Abstract Number: 025-1379

Abstract Title: **An Empirical Examination of Scheduling Decisions in Peri-operative Services**

Sriram Venkataraman

Department of Management, 100 Sirrine Hall

Clemson University, Clemson, SC 29634. Email: svenkat@clemson.edu

Lawrence D. Fredendall

Department of Management, 100 Sirrine Hall

Clemson University, Clemson, SC 29634. Email: flawren@clemson.edu

Kevin M. Taaffe

Department of Industrial Engineering, 130A Freeman Hall

Clemson University, Clemson, SC 29634. Email: taaffe@clemson.edu

Nathan Huynh

Department of Civil and Environmental Engineering, 300 Main Street, Room C116

University of South Carolina, Columbia, SC 29208. Email: huynhn@cec.sc.edu

POMS 23rd Annual Conference Chicago, Illinois, U.S.A. April 20 to April 23, 2012.

Title: **An Empirical Examination of Scheduling Decisions in Peri-operative Services**

Abstract

We examine variability buffers in peri-operative system (POS) of a hospital using the Newsvendor Model and an econometric framework. The implications of managers actions based on their perceptions of the value of POS resources are explored using this econometric framework. Preliminary results indicate that the managers' view idle time costs as being less important than over-time costs in operating room (OR), and in Post Anesthesia Care Unit (PACU), while they view idle time costs as being more important than over-time costs in Pre-Op. Further, we use this econometric framework to examine the relationships between three buffers - inventory, capacity, and lead time - and system variance into the analysis of the POS.

1 Introduction

This paper makes three contributions to our understanding of healthcare services. First, it generalizes the model of a manufacturing system as four variables - three buffers and system variance (Hopp and Spearman, 2004) – to healthcare services. Second, it extends the methodology developed by Olivares et al. (2008) to examine how operating room managers make scheduling decisions to examine decisions about the entire Perioperative Services (POS). Third, it enriches the model proposed by Hopp and Spearman (2004) by integrating variables proposed by Hayes et al. (1988) to achieve a richer understanding of the decisions managers make about the POS in particular and healthcare services in general. The data for the study was gathered during a case study of a POS department of a major teaching hospital in the southeast to understand how to improve process efficiency.

Hopp and Spearman (2004) model decisions made about operations in manufacturing firms as affecting just four variables. The variables are the levels of three buffers - inventory, capacity and lead time - that protect the system from the fourth variable which is variance. They argue that as the variance in the system increases, the levels of all three buffers will increase and the managers choice is to determine which buffer(s) will increase. Healthcare is a very complex system with a large amount of variance internally and with large variance in demand (e.g., Argote, 1982). At the same time the resources within the healthcare system are very expensive so that there is pressure on managers to utilize them fully. There is also pressure on hospital managers to increase patient satisfaction and to publically report satisfaction scores gathered by independent firms. So, healthcare managers are experiencing pressure to reduce the inventory buffer (which consists of patients waiting for the service), the lead time buffer for the completion of the service and the capacity buffer while they increase demand variance by maximizing revenue

by providing a wide range of services.

This paper uses the methodology developed by Olivares et al. (2008) to evaluate the relative value that managers of a POS unit placed on undertime and overtime in the operating room (OR). In their research, Olivares et al. (2008) demonstrated that the "newsvendor problem" could be used to analyze data about the actual time to perform procedures versus the scheduled time to perform procedures. This was one of the first uses of the newsvendor model to determine managers' perceived valuation of costs. The output of their econometric analysis suggested that POS managers considered undertime costs to be less expensive than overtime costs in the OR. We confirm that this technique measures managers' perception of the relative costs of overtime to undertime in the OR and we extend their analysis to study the scheduling decisions for the Pre-Op and Post Anesthesia Care Unit (PACU) departments which are integral components of the POS.

This paper contributes to our understanding of operations theory by using the insights from the decision analyses conducted in this case to understand how managerial decisions about the ten decision variables that Hayes et al. (1988) stated that managers must address to achieve world class manufacturing status affect the three buffers and system variance used by Hopp and Spearman (2004) to model the system. As will be explained in section 2, while some of these decision variables overlap with Hopp and Spearman's (2004) four variables (i.e, three buffers and system variance), several are unique. By doing so, this case study places the concepts of Hopp and Spearman (2004) within the context of service operations strategy. The rest of this paper is organized in the following way: Section 2 is literature review, section 3 explains data, methodology, and hypothesis, section 4 covers results and discussion, and section 5 includes conclusions, and limitations.

2   Literature Review

Three streams of research are reviewed here. First, the research that places the work of Hopp and Spearman (2004) within the broad context of work design within the operations management field is discussed. Second, the use of the newsvendor model within analytical models in operations research is discussed and then the application of this model by Olivares et al. (2008) is explained in detail. Finally, the stream of literature related to Hayes et al. (1988) framework about managerial decision making is discussed.

Hopp and Spearman (2004) provide an overall framework to analyze the work design of a system by theorizing that the three buffers are related to each other and to the variance in the system. If management chooses not to have a safety capacity buffer, then increases in demand result in a larger inventory buffer and a longer lead time is required. Management choose to limit the work entering the system, which means they can maintain a safety capacity buffer which they argue makes the system easier to control than those systems without a limit on the work.

