

## Research Article

# The Integration of Group Technology and Simulation Optimization to Solve the Flow Shop with Highly Variable Cycle Time Process: A Surgery Scheduling Case Study

T. K. Wang,<sup>1</sup> F. T. S. Chan,<sup>2</sup> and T. Yang<sup>1</sup>

<sup>1</sup> Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan 70101, Taiwan

<sup>2</sup> Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Correspondence should be addressed to T. Yang; [tyang@mail.ncku.edu.tw](mailto:tyang@mail.ncku.edu.tw)

Received 7 July 2014; Revised 22 August 2014; Accepted 26 August 2014; Published 11 September 2014

Academic Editor: Chiwoon Cho

Copyright © 2014 T. K. Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Surgery scheduling must balance capacity utilization and demand so that the arrival rate does not exceed the effective production rate. However, authorized overtime increases because of random patient arrivals and cycle times. This paper proposes an algorithm that allows the estimation of the mean effective process time and the coefficient of variation. The algorithm quantifies patient flow variability. When the parameters are identified, takt time approach gives a solution that minimizes the variability in production rates and workload, as mentioned in the literature. However, this approach has limitations for the problem of a flow shop with an unbalanced, highly variable cycle time process. The main contribution of the paper is to develop a method called takt time, which is based on group technology. A simulation model is combined with the case study, and the capacity buffers are optimized against the remaining variability for each group. The proposed methodology results in a decrease in the waiting time for each operating room from 46 minutes to 5 minutes and a decrease in overtime from 139 minutes to 75 minutes, which represents an improvement of 89% and 46%, respectively.

## 1. Introduction

Currently, the US healthcare system spends more money to treat a given patient whenever the system fails to provide good quality and efficient care. As a result, healthcare spending in the US will reach 2.5 trillion dollars by 2015, which is nearly 20% of the gross domestic product (GDP). A similar trend is observed by the Organization for Economic Cooperation and Development (OECD), which included Taiwan. The cost of increased healthcare spending will become more important in the coming years. One way to decrease the cost of healthcare is to increase efficiency.

The demand for surgery is increasing at an average rate of 3% per year. To increase access, operating rooms (ORs) must invest in related training for specialized nursing and medical staff. ORs will be a hospital's largest expense at approximately \$10–30/min and will account for more than 40% of hospital revenue [1]. Two types of surgical services

are provided by ORs: reaction to unpredictable events in the emergency department (ED) and elective cases, where patients have an appointment for a surgical procedure on a particular day. This paper considers elective cases because an important part of the variance can be controlled by reducing flow variability [2]. The efficiency of ORs not only has an impact on the bed capacity and medical staff requirement but also impacts the ED [3]. Therefore, increasing OR efficiency is the motivation for this study.

Utilization is usually the key performance indicator for OR scheduling. Maximum productivity requires high utilization. However, in combination with high variability, high utilization results in a long cycle time, according to Little's Law [4], as shown in Figure 1. High utilization and low cycle times can be achieved by reducing the flow variability, as shown in Figure 2. Therefore, the identification and reduction of the main sources of variability are keys to optimizing the compromise between throughput and cycle

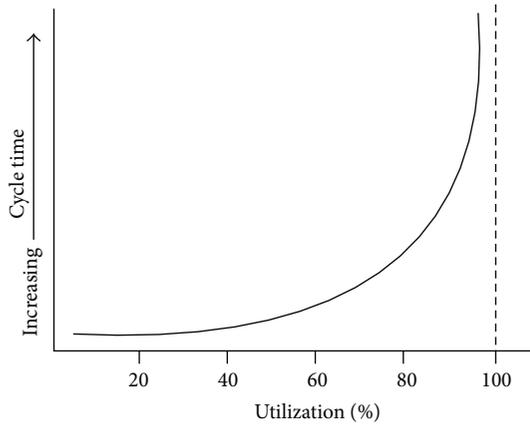


FIGURE 1: Cycle time versus utilization.

