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

EVALUATING HEALTHCARE SYSTEMS WITH INSUFFICIENT CAPACITY TO MEET DEMAND

Sachin R. Pendharkar

Departments of Medicine and Community Health Sciences
University of Calgary
7007-14th Street Southwest
Calgary, AB, T2V 1P9, CANADA

Diane P. Bischak

Haskayne School of Business

University of Calgary
2500 University Drive Northwest
Calgary, AB, T2N 1N4, CANADA

Paul Rogers

Department of Mechanical and Manufacturing Engineering
University of Calgary
2500 University Drive Northwest
Calgary, AB, T2N 1N4, CANADA

ABSTRACT

Modeling healthcare systems using discrete-event simulation (DES) provides the flexibility to analyze both their steady-state and transient performance. However, there has been little work on how best to measure healthcare system performance in cases where there is at least one unstable and lengthening queue in the system, so that traditional steady-state measures such as mean queue length or mean time in queue are meaningless. Using the example of an academic sleep disorders clinic, the authors discuss some of the challenges in constructing a DES model of a healthcare system that has a growing waiting list due to insufficient capacity in one or more areas. Specific considerations include: bottleneck identification through pre-analysis, how to determine a meaningful warm-up period, and the selection of performance measures given system instability.

1 INTRODUCTION

Due to constrained resources, many healthcare systems are unable to meet demand for care, especially when such demand is itself difficult to constrain. Consequently, patients experience delays for care, and delays are likely to worsen over time (Flemons et al. 2004, White et al. 2005, Barratt et al. 2010), at least for some classes of patient.

Since most of the work on healthcare performance measurement employing discrete-event simulation (DES) modeling emphasizes steady-state analyses, problems concerning healthcare systems in which queues are unstable have not been as thoroughly explored. Steady-state output analyses of DES models use average values for traditional performance measures, such as mean queue length or mean time in system, but these performance measures are not defined for systems with unstable queues that are lengthening due to insufficient capacity. Thus, the analysis of unstable systems requires special consideration of the type of analysis to be performed and of the particular model performance measures to be chosen.

Sleep disorders are one example of a disease group where there is often high demand but limited resources to meet that demand. The focus of this study was the Sleep Centre (SC) at the Foothills Medical Centre in Calgary, Alberta, Canada, where waiting times from referral to assessment increased by approximately 50% from 2007-2010. This paper outlines some important methodological issues related to system instability that arose during a larger DES study of the SC (Pendharkar 2011). The main contribution of this paper is to highlight these issues and to describe the necessary departure from steady-state analysis that was required in the assessment of model performance.

This paper is organized as follows: Section 2 presents a review of the application of DES to problems in healthcare systems, as well as a discussion of approaches to the analysis of systems with insufficient capacity using DES; Section 3 provides an overview of the general management of sleep disorders and a description of the SC, including the confirmation that it was an unstable system; Sections 4 and 5 describe the model and the specific considerations for performance analysis; Section 6 outlines some key results from the DES model; and Section 7 provides some conclusions and directions for future work on unstable systems.

2 LITERATURE REVIEW

There have been many applications of DES to healthcare systems, including a considerable number with appointment scheduling problems as a prime focus. Through simulation, Vasilakis et al. (2007) demonstrated a reduction in waiting time from referral to initial assessment in a surgical clinic by using a pooled appointment scheduling system rather than scheduling by individual surgeon. In an outpatient primary care clinic, Klassen and Rohleder showed that by scheduling patients whose expected appointment duration was likely to have the least variation at the beginning of a clinic, the amount of time spent waiting by patients and physicians could be reduced (Klassen and Rohleder 1996). The same authors found that clinic scheduling for single-day and multi-day appointment booking systems resulted in similar scheduling rules to minimize patient waiting times and physician idle time (Klassen and Rohleder 2004).

Resource allocation in healthcare systems has also been explored via simulation. In their DES analysis of the British Columbia Cancer Agency Ambulatory Care Unit, Santibáñez and colleagues showed that flexible allocation of rooms to different oncologic specialty clinics, rather than a 'pod-based' allocation system, could result in a 25% decrease in room requirements for a given patient volume as well as a 70% reduction in waiting times (Santibáñez et al. 2009). Rohleder and colleagues demonstrated that compared to the current system of 25 clinical laboratory service centers in the Calgary Health Region, redistributing available resources across 12 service centers would reduce mean waiting times by approximately 50 minutes – including a 98% reduction in the number of patrons experiencing prolonged waits – due to matching of pooled resource capacity and patient demand (Rohleder et al. 2007).

Much of the previous work using DES has focused on single clinics with fairly uniform patient populations. Elkhuizen and colleagues sought to generalize the modeling process by applying a DES model produced for a neurology clinic to gynecology and preoperative clinics in the same hospital (Elkhuizen et al. 2007, Edward et al. 2008). However, Elkhuizen merely applied the same model to another uniform patient population, rather than simultaneously addressing a variety of patient types. Additionally, the only performance measure studied was access for new patient consultations, without consideration of follow-up assessments or other operational aspects of the clinics. In contrast, Santibáñez and colleagues built an outpatient DES model that incorporated various oncologic specialties with different resource requirements and appointment lengths (Santibáñez et al. 2009).