The term work design was used above because work design is a larger construct than job design. Work design includes all of the systems and processes used to organize the work in an organization including the support services, while job design is concerned about the individual jobs (Sinha and de Ven, 2005). Work design includes the allocation of tasks and responsibilities and the design of systems to ensure effective communication and integration of the various tasks and responsibilities (Daft and Lengel, 1986). Sinha and de Ven (2005) argue that work design is of fundamental importance because it "directly affects the behavior and performance of individual workers and organizations each day, as well as the aggregate productivity and well-being of economies" (p. 389).

Because decisions about each buffer are made at different time points, Hopp and Spearman (2004) say that many managers do not understand the interactions of their decisions on these buffer levels. For example, if a manager wants short lead times for customers, the manager will increase overtime if the lead times grow too large. This overtime increases the capacity buffer which reduces the lead time (i.e., reduces the lead time buffer). As the lead-time gets smaller, the inventory buffer becomes smaller and the manager may respond by reducing the size of the capacity buffer. Hopp and Spearman (2004) expand on this by conceptualizing any operations system as a set of three interrelated buffers to protect the system from variability. The three buffers are an inventory buffer, a capacity buffer and a lead time buffer. The interrelations of these variables is seen through Little's Law. For example, if the inventory buffer is increased by allowing more work into the system then according to Little's Law the lead time is simultaneously increased unless the system throughput is increased by increasing the capacity buffer.

If managers design a process so that there are many waiting points, the lead time buffer has to be longer and the inventory also grows. All systems have at least one bottleneck process wich is usually at the location of the most expensive resource. The safety capacity buffer is at the bottleneck and is the amount of bottleneck capacity that exceeds demand. Queueing theory demonstrates that the safety capacity must be strictly greater than zero (Anupindi et al., 2005). Hayes and Wheelwright (1984) and Hayes et al. (1988) argued that effective execution of operations strategy requires that managers understand how to make coordinated structural and infrastructural decisions to obtain a competitive advantage. Much of the work design research conducted in operations has been concerned with the issues of coordinating work both vertically and horizontally. Depending on the type of environment management controls the workload in the shop using schedules or workload control (see Fredendall et al., 2010 for a complete

4

review).

Hayes et al. (1988) identified four structural and six infrastructural decision variables (see Table 1) that must be coordinated for management to positively affect an operation's capability. As shown in Table 1, six of these decision variables directly affect one or more of the capacity, inventory and lead time buffers and the level of system variance identified by Hopp and Spearman (2004) that create the environment within which work is done. Both the amount of total capacity provided and the allotment of the capacity to different facilities or specialties affect the capacity buffer. These decisions partially determine the bottleneck and consequently the safety capacity buffer and determine the protective capacity in the system. Both the Equipment and Systems provided and the human resource policies and procedures (HR) also affect the level of the capacity buffer. The human resource policies and procedures can affect the capacity buffer. For example policies, which provide training to workers, increase worker capacity and allow for a larger capacity buffer. Also, policies that reduce labor costs by sending workers home whenever there is idle time decrease the labor capacity buffer. The supplier relationship (internal or external) affects the amount of variance in the system since delays in deliveries create variance. If a workload control system or schedule limits work in the system then the non-bottlenecks can create a controlled WIP buffer in front of the bottleneck. A known processing time and WIP buffer allow a relatively accurate prediction of the system flow time and consequently calculation of the led time buffer. The controls used in the system are part of the work design as well as the variance and the buffers. Schedules are a work flow device that can limt both the type and amount of work in the system at any particular time. The production planning and inventory control systems affect the capacity buffer by allocating it to different products; affect the inventory buffer by determining the amount of inventory to hold; and, directly affect the lead time buffer by setting due dates and indirectly affect it through decisions about inventory levels.

5

If the supplier places a low priority on our relationship it is likely that their lead times will be longer which will inflate our lead time and our inventory buffer. In addition if the supplier does not value our relationship there may be variance in the delivery. The quality assurance and control system decisions directly affect the system variance by reducing the number of defects that lead to system variability. The new product development process does not directly affect any of the buffer variables or the system variance. The performance measurement and reward systems can affect the capacity buffer for example by rewarding managers who immediately send workers home to avoid idle time, which would reduce the capacity buffer. In similar manners it can affect lead time and inventory buffers. The organizational structure and design creates a hierarchy to process information and make decisions, which may directly affect the lead time buffer.

Multiple authors have demonstrated the relationship between variability and the work-in-process inventory (e.g., Anupindi et al., 2005) and the time for a job to serviced (e.g., Hopp et al., 2007) by using variations of the Pollaczek-Khintchine formula (Heyman and Sobel, 1982). Indeed, Lovejoy (1998) suggests that these queueing relationships could form the foundation of a theory of operations management that will allow findings from other theories to be integrated. However, a moderating influence on the use of queueing theory to explain performance of a process is the task environment. As the need for problem solving increases there is the need for more decentralized decision-making, which requires different ways of coordinating labor resources. In a complex system managers must coordinate multiple resources to achieve their desired outcomes (Galbraith, 1973). Work design typically divides tasks to create specialized labor resources, and then utilize some mechanisms to reintegrate or coordinate these specialized resources (Lovejoy, 1998). A key responsibility of managers is to coordinate the multiple resources used to accomplish work. The level of coordination required depends on the level of variance