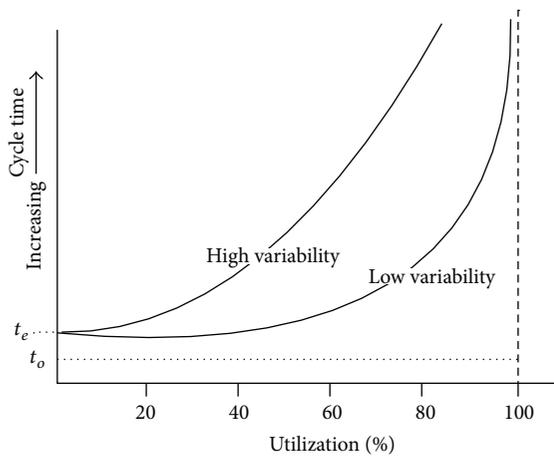


FIGURE 2: The corrupting influence of variability.

time. Unfortunately, a few measures for flow variability are used in ORs. Such a measure would be highly valuable in reducing variability and would allow more efficient study.

The flow variability determines the average cycle time. There are different sources of variability, such as resource breakdown, setup time, and operator availability. An approach proposed by Hopp and Spearman used the VUT equation to describe the relationship between the waiting time as the cycle time in queue ( $CT_q$ ), variability ( $V$ ), utilization ( $U$ ), and process time ( $T$ ) for a single process center [5]. The VUT is written in its most general form as (1). This study determines the parameters and the solutions of this equation:

$$CT_q = VUT. \quad (1)$$

This paper is structured as follows. The analytical VUT equation is applied to a workstation with real surgical scheduling data. The algorithm quantifies the patient flow for the entire OR system and makes the cycle time longer than predicted, due to several parameters. An example then shows the potential of the VUT algorithm for use in cycle time reduction programs. The solution depends on finding the

parameters that cause the cycle time variability. A simulation model is used to demonstrate the feasibility of the solution. Finally, the main conclusions and some remarks on future work are given.

## 2. Literature Review

Timeframe-based classification schemes generally include long, intermediate, and short term processes as follows: (1) capacity planning; (2) process reengineering/redesign; (3) the surgical services portfolio; (4) estimation of the procedural duration; (5) schedule construction; and (6) schedule execution, monitoring, and control [6]. This study focuses on short-term aspects because the shop floor control makes adjustments when the process flow is disrupted by the variability of patients' late arrivals, surgery durations, and resource unavailability in the real world.

The sequencing decision which can be thought of as a list of elements with a particular order and its impact on OR efficiency are addressed in the literature [7, 8]. Most of the studies use a variety of algorithms to improve the utilization under the assumption that the cycle time is deterministic. Studies developed a stochastic optimization model and heuristics to compute OR schedules that reduce the OR team's waiting, idling, and overtime costs [9, 10]. Goldman et al. [11] used a simulation model to evaluate three scheduling policies (i.e., FIFO, longest-case first, and shortest-case first) and concluded that the longest-case first approach is superior to the other two.

Scheduling always struggles to balance capacity utilization and demand in order to let the arrival rate  $r_a$  not exceed the effective production rate  $r_e$  [12–14]. Then, the utilization at each station is given by the ratio of the throughput to the station capacity ( $u = r_a/r_e$ ). Under the assumption that there is no variability, which includes the assumption that cases are always available at their designated start time, the surgery durations are deterministic and resources never break down. However, it is not possible to predict which patients or staff will arrive late, precisely how long a case will take to perform, or what unexpected problems may delay care [15]. This is why none of a variety of research models has had widespread impact on the actual practice of surgery scheduling over the past 55 years [6]. Therefore, this study will consider these flow variability issues.

Studies show that the management of variability is critical to the efficiency of an OR system. McManus et al. [16] noted that natural variability can be used to optimize the allocation of resources, but no empirical model was included in the study. Managing the variability of patient flow has an effect on nurse staffing, quality of care, and the number of inpatient beds for ED admission and solves the overcrowding problem [17, 18]. However, there is a lack of quantitative analysis to demonstrate which flow variability parameter causes the impact. In summary, this study quantitatively analyzes flow variability, determines which parameters have an impact, and provides relevant solutions for empirical illustration.