While the literature abounds with papers employing DES to make important contributions to the understanding of healthcare system operations, there are no published studies of outpatient care that incorporate diagnostic testing and clinical care pathways. An example of a system requiring integrated services is the care of patients with sleep disorders. Current challenges in the field of sleep medicine include long waiting times for diagnostic testing and specialist assessment. Flemons and colleagues described deficiencies in resources for the care of patients with obstructive sleep apnea (OSA) in a number of countries, highlighting the long waiting times that resulted (Flemons et al. 2004). One solution that has been accepted in many jurisdictions is the use of ambulatory testing with portable monitors (PM) rather than laboratory-based diagnosis of OSA. Diagnosis using PM has been found to be comparable to in-laboratory testing for the diagnosis of OSA (Whitelaw et al. 2005) and demonstrated similar clinical outcomes when used in clinical pathways for the management of patients with OSA (Mulgrew et al. 2007). For uncomplicated OSA, Antic and colleagues reported that clinical care by specialist nurses could result in similar clinical outcomes as specialist physician assessment (Antic et al. 2009), with a possible reduction in cost.

The literature examining alternative pathways for sleep care has thus far emphasized clinical rather than operational performance measures, and has focused on OSA rather than the full spectrum of sleep disorders seen at most sleep centers. More recently, operational analysis was used to facilitate the amalgamation of two sleep centers in Manitoba (Hathout et al. 2011). This system redesign resulted in access improvements for Manitobans with sleep disorders. However, this project involved the simultaneous implementation of multiple system changes, making it difficult for the effects of any one improvement strategy to be ascertained. Thus, while new efforts are being made to understand the operational aspects of sleep medicine, there is still a lack of generalizable research in this area.

Unlike the physical queues in manufacturing systems that necessarily create an upper bound on the number of waiting items due to lack of space, waitlists in healthcare systems are virtual queues that can grow indefinitely long. Hence, it may be that healthcare systems with insufficient capacity will become more prevalent as the demands placed on them increase in the future. Systems with insufficient capacity have attracted some attention in non-healthcare settings, but have mostly explored the identification of stability conditions that must hold so as to avoid designing unstable systems, rather than the important questions of how to analyze the performance of an existing unstable system and how to analyze the output of a simulation of such a system. As Schmeiser (2004) pointed out, there is actually no method to determine whether a simulation performance measure, such as an average queue length, is finite, and it is very difficult to distinguish a system that is in the initial transient phase from one in which queues are growing without bound. Wieland et al. (2003) discussed the difficulty of correctly classifying a queueing network as stable or unstable; they presented a stability-checking algorithm but then stated that their algorithm performed poorly when judged on the basis of their own criteria.

Nazzal et al. (2008) developed analytic stability conditions for the design of a conveyor-based automatic material handling system and used DES to validate them, but for unstable configurations the only validation possible was to observe growing queues. Brandão and Porta Nova (2003) discussed the absence of literature on the output analysis of non-stationary DES, and found that ARIMA (p, d, q) models could be effective as meta-models of the differentiated time series of average sojourn times and average queue lengths in unstable M/M/1 and M/M/2 queuing simulations. Kaczynski et al. (2012) developed computer code to calculate measures of performance for the transient M/M/s queue under conditions of stability or instability. The measures of performance included the probability distribution of the number of customers an arrival sees and the sojourn time distribution for a given customer in queue when $k \geq 0$ customers are in the system initially. However, such results do not yet exist for more complex systems for which simulation is the only practical approach to performance analysis. Below, we provide evidence on the instability of the SC healthcare system and develop appropriate performance measures for the simulation of the system, given that it is unstable.

3 SYSTEM DESCRIPTION

3.1 Overview of the Management of Sleep Disorders

Sleep disorders are highly prevalent; population estimates for the two most common sleep disorders, insomnia and obstructive sleep apnea syndrome, are as high as 10% and 4%, respectively (Young et al. 1993, Morin et al. 2006). While primary care physicians often diagnose and manage patients with sleep disorders, referral for assessment by a sleep specialist physician or for advanced sleep diagnostic testing is common. Sleep specialist physicians typically come from various medical subspecialties but usually have advanced training in sleep medicine.

Sleep disorders are broadly classified as respiratory or non-respiratory. Respiratory sleep disorders are diagnosed using either in-laboratory polysomnography (PSG) or portable monitoring of breathing during sleep (PM). The clinical management of patients with respiratory sleep disorders is generally performed by sleep specialist physicians, with more complicated patients being managed by those with specific training in respiratory sleep medicine. Respiratory sleep disorders are often treated with positive airway pressure (PAP) therapy. Some patients with more complex respiratory sleep disorders are also treated with oxygen therapy.