6

Table 1: Decisions Related to Buffers and Variance

	Variability Buffers (Hopp and Spearman, 2004)			
Decision Variable (Hayes et al., 1988)	Capacity	Inventory	Lead Time	Variance
Amount of Total Capacity	X			
Allotment of Capacity	X			
Equipment and Systems	X			
Human Resource Policies and Procedures	X			
Supplier Relationship				X
Quality Assurance and Control Systems				X
Production Planning and Inventory Control Systems	X	X	X	
New Product Development Processes				
Performance Measurement and Reward Systems	X			
Organizational Structure and Design			X	

and the demands on the resources, but given the demand and variance in the system the coordination methods will determine how large the buffers need to be.

There have been many papers in the Operations Research and Operations Management literature that have analytically used the "newsvendor model". Newsvendor model is used when the demand is a random variable. Using the notations by Olivares et al. (2008), let D be a random variable with a known distribution F(·) and let Q be the decision made by the newsvendor before D's realization takes place. If $Q > D$, then the newsvendor incurs a cost of $C_o(Q - D)^+$ and if $D > Q$, then the newsvendor incurs a cost of $C_u(D - Q)^+$; where C_o and C_u are overage and underage cost parameters respectively, and C_o and $C_u > 0$; and $(z)^+ = \max\{z,0\}$. The objective of the newsvendor problem is to find the optimal Q^* that minimizes the expected total cost, i.e.

$$Q^* = \underset{Q}{\operatorname{argmin}} \; \mathrm{E}[C_o(Q - D)^+ + C_u(D - Q)^+] \tag{1}$$

It is known (see Zipkin, 2000) that the optimal solution to equation 1 satisfies

$$F(Q^*) = \frac{C_u}{C_u + C_o} = \frac{1}{1 + \gamma} \tag{2}$$

where $\gamma = C_o/C_u$ is the cost ratio. γ is used instead of the explicit costs C_o and C_u to avoid identification problems later in the analysis.

Most of the existing papers have used the newsvendor problem in analytical settings, where (generally) the distribution function F(·) of the random variable D is known, cost parameters C_o and C_u are known, and the optimal value of Q (Q^*) is computed. However, there are a few exceptions. Cohen et al. (2003) studied the order-fulfillment process of a company producing customized devices such as defense equipment and medical devices etc. A company producing this type of product (hereon referred to as

the supplier) usually receives "soft orders" consisting of shared forecasts. This helps the supplier decide when to start fulfilling the orders given its "stochastic internal manufacturing lead time" (p. 1653, Cohen et al., 2003). This decision of when to start fulfilling the orders is similar to a newsvendor problem since if the supplier starts fulfilling the orders too early, then it will incur an extra holding cost; whereas if the supplier starts fulfilling the orders too late, then it will incur a delay cost. Based on the assumption that the supplier acts "rationally" the authors found that the supplier views the "holding cost to be about three times higher than the delay cost" (p. 1653, Cohen et al., 2003).

Olivares et al. (2008) used the assumption that the decision-maker (newsvendor) is rational (so, Q^* is known), and the distribution function $F(\cdot)$ of the random variable D is known to empirically estimate the managers' preference cost ratio (γ), the costs of over-reserving (i.e., staff have idle time) versus under-reserving time (i.e., staff work overtime) for surgical procedures in an OR. Their data set was limited to decisions about the OR for cardio-vascular surgery. They did not consider the other functions inside POS. They found that the median cost ratio γ was less than 1.0 suggesting that the costs of undertime were considered by managers to be substantially less than the costs of overtime (i.e., managers were more willing to pay for undertime than overtime), so managers preferred to incur undertime costs rather than overtime costs.

3   Data, Methodology, and Hypothesis

This paper extends the work of Olivares et al. (2008) to the entire POS and to all types of surgical procedures within the POS. The first step in the analysis was to study the POS in detail to understand the system. Staff in each of the three POS departments were shadowed by members of the research team and

time studies of each activity performed at each step in the POS's set of activities were conducted. The researchers followed the staff member for a period of two to eight hours depending on the repetition in their activities. These were used to create process flow maps of each unit in the POS to identify work flows, the sources of variability and the coordination mechanisms used to achieve coordination at each stage of the POS. The technique of shadowing to gather detailed data about the actual activities of a job has been used by many researchers (e.g., Fredendall et al., 2009; Yule et al., 2006). The flow of patients within the POS (see Figure 1) is from Preoperative (Pre-Op), to the Operating Room (Intraoperative) to Post Anesthesia Care Unit (PACU). The supporting departments schedule the OR, supply medicine, supply sterile supplies and sterilize instruments after surgeries. It was determined that instrument sterilization, pharmacy and sterile supplies were not affecting OR operations since they were able to maintain an inventory buffer between themselves and the OR and they had very short lead times for special requests from the OR and they were not studied further.