Womack et al. [19] stated that high utilization with relatively low cycle time requires a minimum variability.

Although this originates from the Toyota Production System (TPS), its potential applications and in-depth philosophy are not well defined [20]. Different industries apply these principles and develop customized approaches to optimize shop floor processes. The methodology of the study refers to Ohno [21], Monden [22], and Liker [23] for details of development. The five-step process is as follows.

The first step defines the current needs for improvement. Key performance indicators are selected. Performance measures for the OR system fall into two main categories: patient waiting time and staff overtime. Patient waiting is associated with two activities: patients waiting for the preparation of a room and waiting for surgery. There is no waiting time for the recovery process because recovery begins immediately after surgery. Late closure results in overtime costs for nurses and other staff members. A reduction in overtime has a positive effect on the quality of care, decreases surgeons' daily hours, produces annualized cost savings, makes inpatient beds available for ED admission, and positively affects ED overcrowding [17].

The second step incorporates an in-depth analysis of the production line. Before starting detailed time studies, standard movements are observed and mapped. Value stream mapping (VSM) is used to design and analyze an OR's process layer [24]. VSM has a wide perspective and does not examine individual processes. The average cycle time is determined by variability, but VSM does not provide quantifiable evidence and fails to determine how methods can be made more viable. Hopp and Spearman proposed the use of the VUT equation. Equation (2) represents the variability as the sum of the squared coefficients of the variation in the interarrival times,  $C_a^2$ , the squared coefficients of the variation in the effective process time,  $C_e^2$ , the utilization,  $U$ , and the squared coefficients of the variation in departure,  $C_d^2$ . The squared coefficient of variation is defined as the quotient of the variance and the mean squared. Therefore,  $C_a^2 = \sigma_a^2/t_a^2$  and  $C_e^2 = \sigma_e^2/t_e^2$ , where  $t_a$  and  $t_e$  are the mean interarrival time and the mean process time, respectively. The effective process time paradigms,  $t_e$  and  $C_e^2$ , include the effects of operational time losses due to machine downtime, setup, rework, and other irregularities. Compared with the theoretical process time,  $t_o$ ,  $t_e > t_o$  and  $C_e^2 > C_o^2$ .  $C_e^2$  is considered low when it is less than 0.5, moderate when it is between 0.5 and 1.75, and high if more than 1.75. Equation (3) shows that, for low utilization, the flow variability of the departing flow equals the variability of the arriving flow, and, for high utilization, the flow variability of the departing flow equals the effective process time variability. The equations give quantifiable evidence of variability:

$$CT_q = \left( \frac{C_a^2 + C_e^2}{2} \right) \left( \frac{u}{1-u} \right) t_e, \quad (2)$$

$$C_d^2 = u^2 C_e^2 + (1-u^2) C_a^2. \quad (3)$$

The third step consolidates the current performance data and determines the baseline for efficiency improvement. Because the period of operating time for this study is from

8:00 a.m. to 5:00 p.m., the total overtime after 5:00 p.m. as the baseline per day is 3,336 minutes.

The fourth step defines implementation methods that satisfy the abovementioned subtargets and use the detailed time studies and data analysis from earlier steps. In summary, (2) and (3) clearly show the contribution of variability. The leveling approach minimizes the variability in production rates and work load [25]. However, a leveling approach that only considers a single production level is not applicable to the problem of low volume and high mix production [26]. Only a few papers outline leveling approaches for flow shop environments [27]. The flow shop with an unbalanced, highly variable cycle time process can be solved by takt time grouping [28]. However, this method assumes that the process time for each batch is the same and is not applicable to this study. This study uses a new method of takt time based on group technology to implement the flow environment.

When all of the improvement items are chosen, the fifth step ensures their sustainable implementation. Discrete-event simulation is used to model the behavior of a complex system. By simulating the process, the system behavior is observed and the potential improvements after changes can be evaluated [29]. However, grouping and leveling are still required to achieve the optimal solution for a given problem.