Non-respiratory sleep disorders are frequently managed by primary care physicians, but assessment by a sleep specialist physician may be indicated for disease that does not respond well to treatment, or in patients with more complex medical problems. Treatments for non-respiratory sleep disorders include psychological and cognitive behavioral therapy for insomnia, and pharmacologic agents for other non-respiratory sleep disorders.

3.2 Foothills Medical Centre Sleep Centre

The SC is a publicly-funded academic sleep centre in Calgary, Alberta. Approximately 2500 referrals are received annually, and over 5000 patient visits occur each year. Diagnostic services include over 1000 PSG tests and almost 2000 portable monitoring tests.

Diagnostic testing occurs on five nights per week, resulting in a weekly capacity of 20 PSG and 50 PM tests. Clinical services are provided by six respiratory physicians, one neurologist, two general internists and one general practitioner. Each physician holds one or two clinics per week, which do not occur when the physician has on-call responsibilities or vacation scheduled. Alternate care providers (ACP) include a psychologist for patients with insomnia and respiratory therapists (RT) for patients receiving PAP therapy for a respiratory sleep disorder. These alternate care providers commonly assess patients who are referred by other practitioners at the SC, although they may assess newly referred patients who meet very specific referral criteria allowing them to bypass a physician assessment.

An overview of patient flow through the SC is provided in Figure 1. All patients arrive to the SC by referral from another physician and are processed through a two-step central intake and triage system. New referrals are prioritized as urgent, semi-urgent or non-urgent based on the severity of daytime sleepiness, co-morbidities, previous sleep testing and occupation. Referrals are first assigned an urgency rating through the "Primary Triage" process, which uses only the information provided in the referral. Patients who are classified as "Primary Urgent" based on the primary triage process are scheduled for physician assessment with highest priority and do not undergo any testing in advance of this appointment. All remaining referred patients are sent a questionnaire that is used to determine triage priority through a "Secondary Triage" process. All patients processed through the "Secondary Triage" process who have not had PM testing prior to referral undergo PM testing before the initial clinician visit occurs. There is one appointment slot with a respiratory physician each week that is reserved for an urgent patient; this slot can be used by a "Primary Urgent" patient or by a patient who is classified as urgent through the "Secondary Triage" process, but is not available to be used by semi-urgent or non-urgent patients. The allocation of patients to a particular physician is based on the suspected diagnosis (patient type) and the physician's expertise. For some suspected diagnoses, more than one physician has the necessary expertise; for these situations, the patient is assigned to the next available of these physicians.

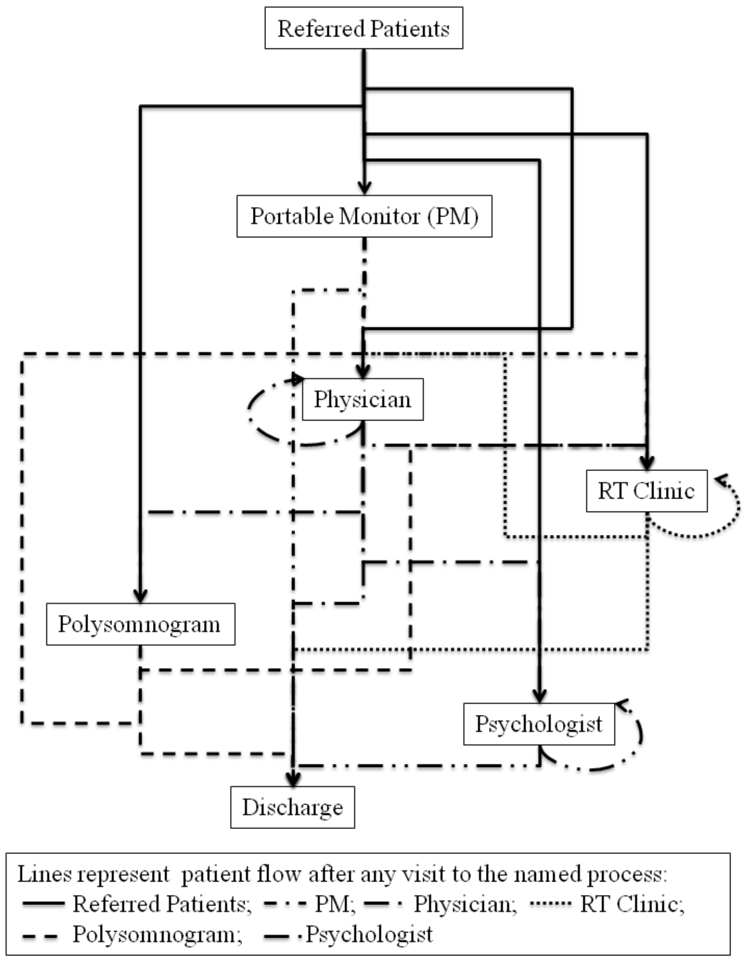

Figure 1: Overview of Patient Flow at the SC

Respirologists who do not practice sleep medicine at the SC are able to order portable monitoring without a referral for assessment. Polysomnography can only be ordered by sleep specialists who practice at the SC, although they may order and interpret PSG tests for certain non-sleep respirologists without a referral for specialist assessment. PSG tests are prioritized as high or normal urgency at the discretion of the ordering physician. Patients who are referred by their respirologist for PM or PSG testing without a referral for sleep specialist assessment do not undergo triage prioritization and usually leave the system after the requested sleep test is completed.