Figure 1: Perioperative Services: Patient and Supporting Flows
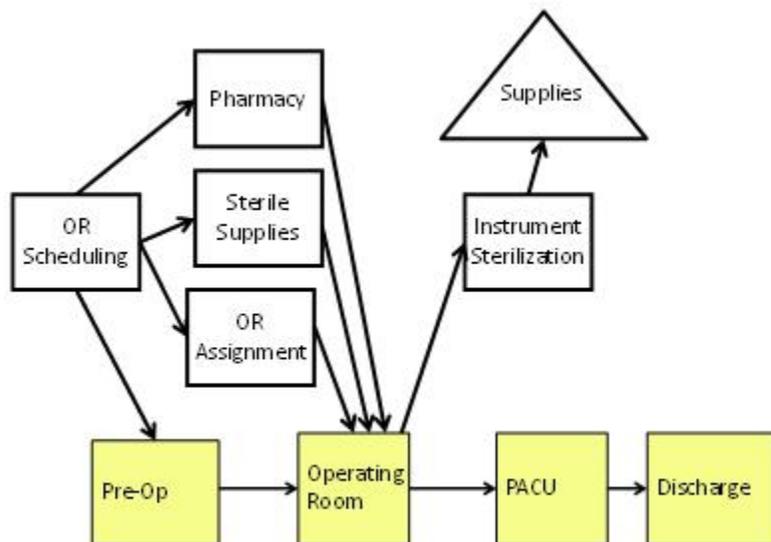
The second step in the analysis was to evaluate the OR performance using archival data. Four weeks of data from May 2011 were used to calculate the average utilization, the total on-time starts and the First-Case on-time starts. During these four weeks, there were 1413 cases scheduled to begin and end between 8 AM and 6 PM. This time period is the busiest time period during the 24 hour period. While each of these 1413 cases was a different patient, 530 patients (37.5%) had multiple procedures performed during the scheduled time period.

This POS had twenty-seven operating rooms available each of the 26 days during the month, so there were $10 * 26 * 27 = 7,020$ hours available to schedule during the four weeks. The schedule for the 1413 cases used 2686.48 hours of this available including scheduled room turn around activities. This is a scheduled 38.27% utilization. Of these 1413 cases, 641 cases started on-time (i.e., within 15 minutes of scheduled start time) for a 45.36% average on time starts. Of the 358 cases scheduled to start at 8 AM, 220 had started by 8:15 AM for a 61.45% First-Case-On-Time start rate.

The next step in the analysis was to examine the scheduling decisions made for the POS during one month to evaluate the managerial decisions. For each unit in the POS - Pre-Op, OR, and PACU, similar to Olivares et al. (2008), we assumed that D_is are drawn independently from $F(\cdot, \theta)$, where $\theta \in \Theta$ is a parameter vector and Θ is a parameter space. Following the notations used by Olivares et al. (2008) and the functional econometric form, let

$$\theta_i = h(X_i, \eta) \tag{3}$$

where X_i represents the covariates and η represents the estimator vector. Similarly, let:

$$\gamma_i = g(Z_i, \alpha) \tag{4}$$

where Z_i represents the covariates and α represents the estimator vector. Since γ_i is not known, we cannot estimate α using linear regression. So, using the framework developed by Olivares et al. (2008) and substituting equation 2 into equation 4, we get:

$$F\left(Q_i^*; h(X_i, \eta)\right) = \frac{1}{1 + g(Z_i, \alpha)}. \tag{5}$$

Assuming a log-normal distribution for the cost ratio which is consistent with the medical literature (Strum et al., 1999), we obtain:

$$\log(\gamma_i) = Z_i\alpha + \psi_i. \tag{6}$$

Since γ_i is not observed, we need to use an estimate of γ_i (see Murphy and Topel, 2002), so we use the following 2-step procedure as proposed by Olivares et al. (2008) and Murphy and Topel (2002):

1. First, we estimate $\hat{\theta} = h(X_i, \hat{\eta})$

2. Second, we estimate $\hat{\gamma} = \dfrac{1}{F(Q_i; \hat{\theta}_i)} - 1$ and finally estimate α by $\hat{\alpha}$ using equation 6

Since maximum likelihood technique and ordinary least squares (OLS) will produce the same estimates (see Ruud, 2000), we used OLS to get the estimates of η. The covariates used in both stages of the estimation procedure for the operating room were the surgeon's experience level, the patient's charge level (i.e., acuity level), the case type code (type of surgery), and the number of procedures scheduled. We had data on these covariates for 1045 surgeries in the OR. The covariates for both stages of the estimation procedure used for Pre-Op were the patient charge level (i.e., acuity level) and the case type code (i.e., type of surgery). We had data on these covariates for 895 patients in the Pre-Op. The covariates for both stages of the estimation procedure used for PACU were the PACU acuity level (which is different from OR acuity level), and the case type code (i.e., type of surgery). We had data on these covariates for 500 patients in the PACU.