### 3. Case Description by the Current-State VSM and VUT Equation

*3.1. The Current-State VSM.* The case studied in this paper is from a Taiwanese medical center that has 21,350 surgical cases per year. The surgical department consists of 24 operating rooms, 15 of which are for specialty procedures. In identifying the overall flow shop procedure using the current-state VSM, which includes the processing time for each process, boxes are used to understand the type of activities that occur in the ORs. VSM allows a visualization of the processes for an entire service rather than just one particular process. This result is plotted in Figure 3. The current value stream mapping shows the cycle time, which includes value-added time and non-value-added time. The non-value-added time is the waiting time, which is 46 minutes.

*3.2. The VUT Equation Analysis.* To describe the performance of a single workstation, the following parameters are assumed:

- $t_o$ : the mean natural process time,
- $r_a$ : the arrival rate,
- $\sigma_o$ : the standard deviation for the natural process time,
- $c_o$ : the coefficient of variability for the natural process time,
- $N_s$ : the average number of cases between setups,
- $t_s$ : the mean setup time,
- $\sigma_s$ : the standard deviation for the setup time,
- $t_e$ : the mean effective process time,

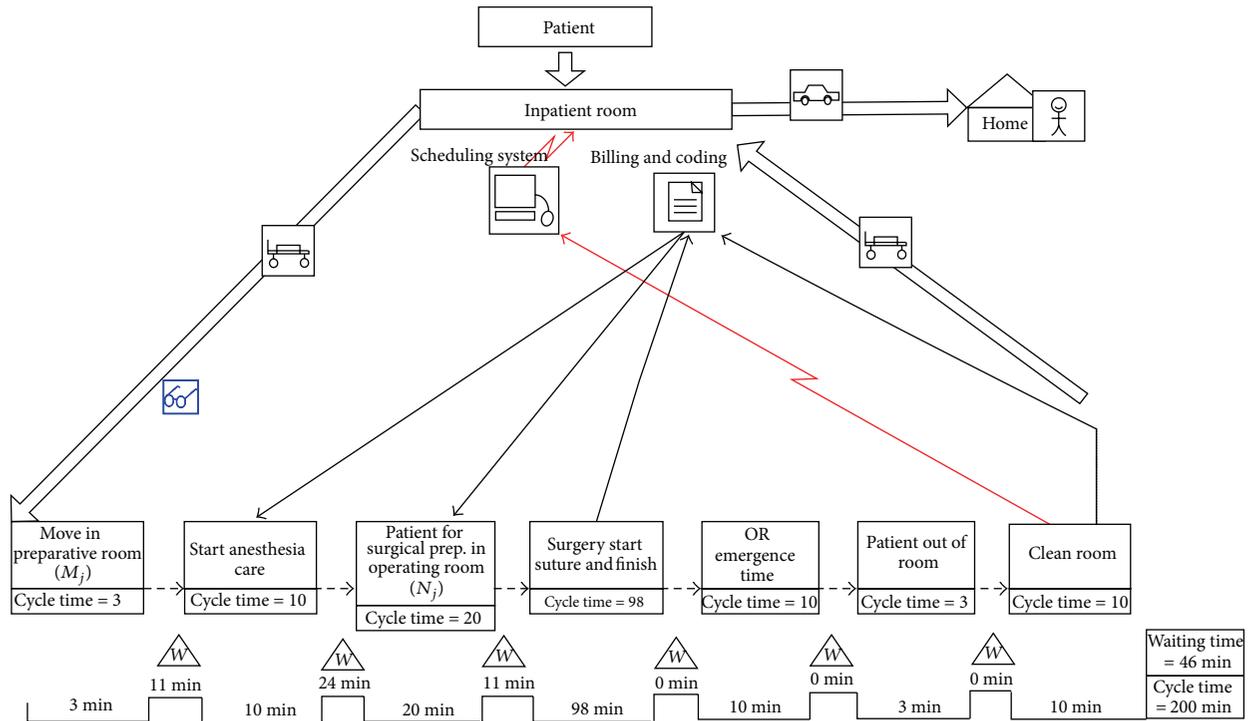


FIGURE 3: The current-state VSM.