After the initial assessment by a physician or ACP, patient trajectories vary significantly depending on the presenting complaint and medical history. Many patients who were initially seen by a physician are referred to see an ACP for further management of their sleep disorder. Additionally, patients often undergo repeated PM testing or PSG testing to clarify the nature and severity of their sleep disorder. Other return visits include further visits for testing or clinical assessment as is deemed necessary by the clinical provider and patient. These visits are usually for discussion of a sleep test result, follow-up of the patient's response to treatment for a sleep disorder, or to explore new sleep problems or treatment difficulties experienced by the patient. Depending on the diagnosis, patients whose sleep problems have resolved with treatment are either discharged from the clinic or followed indefinitely.

3.3 System Data

System data was obtained from computerized clinical databases and an electronic appointment scheduler. This data was used for preliminary system analysis and as input data for the DES model. Two types of input data were obtained from the computerized databases. First, patient clinical and visit data was obtained for a representative sample of 150 out of 1582 actual SC patients who were referred in 2007. The charts for these 150 patients were reviewed to obtain triage urgency, referral diagnosis and the results of sleep tests that accompanied the referral. All visit data and any data from testing performed at the SC was obtained from electronic databases. Final sleep diagnosis was not available in clinical databases or in the patient's chart; however, since the allocation of patients to a physician is primarily determined by triage information such as the referring diagnosis and urgency rating, it was deemed acceptable to continue without this information.

Second, system data such as physician availability for clinics and probabilities of cancelled or missed appointments were calculated using at least 20 months of historical data in 2008-10. This timeframe for data collection was judged to be appropriate as it was recent but also extended far enough in the past to provide longer term estimates of these variables. For data that was unavailable from these sources, simplifying assumptions and approximations were made in consultation with expert providers and managers at the SC. Examples included the number of days of advanced notice given by a patient when cancelling an appointment, and the desired follow-up interval for patients who were to return to clinic. For the former, the cancellation date was defined using a uniform distribution, whereas for the latter, the desired interval between visits was rounded down from the actual interval between visits, to the nearest of six weeks or one, two, three, six or twelve months.

3.4 Confirmation of Growing Queues at the Sleep Centre

A preliminary data review was undertaken to provide some early insights into system performance. After discussion with SC managers, one important insight that arose from this analysis was the increase in queue lengths and waiting times that had occurred over time. Whereas patients who were referred to the Sleep Centre and classified as non-urgent waited approximately 12 months for initial assessment in 2007, the average delay had increased to 17 months by 2009. Over the same time period, the waiting time for an appointment with a respiratory therapist had increased from two weeks to 12 weeks for patients on PAP therapy. Once assessed by a physician, patients requiring non-urgent PSG testing experienced a delay of approximately 6 months. The delay for PSG testing had not changed from 2007-9. Clerical staff at the SC identified further increases in delays for patients who required a follow-up visit with a sleep physician after PSG testing. Waiting times for follow-up visits could not be measured using the existing system data, but the increasing delays were consistently noted by many staff and managers at the Sleep Centre. Thus, increasing delays were observed at two points in the patient flow pathway.

SC administrators believed that growing queues were due to inadequate capacity of both physicians and testing beds for in-laboratory polysomnography. To explore this assertion, utilization predictions were calculated for each diagnostic testing process, physician group and ACP type. Table 1 shows the resource utilization estimates, which were based on demand estimates for the input dataset of 150 patients and capacity estimates from the 2008-10 system data. These revealed estimates greater than one for two physician groups (obstructive sleep apnea (OSA)-only physician and respirologist) and for the PSG process. Additionally, utilization estimates for the psychologist and the respiratory therapists were over 0.90.

According to the SC administrators, the main source of system congestion had historically alternated between the respiratory physician and PSG process, depending on the number of respiratory physicians practicing at the SC. The OSA-only physician group, which was identified as the most highly utilized resource, had never been a major source of congestion. Furthermore, although the psychologist position was relatively new, there were no problems identified with access to this provider despite a utilization estimate of 0.96. Thus, there were concerns that the utilization predictions might not accurately represent what had been observed at the SC. Further discussion with providers and managers at the SC revealed that this discrepancy was likely due to differences in the time periods for the demand and capacity estimates. Specifically, patients from the dataset were referred in 2007, at which time the respirologist and OSA-only physicians held more clinics and were thus more likely to see referred patients compared to 2008-10. Since patients always have follow-up visits with the same physician to whom they were initially assigned, the demand estimate would have remained disproportionately higher than the capacity estimate from 2008-10. However, using 2007 data for physician clinics would also have been inappropriate, since many patients who were referred in 2007 would have waited at least one year to be assessed.