3.1 Hypotheses

The estimated variable labor cost of an operating room with minimum staffing which includes a CRNA, a circulating nurse and a surgical technician, but excludes the M.D. anesthesiologist is a minimum of $150 an hour. The variable labor costs of the Pre-Op for one RN and one technician is under $50 an hour. The variable labor costs for the PACU for one RN is under $35 an hour. While the variable labor costs for overtime can be estimated at time-and-a-half for each department, the managers evaluation of the ratio of overtime to undertime costs is expected to also include costs that are difficult to quantify, such as the loss of goodwill from a surgeon who must wait to begin a scheduled surgery and then must work late to complete a surgery which started late if there were not adequate staff. So it is expected that mangers will perceive OR overtime to be much more expensive than undertime costs, which means that they will prefer undertime costs to overtime costs. In this situation a γ that is estimated from actual data about the amount of overtime and undertime scheduled will be less than 1.0 if managers prefer undertime to overtime. This expectation is supported by prior researchers who have found that gamma for the OR is less than 1.0 (Strum et al., 1999; Olivares et al., 2008). So, it is hypothesized that:

Hypothesis 1: For operating room, $\gamma < 1$.

As shown in figure 2, the OR is the bottleneck in this system and prior researchers (Hillier and Boling, 1979) have found that throughput is increased if there is protective capacity at non-bottlenecks immediately before and after the bottleneck. So, Pre-Op and PACU should have capacity buffer to increase throughput and hence the scheduling managers of both Pre-Op and PACU would tend to consider idle time to be a lower cost than overtime. As stated above the variable labor costs of the Pre-Op are low compared to the variable labor costs of the OR, so it is expected that management will see undertime

costs in Pre-Op as being less important than overtime costs, since overtime worked in Pre-Op would be in response to Pre-Op causing a delay in the OR. To avoid delaying the OR management will prefer undertime to overtime. So, we hypothesize that:

Hypothesis 2: For pre-op, $\gamma < 1$.

Figure 2: Bowl Phenomenon (Hillier and Boling, 1979)
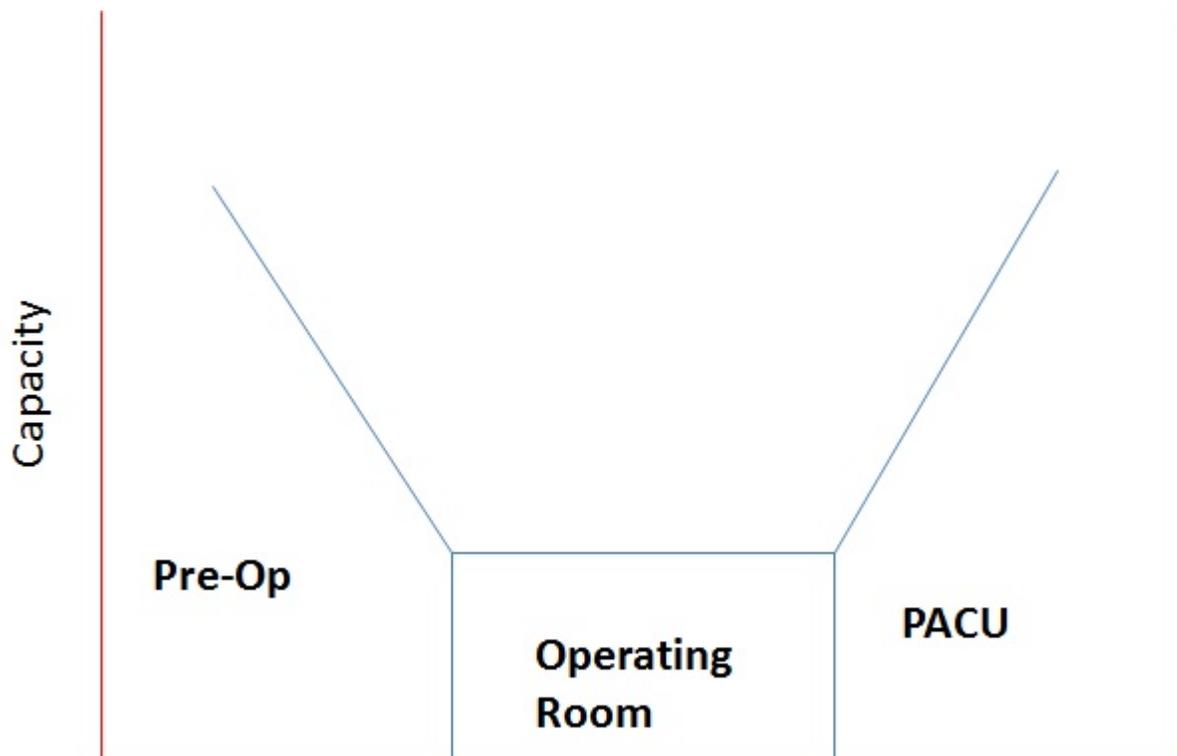
Again, since the variable labor costs of PACU are much lower than the variable labor costs in the OR, management will ensure that PACU does not create delays for the OR. For this reason management will view undertime costs in the PACU to be more allowable than overtime costs. So their preference for undertime will be greater than for overtime, and it is hypothesized that:

Hypothesis 3: For PACU, $\gamma < 1$.

4 Results and Discussion

As stated earlier, the first step in the study was a case study where we examined the performance of each department in the POS. The first observation was that the workload into the system was controlled. While there are some emergency surgeries that occur when needed that are not controlled, the OR scheduling department strives to level the number of surgeries that occur daily. This leveling creates a stable POS workload and variance from this source is reduced, which allows the relevant buffers to be smaller and to be stable themselves (Hopp and Spearman, 2004). However, there was variance from other sources. The Pre-Op and PACU had more capacity than the OR. The Pre-Op department had enough capacity to have the patients prepared well in advance of their scheduled surgery. All the observed delays in the Pre-Op department were because events out of their control such as the late arrival of lab reports or delays by the surgeons or anesthesiologists in visiting the patient to mark the site of the surgery or to update the Health and Physical report.