- $\sigma_e^2$ : the variance of the effective process time,
- $c_e^2$ : the squared coefficient of the variation in the effective process time,
- $c_a^2$ : the squared coefficient of the variation in demand arrivals.

The daily surgical scheduling has 80 elective cases on average, according to the effective capacity from 8:00 a.m. to 5:00 p.m. Namely, the arrival rate,  $r_a$ , is 8.9 cases/hour. Each patient will go through the two series of stage ( $W_i$ ), which included the process of preparation ( $W_1$ ) and operation ( $W_2$ ). For the worst-case example, at the starting time, patients move into the OR system from wards when the operating room ( $W_2$ ) is ready. Because the ward and the surgical department are far from each other, the interarrival time is assumed to be exponential ( $c_a^2 = 1$ ). The characterizing flow in the ORs' system passes through the two stages ( $W_i$ ) shown in Figure 4. The first stage ( $W_1$ ) checks the patient's documentation, nursing history, and laboratory data. The natural process time mean  $t_o$  is 20 minutes, and the natural standard deviation  $\sigma_o$  is 2 minutes. These result in a natural CV of  $c_o = \sigma_o/t_o = 0.1$ . The capacity of the preparation room ( $M_j$ ) in the first stage is 12, which is less than the value of 24 for the second stage ( $N_j$ ), and this is so for all cases. Using a dispatching rule of first-come-first-served (FCFS) in the first stage ( $W_1$ ), the first stage ( $W_1$ ) can breakdown under certain conditions (e.g., the patient does not arrive at the start time, when the preparation room ( $M_j$ ) is ready, or when the number of patients is greater than 12). These situations are called nonpreemptive outages. Specifically,  $W_1$  has a mean time to failure (MTTF),  $m_f$ , of 60 minutes and a mean time to

repair (MTTR),  $m_r$ , of 35 minutes. MTTF is the elapsed time between failures of a system during operation, and MTTR is the average time required to repair a failed operation. The average capacity of  $W_1$  for nonpreemptive outages can be calculated using (4), where the availability  $A = 60/(60 + 35) = 0.63$ . The effective mean process time,  $t_e$ , calculated using (5) is 31.75 minutes. The utilization of the first stage ( $W_1$ ) is calculated using (6) to be 0.27, and  $c_e^2$  is calculated using (7) as 0.83:

$$A = \frac{m_f}{m_f + m_r}, \tag{4}$$

$$t_e = \frac{t_o}{A}, \tag{5}$$

$$u = \frac{r_a}{r_e} = \frac{r_a t_e}{m}, \tag{6}$$

$$C_e^2 = C_o^2 + 2A(1 - A) \frac{m_r}{t_o}. \tag{7}$$

After the previous patient has left the operating room and following the setup time, the current patient then starts at the second stage ( $W_2$ ). Both the process time and setup time are stochastic and will be commensurate with the complexity of the disease. The natural mean process time  $t_o$  is 120.17 minutes, and the natural standard deviation  $\sigma_o$  is 80.25 minutes. The setup time is regarded as a preemptive outage when they occur due to changes in the following surgery. Trends in the setup time are associated with the type of surgery, and the mean of the setup time  $t_s$  is 25.26 minutes and the standard deviation of the setup time  $\sigma_s$  15.43 minutes.