Table 1: Predicted Utilization

Resource	Utilization
Diagnostic Tests	
Portable Monitor	0.93
Polysomnogram	1.14
Provider	
Respiratory Therapist	0.91
Respirologist	1.27
Neurologist	0.50
OSA-only Physician	1.59
General Internist	0.88
Psychologist	0.96

4 MODEL DETAILS

The DES model was constructed using the Arena version 13 (Rockwell Automation Technologies, Inc., Wexford PA) software package. This is the most commonly used DES platform at the University of Calgary.

Patients were represented as entities in the model, while test resources and clinical providers were modeled either explicitly as resources or as model variables representing resource availability. Based on historical data revealing a mean of approximately 51 referrals per week without any seasonal variation, arrivals were modeled using a Poisson random variable with a mean of 51 patients per week. Arriving patient entities were assigned clinical characteristics and visit requirements corresponding to a randomly selected patient from the dataset.

As patients flowed through the model, they entered submodels for each visit. Within a submodel, a patient underwent appointment scheduling using known scheduling rules, waited for the appointment and proceeded to have the scheduled visit. All cancellations and no-shows also occurred within the submodels. Once a patient entity completed a visit, the entity emerged from the relevant submodel and any relevant performance data pertaining to the visit was collected. The assigned visit trajectory was then evaluated to determine whether the patient should be routed through subsequent visits or discharged from the model.

Due to the complexity of certain scheduling procedures, simplifying assumptions were made when they were considered unlikely to have a significant effect on model performance. The DES model of the SC was verified using a number of established verification techniques (Law 2007), including: modular modeling, with isolated verification of each submodel before adding it to the larger model; variation of model input parameters such as arrival rates, patient entity characteristics and clinical pathways; interactive debugging of the model during a simulation replication; and review of a dashboard that presented model results during the simulation. Validation was performed through regular review of the input data, conceptual and actual model, and model performance measures with a content expert panel of SC managers, physicians and staff.

One of the challenges in comparing system data from the SC to model results was the unavailability of accurate or complete system data for the performance measures selected for this analysis. This problem has been previously described (Sargent 2010). Although the database was modified to collect such data, the long delay between referral and assessment for most patients did not allow for this data to be used in the current study. We reviewed model results, and comparisons of these results to system data for the 150 patients in the input dataset, with the expert panel as a group and individually, and model outputs were thought to accurately reflect system performance.

The base case model was compared to a case in which prioritization by triage urgency was replaced by a 'first-in-first-served' queue policy for initial triage and for all resources. Additionally, while the system and base case model had a reserved slot on Wednesday for an urgent patient, this scenario did not have any reserved slots based on urgency. The scenario was called NoTriage. The DES model was also used to test other improvement strategies, which have been excluded from this discussion of methodologic issues.

5 CONSIDERATIONS FOR UNSTABLE SYSTEMS

5.1 Warming Up the Model

Although the system that was modeled did not reach a steady state, it was necessary to warm up the model in order to reduce, as far as possible, any initialization bias due to the model starting in an empty and idle state. By graphing the number of patients 'in-progress' (WIP) for the entire model and for each resource over 50 simulation runs of 10 years' duration, the WIP for two physician resources and for the overall model could be seen to increase linearly after an initial rapid increase, whereas a steady-state was achieved for all other resources. The WIP levels for the respirologist, neurologist and entire model were observed to increase linearly after an initial rapid increase. Thus, the warm-up period was defined as the point after which the number of patients within a process step reached a plateau or began to increase linearly. Since the processes reached a plateau or linear increase at different times during the simulation, the longest of these times (1500 days) was used as the warm-up period.

5.2 Selection of Performance Measures

As already noted, the unstable nature of the system meant that outcome measures that assumed a steady-state condition, such as mean wait time or mean queue length, would be inappropriate. Instead, the analysis of performance had to be based on measures whose values were captured during a narrow interval of time, as well as the change seen in those measures over wider interval of time. We now describe the selection of the performance measures and these time intervals.

Discussions with SC administrators revealed that the most important point in a patient's visit trajectory is the initiation of treatment. Thus, the primary outcome measure chosen was the mean time taken for a new patient to flow through the system from referral to initiation of treatment (T_{Treat}) for all patients initiating treatment during the last month of the simulation. Since specific treatments were not recorded in the clinic databases, treatment initiation was defined as one of the following:

- The first follow-up physician visit after an initial physician visit
- The first visit to an ACP (psychologist for insomnia or respiratory therapist for a respiratory sleep disorder).

Two other outcomes were of interest, and were also measured for patients reaching the outcome in the last simulated month:

- The time to initial physician visit for all patients ($T_{Initial}$)
- The time to polysomnogram testing for those that required it (T_{PSG})

As historical data demonstrated that most non-urgent patients flowed through the system within one to two years, a replication length (after warm-up) of three years was selected. In addition to measuring the outcomes for patients reaching the outcome during the last month of a replication, the change in each of these performance measures between the first and last months of the three-year period was also measured as an estimate of the rate at which the system's performance was worsening over time.