Similar observations of the PACU during this time period showed that it had more capacity than the OR. This protective capacity meant that the PACU was able to accept patients from the OR with less than 15 minutes notice. The staff reported that the PACU had not stopped an OR in the last five years of POS operation. To do this the POS uses extra capacity from Pre-Op as flexible, surge capacity for PACU. The Pre-Op nurses have the ability to move from Pre-Op to PACU and the supervisory staff can also work with patients if required.

The POS uses lead-time, inventory and capacity buffers. The first lead-time buffer is in the scheduling of day of the surgery. Prior to the scheduled day of their surgery, the scheduling office will request

15

patients and surgeons to move their scheduled surgery forward if the POS scheduling office identifies an opening in the OR schedule. To do this the POS scheduling office contacts surgeons and patients to obtain their approval to move the surgery forward. Once the patients arrive in Pre-Op 120 minutes prior to the scheduled start of their surgery, they are told that the current expected start of their surgery is 150 to 400 minutes later (note that the procedures for pediatric surgery are different). This second lead-time buffer results in an inventory buffer of patients. The patients who are prepared to go into the OR serve as an inventory buffer. Many patients wait in the Pre-Op past the scheduled start of their surgery, but some patients start their surgery early when there is room in the OR and the surgeon and staff are available. A second POS inventory buffer is instruments and sterile supplies. These are typically prepared and in the OR core by 3 AM of the morning of the surgery. The surgical technician can then quickly setup the instruments in the room. However, despite these buffers there may be delays in the start of surgery due to a shortage of some special equipment such as portable x-ray machines. The POS also uses a capacity buffer. As calculated earlier, it has a 38.27% utilization. To reduce undertime costs, if there are no problems, the POS managers send staff home on flex time before the end of their scheduled shift. If after some staff are sent home, and overtime from the remaining staff is needed to complete scheduled set of procedures in the OR then overtime is used. So, management decisions about the size of these buffers are confounded and cannot be analyzed individually, since a decision to insert a large lead time buffer in Pre-Op directly leads to an inventory of patients in Pre-Op who are ready for surgery, but whose surgeon is not ready to start surgery.

To examine the thoughts of the managers about overtime versus undertime in the OR, the OR data were analyzed using the approach explained in Section 3. To model the first stage of estimation of

16

operating room, we regressed:

$$\ln(D) = X\beta + \epsilon \tag{7}$$

where D is a vector of actual surgery time and X is the matrix of covariates [surgeon experience level, charge level (acuity level), case type code (type of surgery), and number of procedures]. Then, we used the estimates of $\hat{\beta}$ to calculate

$$\hat{\gamma}_i = \frac{1}{\Phi\left(\frac{\ln(Q_i) - X_i\hat{\beta}}{\hat{\sigma}_i}\right)} - 1. \tag{8}$$

Then, we regressed:

$$\ln(\hat{\gamma}) = Z\alpha + \psi. \tag{9}$$

where $\hat{\gamma}$ is a vector of estimated cost-ratio and Z is the matrix of covariates [surgeon experience level, charge level (acuity level), case type code (type of surgery), and number of procedures].

The coefficient of determination (R^2) for step 1 estimation procedure of OR was 0.3778 with p-value < 0.001; so the covariates capture around 38% of variance in the actual time of surgery. This result is consistent with that reported by Olivares et al. (2008), although they used only cardio-vascular surgeries. Further, all four covariates were found to significantly affect actual surgery time with a 95% significance level. As the experience level of surgeon increases, actual surgery time goes down which is intuitive. Conversely, as the acuity level goes up, the actual time of surgery goes up which again makes sense since higher acuity levels are associated with more complex surgeries. The actual time of surgery also goes up as the number of procedures go up, which is consistent with the findings reported by Olivares et al. (2008). Table 2 shows the summary statistics of $\hat{\gamma}$. The median value of $\hat{\gamma}$ was $0.4133 < 1$. This result is consistent with that of $\hat{\gamma}$ obtained by Olivares et al. (2008). Further, this result supports our first hypothesis and hence the hospital indeed pays more attention to over-time costs as opposed to idle-time costs

17

in case of the OR. R^2 for step 2 estimation procedure of OR was 0.2021 with p-value < 0.001; so the covariates capture around 20% of variance in the estimated cost-ratio. Further, all four covariates were found to significantly affect actual surgery time with a 95% significance level. As the experience level of surgeon increases, estimated cost-ratio goes down, which implies that the hospital tries to overtime costs than idle-time costs, which makes sense because the overtime cost of experienced surgeons are higher. Conversely, as the acuity level goes up, the estimated cost ratio goes up, which means that the hospital tends to avoid idle-time costs for higher acuity surgeries. The estimated cost-ratio also goes up as the number of procedures go up, which again implies that the the hospital tends to avoid idle-time costs for surgeries involving multiple procedures. These results imply that the cost-ratio indeed is heterogeneous and varies on a case-to-case basis in the OR.