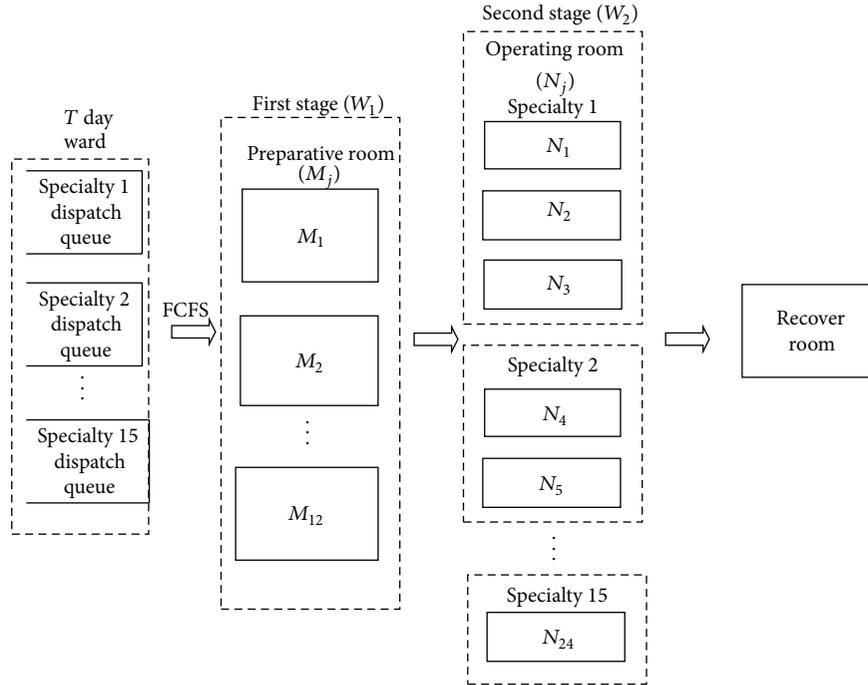


FIGURE 4: The charactering flow in the ORs' system.

The effective mean process time  $t_e$  from (8) is 145.43 minutes. The capacity is 9.9 cases/hour. The utilization of  $W_2$  by (6) is 0.89. Using (9), we can compute  $c_e^2 = 7.49$ . From the VUT equation, we conclude that this is a stable system in the flow shop with an unbalanced, high variation cycle time process. Consider

$$t_e = t_o + \frac{t_s}{N_s}, \quad (8)$$

$$\sigma_e^2 = \sigma_o^2 + \frac{\sigma_s^2}{N_s} + \frac{N_s - 1}{N_s^2} t_s^2, \quad (9)$$

$$c_e^2 = \frac{\sigma_e^2}{t_e^2}.$$

**3.3. The Baseline for Efficiency Improvement.** The third step consolidates the current performance data and determines the baseline for efficiency improvement. Then, the VUT equation for computing queue time  $CT_q$  of  $W_1$  is 10.81 minutes, and  $c_d^2$  is 0.99; however,  $CT_q$  of  $W_2$  is 764.74 minutes. After analysis of the VUT (2), we found that the relative differences among the mean of the effective process time  $t_e$  and utilization compared to the variability are small. The value of 4.24 comes from two parts: the first is  $c_e^2 = 7.49$ , which is highly variable based on the process time in the second stage ( $W_2$ ); the second is  $c_a^2 = 0.99$ , which is equal to  $c_d^2$  from the first stage ( $W_1$ ). The departure variability of  $W_2$  depends on the arrival variability of  $W_1$ . The  $c_e^2 = 0.83$  in the  $W_1$  due to the nonpreemptive outages, which are caused by the interarrival rate from the inpatient ward to the ORs' system. Equations (2) and (3) provide useful models for a

deeper understanding of the worst case of natural and flow variability when access to resources is limiting. In practice, balancing the average utilization and the systemic stresses results in a smoother patient flow. Consider

$$\begin{aligned} CT_q &= \frac{C_a^2 + C_e^2}{2} \frac{u}{1-u} t_e \\ &= \frac{(0.99 + 7.49)}{2} \left( \frac{0.89}{1 - 0.89} \right) 145.43 \\ &= (4.24) (8.09) (145.43). \end{aligned} \quad (10)$$

These are some assumptions in this case study.

- (i) The data in analysis of surgical-specific procedure time is the year of 2002.
- (ii) Each preparation room ( $M_j$ ) and operating room ( $N_j$ ) can process only one case at a time.
- (iii) For this study, there should be totally 24 rooms strictly assigned to the different surgical cases. Each case can be carried out in any of the 24 rooms, but each room must be assigned one group at most.
- (iv) The period of opening of operating room is from 8:00 a.m. to 5:00 p.m. and the overtime is counted after 5:00 p.m.
- (v) Emergency surgeries are not considered. Either patients must have appointments on certain OR days for a medical reason, or any period during which surgeons cannot perform is ignored. In other words, no surgeries are cancelled or added.