The decision to measure T_{Treat}, $T_{Initial}$ and T_{PSG} in the last simulated month to describe the predicted performance of the SC after three years was well received by the SC managers and clinicians, who were interested in what the system would look like in absolute rather than relative terms. For this type of measurement, it was important that all model results could be compared to the system in both its current state and its predicted state if no system changes were made over three years. Thus, the same warm-up conditions were used for each scenario, with any configuration changes occurring after the warm-up was completed. Furthermore, during model validation it was demonstrated that model performance in the first simulated month was similar to the actual system, thereby making any changes in performance under each scenario suggestive of how the system would change in the future.

5.3 Identification of Bottlenecks and Alternative System Configurations

Due to concerns regarding the accuracy of the utilization estimates, other means of identifying sources of congestion were explored. For each resource, the total WIP during the simulation run was measured. This measure was observed to increase over time for the respirologist and neurologist processes.

As shown in Table 1, the utilization estimate for the polysomnogram resource suggested that it also had insufficient capacity to meet demand. However, there was no time-dependent increase in WIP for the PSG process, suggesting that either the utilization estimate was inaccurate or that the resource constraint at the PSG process was masked by congestion at another constrained resource, upstream of this process.

6 MODEL RESULTS AND DISCUSSION

Tables 2 and 3 show the simulation results for T_{Treat}, $T_{Initial}$, and T_{PSG} in the last simulated month for the BaseCase and NoTriage scenarios, for all patients and urgent patients, respectively. The absolute and percentage change in the values of T_{Treat} and $T_{Initial}$ over the three-year simulation period are reported for all patients in Table 4 and for urgent patients in Table 5. For each scenario, we report the average value obtained across replications along with a 99% confidence interval for the true mean value. The columns labeled "Difference" show the results of two-sample t-tests with 99% confidence intervals.

As seen in Table 2, the average time to the initiation of treatment in the base case scenario in the last simulated month was 273 days. The average time that a patient waited for an initial physician visit was 260 days, and the average time to obtain a polysomnogram was 83 days. Removing the prioritization by triage category did not significantly change T_{Treat}, but it increased the average initial time required for a patient to see a physician to 272 days (although this increase was statistically insignificant). Conversely, the mean time to obtain an initial polysomnogram was reduced to 62 days by removing triage. The combination of the reduction in T_{PSG} and the trending increase in $T_{Initial}$ in the NoTriage scenario confirmed the SC administrators' suspicion of "shifting bottlenecks" between the respirologist and PSG processes. It also confirmed that the utilization calculation was correct in demonstrating the constraint on the polysomnography resource, and that the lack of increase in WIP for the PSG process in the model was due to a masked bottleneck that was exposed when the constraint on respirologist capacity was relieved. The phenomenon of bottleneck servers demonstrating lower than expected utilization due to upstream congestion has been described previously (Wieland et al. 2003).

Table 3 demonstrates that T_{Treat} and $T_{Initial}$ increased significantly for urgent patients in the NoTriage scenario. Since there was a large number of urgent patients in the patient sample that was used as input data for patient arrivals, the overall increase in $T_{Initial}$ when prioritization by triage urgency was removed was likely related to the increase in $T_{Initial}$ for urgent patients. Specifically, most urgent patients can only be assessed by a respiratory physician due to the complexity of their condition. In contrast, non-urgent patients can often be assessed by one of a number of physicians. Thus, the removal of prioritization by triage urgency significantly worsened $T_{Initial}$ for a large number of patients that would have been classified as urgent and who were required to visit the most constrained physician type. These patients waited in the queue to see a respiratory physician, but were no longer prioritized by their urgency category, nor were they able to access reserved slots. In contrast, non-urgent patients were able to see the next available of a number of physicians, resulting in no significant change in the time to initial physician visit for non-urgent patients.

As is shown in Table 4, in the base case scenario both the average time to initiation of treatment and the average time to see a physician increased over time, by 72 and 83 days on average, respectively, representing significant increases of 37% and 47%. Under the NoTriage scenario, these measures increased to 43% and 55%. This result is not surprising, as is seen in Table 5, since urgent patients experienced disproportionately longer delays under the NoTriage scenario due to the requirement to see a respiratory physician. Since the respiratory physician was the primary bottleneck, model performance worsened more quickly over the three-year simulation period when the triage process was removed, compared to the BaseCase scenario in which urgent patients were of higher priority and in which there were reserved appointment slots for urgent patients.