As in the case of operating room, we assumed log-normal distribution for actual time spent in Pre-Op. R^2 for step 1 estimation procedure of Pre-Op was 0.1196 with p-value < 0.001; so the covariates capture around 12% of variance in the actual time spent in Pre-Op. The R^2 is smaller than that of OR primarily because we used two covariates only in the case of Pre-Op as against four covariates in the case of OR. Both the covariates were found to significantly affect actual surgery time at 95%. In the case of Pre-Op, as acuity level goes up, actual time spent in Pre-Op goes down since the patients who are under-going more complex surgeries were prepared for surgery before-hand and hence spent lesser time in Pre-Op than those patients who were under-going more complex surgeries. Table 3 shows the summary statistics of $\hat{\gamma}$. After deleting one outlier, the median value of $\hat{\gamma}$ was $1.6626 > 1$. This result does not support our second hypothesis and hence the hospital pays more attention to idle-time costs as opposed to over-time costs in case of Pre-Op. R^2 for step 2 estimation procedure of Pre-Op was 0.7612 with p-value < 0.001; so the covariates capture around 76% of variance in the actual time spent in Pre-Op. Both the covariates

were found to significantly affect actual surgery time at 95%. In the case of Pre-Op, as acuity level goes up, the estimated cost ratio goes up, which means that the Pre-Op manager tends to avoid idle-time costs for higher acuity surgeries in Pre-Op.

We assumed log-normal distribution for actual time spent in PACU based on past literature (Dexter and Tinker, 1995). R^2 for step 1 estimation procedure of PACU was 0.1165 with p-value < 0.001; so the covariates capture around 12% of variance in the actual time spent in PACU. The R^2 is similar to the one for Pre-Op and this is mainly due to the fact that we used two covariates for both Pre-Op and PACU. In the case of PACU, as acuity level goes up, actual time spent in PACU goes up (although not significantly) since the patients who under-went more complex surgeries tend to spend more time in PACU than those patients who under-went less complex surgeries. Table 4 shows the summary statistics of $\hat{\gamma}$. After deleting two outliers, the median value of $\hat{\gamma}$ was $0.8673 < 1$. This result supports our third hypothesis and hence the hospital pays more attention to over-time costs as opposed to idle-time costs in case of PACU. An ongoing problem in the POS is coordination with the nursing floors to free up beds on the floors so patients can be moved out of the PACU. There seems to be intenal disagreement about managing the PACU. At this hospital, the PACU has not blocked the OR in the past five years, but the nurse manager says this is because of staffing flexibility. Whenever there is danger of blocking an OR, they bring in more staffing from other units. R^2 for step 2 estimation procedure of PACU was 0.1053 with p-value < 0.001; so the covariates capture around 11% of variance in the actual time spent in PACU. In the case of PACU, as acuity level goes up, the estimated cost ratio goes up (although not significantly), which means that the PACU manager tends to avoid idle-time costs for higher acuity surgeries in PACU.

5   Conclusions and Limitations

This paper made two important contributions to our understanding of scheduling in a POS. First, it extended the work by Olivares et al. (2008) to the entire peri-operative services unit of a leading hospital in South Carolina. Second, it incorporated the relationships between three buffers and system variance Hopp and Spearman (2004) into the analysis of the POS. This is one of the first attempts (atleast to our knowledge) to tie structural estimation procedure of econometrics with buffering theory of operations management.

The first hypothesis that the POS managers do avoid overtime in the OR and are willing to accept the costs of undertime in the OR was supported. This is a significant management decision, since the cost to keep an OR open is high. Each OR has high labor costs (e.g., at a minimum it is staffed with 1 CRNA, 1 Circulating Nurse and 1 Surgical Technician which costs approximately $150 per hour) and an MD Anesthesiologist is assigned to multiple ORs. While undertime is a significant cost, the analysis suggests that management considered the total cost of undertime to be much smaller than the costs of overtime. However, based on our observations of the use of flex time by the POS managers, it seems as if the managers who have responsibility for the daily operation of the POS attempt to reduce their undertime costs by sending staff home during the day if the manager predicts that there is little likelihood they will be needed. Since the hospital is incurring overtime costs on some of the same days that these managers sent workers home earlier in the day, it is possible that this layer of management places a different value on undertime versus overtime costs than upper management in the POS. Hypothesis 2 that the Pre-Op considers undertime costs as being less important than overtime costs is not supported. Examination of why this situation exists suggests that there is undertime costs with the nurses in Pre-Op,

but the nurses cannot record the patients as being ready for surgery until they have received their visits by the surgeon and/or anesthesiologist and the timing of these visits is outside of the control the Pre-Op and POS managers. Also, while the nurses may have completed their tasks patients can stay in Pre-Op longer while they wait to receive lab reports. These patient delays that occur in Pre-Op then extend the actual stay of the patient in Pre-Op, so that it appears that overtime is seen as costing less than undertime by management. In addition, when the manager who is responsible for staffing considers Pre-Op to be overstaffed, workers are sent home early. Hypothesis 3 that the overtime in the PACU is considered to be a much higher cost than undertime in the PACU is supported. This was also supported by our observations that the PACU is managed to ensure that it never stops a patient from leaving the OR. All of the staff spoke with pride about the PACU record of not having delayed a patient's departure from the OR for the past five years. The PACU may borrow staff if needed from Pre-Op to ensure that it does not block an OR, which indicates that undertime costs are considered less important than overtime costs.

