:00</td>
</tr>
<tr>
<td>20:00</td>
<td>4:00</td>
<td>20:00</td>
<td>4:00</td>
<td>20:00</td>
<td>4:00</td>
<td>20:00</td>
</tr>
<tr>
<td>22:00</td>
<td>6:00</td>
<td>22:00</td>
<td>6:00</td>
<td>22:00</td>
<td>6:00</td>
<td>20:00</td>
</tr>
</tbody>
</table>

5 CONCLUSION

The successful implementation of this simulation model showed that it was possible to save $187,200 annually (Bureau of Labor Statistics 2012) over a three year period by reducing the initially predicted number of emergency room physicians. By simply changing the existing schedules slightly it would be possible to maintain the same level of service while expanding the size of the emergency department. The use of the simulation model was key in presenting face validity by being able to replicate the current state. And once validated, the use of optimization was able to determine the optimal future scheduling parameters.

In this paper, we summarize the needs of simulation based optimization in healthcare. We also introduce the key concepts and practical implications of using simulation based optimization to help the users identify the need and model the problems appropriately. Clearly, simulation based optimization seems to have tremendous value in identifying an optimal solution among a diverse set of alternatives. In spite of simulation based optimization being a young field it has seemed to find significant value already in other industries. As healthcare continues to identify ways to operate efficiently, simulation based optimization can prove to be very valuable in staffing studies, facility redesign to determine number of exam rooms, clinic appointment calendar optimization to increase provider efficiencies, supply chain management for balancing the inventory levels in pharmacy, medical decision making to identify the right angle of beam treatment and many others areas. Hopefully the continued growth and refinement of optimization packages within simulation modeling software create more opportunities for its use.

REFERENCES


Mohan Lal, Roh, and Huschka


**AUTHOR BIOGRAPHIES**

**TARUN MOHAN LAL** is a Principal, Practice Redesign and Analytics for the Mayo Clinic Outpatient Practice Operations. Prior to this role he worked as a research analyst for center for science of healthcare delivery ad internal business consultant at Mayo Clinic where he supported organizational strategic and operational planning. His current work at Mayo Clinic is focused on application of advanced analytics and operations research to improve care delivery operations. He has a B.S. in Industrial Engineering from the Manipal University, India, an M.S. in Industrial Engineering from the Texas A&M University and is currently pursuing his Ph.D in Industrial Engineering and Systems Science from State University of New York. His email address is mohanlal.tarun@mayo.edu.

**THOMAS P. ROH** is a Senior Health Services Analyst for the Mayo Clinic Robert D. and Patricia E. Kern Center for Science of Health Care Delivery. Both his prior work at Emory University Hospital and current work at the Mayo Clinic has centered on operations research. He has a M.S. in Health Systems Engineering from the Georgia Institute of Technology and a B.S. from the University of Nebraska. His research has primarily focused on appointment scheduling and resource staffing in healthcare operations. His email address is roh.thomas@mayo.edu.

**TODD R. HUSCHKA** is a Principal Health Systems Analyst for the Mayo Clinic Robert D. and Patricia E. Kern Center for Science of Health Care Delivery. His primary interests are simulation modeling, optimization and statistical analysis relating to improvements in health care systems. He completed his MS in Industrial Engineering Decision Science/Operations Research at the University of Wisconsin in Madison, Wisconsin. His email address is huschka.todd@mayo.edu.