Table 2: Simulation results (in days) for all patients reaching outcomes in the last simulated month (mean and 99% confidence interval)

Outcome	BaseCase	NoTriage	Difference
T_{Treat}	273 (262,284)	274 (265,283)	1 (-13,15)
$T_{Initial}$	260 (244,276)	272 (263,281)	12 (-7,31)
T_{PSG}	83 (76,90)	62 (55,69)	-21 (-31,-11)

Table 3: Simulation results (in days) for urgent patients reaching outcomes in the last simulated month (mean and 99% confidence interval)

Outcome	BaseCase	NoTriage	Difference
T_{Treat}	227 (218,235)	339 (326,352)	112 (97,127)
$T_{Initial}$	150 (147,154)	282 (272,291)	132 (122,142)

Table 4: Absolute (in days) and percentage change in T_{Treat} and $T_{Initial}$ over 3-year simulation period for all patients (mean and 99% confidence interval)

Outcome	BaseCase	NoTriage	Difference
T_{Treat} (Absolute Change)	72 (60,84)	82 (71,92)	10 (-7,26)
T_{Treat} (% Change)	37 (30,43)	43 (37,50)	7 (1,12)
$T_{Initial}$ (Absolute Change)	83 (66,100)	96 (85,106)	13 (-7,33)
$T_{Initial}$ (% Change)	47 (37,57)	55 (48,62)	8 (0,16)

Table 5: Absolute (in days) and percentage change in T_{Treat} and $T_{Initial}$ over 3-year simulation period for urgent patients (mean and 99% confidence interval)

Outcome	BaseCase	NoTriage	Difference
T_{Treat} (Absolute Change)	28 (18,38)	138 (125,152)	111 (94,127)
T_{Treat} (% Change)	15 (11,18)	70 (64,76)	55 (50,61)
$T_{Initial}$ (Absolute Change)	17 (12, 23)	148 (138,158)	131 (119,143)
$T_{Initial}$ (% Change)	14 (10,17)	112 (105,119)	98 (92,105)

Since the performance of the system was worsening over time due to insufficient capacity, traditional steady-state measures such as mean queue length or the average time taken for a patient to flow through the system would have been inappropriate. By using performance measures that included the time of measurement or the rate of change over a period of time, meaningful comparisons could be made between scenarios. The rate of change of a performance measure (after warm-up) demonstrated how the system would be expected to change over time under each condition.

7 CONCLUSION

The combination of budget cuts and increasing demands on a limited pool of healthcare resources have left many healthcare systems with insufficient capacity to meet these demands. Such constraints lead to delays that grow over time, causing consternation among healthcare professionals, administrators and most importantly, patients. While DES has been used to improve flow through healthcare systems, the focus has remained on steady-state analyses. Traditional steady-state performance measures are inappropriate for the analysis of unstable systems, in which performance is worsening over time.

There is a need to identify other ways of measuring the performance of unstable systems. We have demonstrated one approach, which involved the selection of performance measures that depend on the time that they are measured, or are measured as a rate of change over time. This type of analysis allowed for alternative scenarios to be compared despite the presence of a number of constrained resources and increasing delays for care.

We have also discussed our approach to other challenges related to the analysis of an unstable system. First, the determination of an appropriate warm-up period in the DES model was still important, given that the SC did not begin in an empty and idle state. Second, bottlenecks could not be identified easily, given that long-term utilization predictions may not have accurately reflected the system constraints over time.

We propose that further consideration be given to the analysis of unstable healthcare systems. While we have demonstrated one approach, there are certainly others that merit consideration for this challenging problem. Additionally, future research could expand on the current work by addressing related issues that also cannot be resolved using traditional steady-state performance measurement, such as the identification of shifting or masked bottlenecks when multiple resources are constrained. As healthcare institutions increasingly look to simulation methodologies to address their problems with access, it will be critical that modelers have experience with systems that have insufficient capacity.

ACKNOWLEDGMENTS

The authors would like to acknowledge the staff at the Foothills Medical Centre Sleep Centre, for their assistance with data collection. This work was supported by the Alberta Heritage Foundation for Medical Research, the Health Quality Council of Alberta, and the Western Regional Training Centre.

REFERENCES

Antic, N. A., C. Buchan, A. Esterman, M. Hensley, M. T. Naughton, S. Rowland, B. Williamson, S. Windler, S. Eckermann and R. D. McEvoy. 2009. “A randomized controlled trial of nurse-led care for symptomatic moderate-severe obstructive sleep apnea.” *Am J Respir Crit Care Med* 179:501-508.

Barratt, H., M. Wilson, F. Moore and R. Raine. 2010. "The implications of the NICE guidelines on neurosurgical management for all severe head injuries: systematic review." *Emerg Med J* 27:173-178.

Brandão, R. M. and A. M. O. Porta Nova. 2003. “Non-stationary queue simulation analysis using time series.” In *Proceedings of the 2003 Winter Simulation Conference*, Edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 408-413. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Edward, G. M., S. F. Das, S. G. Elkhuizen, P. J. M. Bakker, J. A. M. Hontelez, M. W. Hollmann, B. Preckel and L. C. Lemaire. 2008. "Simulation to analyse planning difficulties at the preoperative assessment clinic." *Br J Anaesth* 100:195-202.

Elkhuizen, S. G., S. F. Das, P. J. M. Bakker and J. A. M. Hontelez. 2007. "Using computer simulation to reduce access time for outpatient departments." *Qual Saf Health Care* 16:382-386.

Flemons, W. W., N. J. Douglas, S. T. Kuna, D. O. Rodenstein and J. Wheatley. 2004. "Access to diagnosis and treatment of patients with suspected sleep apnea." *Am J Respir Crit Care Med* 169:668-672.