The extension of the econometric analysis of Olivares et al. (2008) to the entire POS brought to our attention that the decisions made by the different managers in the hierarchy may not be compatible. For example, while the scheduling office considers undertime costs to be less significant than overtime costs, the manager charged meeting the daily budget guidelines attempts to reduce undertime costs despite the risks that this will lead to overtime costs.

5.1 Limitations

The use of Olivares et al. (2008) technique directly involves estimating the managers' perceptions of the cost ratio γ empirically. The method of conducting this analysis focuses on determining the managers'

perceptions of the capacity buffer cost. However, the same management decisions about overtime and undertime that are analyzed simultaneously determine the capacity buffer and partially determine the inventory buffer, since favoring overtime suggests a buffer of work waiting to be completed is in place.

5.2   Future Research

Future research is needed to formally test whether γ is less than 1.0 using a consistent bootstrapping technique (Efron and Tibshirani, 1993). This paper used only the median value of γ and did not formally test the three hypotheses developed.

References

Anupindi, R., S. Chopra, S. D. Deshmukh, J. A. Van Mieghem, E. Zemel. 2005. *Managing Business Process Flows*. 2nd ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Argote, L. 1982. Input uncertainty and organizational coordination in hospital emergency units. *Administrative Science Quarterly* **27**(3) 420–434.

Cohen, M. A., T. H. Ho, J. Z. Ren, C. Terwiesch. 2003. Measuring imputed cost in the semiconductor equipment supply chain. *Management Science* **49**(12) 1653–1670.

Daft, R. L., R. H. Lengel. 1986. Organizational information requirements, media richness and structural design. *Management Science* **32**(5) 554–571.

Dexter, F., J. H. Tinker. 1995. Analysis of strategies to decrease postanesthesia care unit costs. *Anesthesiology* **82** 94–101.

Efron, B., R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. 1st ed. Chapman & Hall, USA.

Fredendall, L. D., J. B. Craig, P. J. Fowler, U. Damali. 2009. Barriers to swift, even flow in the internal supply chain of perioperative surgical services department: A case study. *Decision Sciences* **40**(2) 327–349.

Fredendall, L. D., D. Ojha, J. W. Patterson. 2010. Concerning the theory of workload control. *European Journal of Operational Research* **201**(1) 193–207.

Galbraith, J. R. 1973. *Designing Complex Organizations*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Hayes, R. H., S. C. Wheelwright. 1984. *Restoring Our Competitive Edge: Competing Through Manufacturing*. 1st ed. John Wiley & Sons Inc., United States of America.

Hayes, R. H., S. C. Wheelwright, K. B. Clark. 1988. *Dynamic Manufacturing: Creating the Learning Organization*. 1st ed. The Free Press, New York.

Heyman, D. P., M. J. Sobel. 1982. *Stochastic Models in Operations Research, Vol. 1*. 1st ed. Dover Publications, Inc., Mineola, New York.

Hillier, F. S., R. W. Boling. 1979. On the optimal allocation of work in symmetrically unbalanced production line systems with variable operation times. *Management Science* **25**(8) 721–728.

Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operating systems with discretionary task completion. *Management Science* **53**(1) 61–77.

Hopp, W. J., M. L. Spearman. 2004. To Pull or not to Pull: What is the question? *Manufacturing & Service Operations Management* **6**(2) 133–148.

Lovejoy, W. S. 1998. Integrated operations: A proposal for operations management teaching and research. *Production and Operations Management* **7**(2) 106–124.

Murphy, K. M., R. H. Topel. 2002. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* **20**(1) 88–97.

Olivares, M., C. Terwiesch, L. Cassorla. 2008. Structural estimation of the Newsvendor model: An application to reserving operating room time. *Management Science* **54**(1) 41–55.

Ruud, P. A. 2000. *An Introduction to Classical Econometric Theory*. Oxford University Press, New York.

Sinha, K. K., A. H. Van de Ven. 2005. Designing work within and between organizations. *Organization Science* **16**(4) 389–408.

Strum, D., L. Vargas, J. May. 1999. Surgical subspecialty block utilization and capacity planning: A minimal cost analysis model. *Anesthesiology* **90**(4) 1176–1185.

Yule, S., R. Flin, S. Paterson-Brown, N. Maran, D. Rowley. 2006. Development of a rating system for surgeons' non-technical skills. *Medical Education* **40**(11) 1098–1104.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill.

A Appendix

Table 2: Summary Statistics of $\hat{\gamma}$ of operating room

Sample Size	Median	Minimum	Maximum	Mean	Std. Deviation
1045	0.4133	0.001	140.698	1.395	5.523

Table 3: Summary Statistics of $\hat{\gamma}$ of Pre-Op

Sample Size	Median	Minimum	Maximum	Mean	Std. Deviation
894	1.6626	1.1219	7.5269	1.719	0.2513

Table 4: Summary Statistics of $\hat{\gamma}$ of PACU

Sample Size	Median	Minimum	Maximum	Mean	Std. Deviation
498	0.8673	0.2274	174.805	1.749	8.2798