Hathout, L., T. Tenbergen, E. Giannouli, H. Clark and D. Roberts. 2011. “Applying systems engineering to create a population centred sleep disorders program.” In *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management*, 939-944. Kuala Lumpur, Malaysia, IEOM Research Solutions Pty Ltd.

Kaczynski, W. H., L. M. Leemis and J. H. Drew. 2012. "Transient queueing analysis." *INFORMS Journal on Computing* 24:10-28.

Klassen, K. J. and T. R. Rohleder. 1996. "Scheduling outpatient appointments in a dynamic environment." *J Oper Manag* 14:83-101.

Klassen, K. J. and T. R. Rohleder. 2004. "Outpatient appointment scheduling with urgent clients in a dynamic, multi-period environment." *Int J Serv Ind Manag* 15:167-186.

Law, A. M. 2007. *Simulation Modeling and Analysis*. 4th ed. Boston: McGraw-Hill, Inc.

Morin, C. M., M. LeBlanc, M. Daley, J. P. Gregoire and C. Merette. 2006. "Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors." *Sleep Med* 7:123-130.

Mulgrew, A. T., N. Fox, N. T. Ayas and C. F. Ryan. 2007. “Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study.” *Ann Intern Med* 146:157-166.

Nazzal, D., A. Johnson, H. J. Carlo and J. A. Jimenez. 2008. “An analytical model for conveyor based AMHS in semiconductor wafer fabs.” In *Proceedings of the 2008 Winter Simulation Conference*, Edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 2148-2155. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pendharkar, S. R. 2011. “Improving Patient Access: Using Computer Simulation to Increase the Operational Efficiency of an Academic Sleep Centre.” M.Sc. thesis, Department of Community Health Sciences, University of Calgary, Calgary, Alberta. http://dspace.ucalgary.ca/bitstream/1880/48730/1/2011_Pendharkar_MSc.pdf

Rohleder, T. R., D. P. Bischak and L. B. Baskin. 2007. "Modeling patient service centers with simulation and system dynamics." *Health Care Manag Sci* 10:1-12.

Santibáñez, P., V. Chow, J. French, M. Puterman and S. Tyldesley. 2009. "Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency’s ambulatory care unit through simulation." *Health Care Manag Sci* 12:392-407.

Sargent, R. G. 2010. “Verification and validation of simulation models.” In *Proceedings of the 2010 Winter Simulation Conference*, Edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 166-183. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schmeiser, B. 2004. “Simulation output analysis: a tutorial based on one research thread.” In *Proceedings of the 2004 Winter Simulation Conference*, Edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 162-170. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Vasilakis, C., B. G. Sobolev, L. Kuramoto and A. R. Levy. 2007. "A simulation study of scheduling clinic appointments in surgical care: individual surgeon versus pooled lists." *J Oper Res Soc* 58:202-211.

White, P. J., H. Ward, J. A. Cassell, C. H. Mercer and G. P. Garnett. 2005. "Vicious and virtuous circles in the dynamics of infectious disease and the provision of health care: gonorrhea in Britain as an example." *J Infect Dis* 192:824-836.

Whitelaw, W. A., R. F. Brant and W. W. Flemons. 2005. “Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea.” *Am J Respir Crit Care Med* 171:188-193.

Wieland, J. R., R. Pasupathy and B. W. Schmeiser. 2003. “Queueing network simulation analysis: queueing-network stability: simulation-based checking.” In *Proceedings of the 2003 Winter Simulation Conference*, Edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 408-413. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Young, T., M. Palta, J. Dempsey, J. Skatrud, S. Weber and S. Badr. 1993. "The occurrence of sleep-disordered breathing among middle-aged adults." *N Engl J Med* 328:1230-1235.

AUTHOR BIOGRAPHIES

SACHIN PENDHARKAR is an Assistant Professor of Medicine and Community Health Sciences at the University of Calgary. He is a pulmonary and sleep physician. In 2011, he completed his M.Sc. in health services research with an interest in patient flow. Current research interests include the application of operations research to healthcare systems and the identification of efficient patient flow pathways to improve access to outpatient care. His email address is sachin.pendharkar@albertahealthservices.ca.

DIANE BISCHAK is an Associate Professor of Operations Management in the Haskayne School of Business, University of Calgary. She obtained her Ph.D. in industrial and operations engineering at the University of Michigan. Her research interests include health care operations management, inventory management, and statistical quality control. Her email address is diane.bischak@haskayne.ucalgary.ca.

PAUL ROGERS is an Associate Professor in the Department of Mechanical and Manufacturing Engineering at the University of Calgary. His research interests include models for the design and analysis of healthcare systems, and production planning and control systems in manufacturing. He is a Professional Engineer registered with the Association of Professional Engineers and Geoscientists of Alberta (APEGA), and serves on the Editorial Board of the International Journal of Computer Integrated Manufacturing. He holds Ph.D. and M.Eng. degrees from Cambridge University in England. His email address is rogers@ucalgary.ca.