- (vi) There is no constraint to surgeons or other staff availability. In other words, surgeons are available at any period of the day (i.e., when a case is moved from the morning to the afternoon).
- (vii) Each physician can only accept one patient at a time. Once the surgery is started, the operation is not allowed to be interrupted or cancelled. Surgical breakdowns are not considered.

#### 4. Proposed Methodology

The fourth step defines implementation methods that satisfy the abovementioned subtargets and uses the detailed time studies and data analysis from earlier steps. Leveling based on group technology consists of two fundamental steps. In the first step, families are formed for leveling based on similarities. Clustering techniques are used to group families according to their similarities. Using these families, a leveling pattern is created in the second step. Every family and every interval is arranged for a monthly period.

*4.1. Group Technology Approach.* It has been shown that variability affects the efficiency of the system. Grouping surgeries minimizes the duration variability of surgery [30]. Of these approaches, cluster analysis is the most flexible and therefore the most reasonable method to employ here. *K*-means is a well-known and widely used clustering method [31]. This method is fast but cannot easily determine the number of groups. If the group is arranged randomly, there will be no obvious difference between each group. Anderberg [32] recommended a two-stage cluster analysis methodology. Ward's minimum variance method is used at first, followed by the *K*-means method. This is a hierarchical process that forms the initial clusters. Ward's method can minimize the variance through merging the most similar pair of clusters among *N* elements. Perform those steps until all clusters are merged. The Ward objective is to find out the two clusters whose merger gives the minimum error sum of squares. It determines a number of clusters and then starts the next step. *K*-means clustering uses the coefficient of variation, which is defined as the ratio of the standard deviation to the mean, as measured by (11). The software SPSS was used for cluster analysis. Consider

$$\text{Coefficient of variation} = \frac{\sigma}{\mu}. \quad (11)$$

*4.2. Takt Time Approach.* Leveling allocates the volume and variety of surgeries among the ORs' resources to fulfill the patient demand over a defined period of time. The first step in leveling is to calculate the takt time, which is measured by (12). The takt time is a function of time that determines how fast a process must run to meet customer demand [28]. The second step is a pacemaker process selection and leveling of production by both volume and product mix [33]. The pacemaker process must be the only scheduling point in the production system and dictates the production rhythm for the rest of the system, where the pace is based on a

supermarket pull system further upstream from this point, as well as First In First Out (FIFO) systems further downstream [34–37]. According to the theory of constraints (TOC), one of the most important points to consider is the bottleneck. Thus, the pacemaker process selection must be located in the second stage ( $W_2$ ). However, the number of resources for each grouping must still be determined to achieve the optimal solution for a given problem. Consider

$$\text{Takt time} = \frac{\text{Available monthly work time}}{\text{Total monthly volume required}}. \quad (12)$$

*4.3. Simulation Modeling and Optimization.* The fifth step ensures sustainable implementation. The simulation tool checks the feasibility of integrating the methods into the current system. Simulation is useful in evaluating whether the implementation of the method is justified [38]. Rockwell Arena, a commercial discrete-event simulator, has been used for many studies [39]. To evaluate potential improvements due to the implementation of takt time based on group technology, Rockwell Arena 13.51 was used to build the general simulation model for the OR system. Depending on the nature and the goal of the simulation study, it is classified as either a terminating or a steady-state simulation. This study is a terminating simulation, which signifies that the system has starting and stopping conditions [40].

This study optimizes the capacity buffers against the remaining variability of each surgical group to minimize OR overtime (i.e., work after 5:00 p.m.). Optimization finds the best solution to the problem that can be expressed in the form of an objective function and a set of constraints [41]. Therefore, the difference between the model that represents the system and the procedure that is used to solve the optimization problems is defined within this model. The optimization procedure uses the outputs from the simulation model as an input, and the results of the optimization are fed into the next simulation. This process iterates until the stopping criterion is met. The interaction between the simulation model and the optimization is shown in Figure 5 [42].

#### 5. Empirical Results

*5.1. Takt Time Based on a Group Technology Approach Clustering Method.* This study focuses on 263 surgical-specific procedures using a Pareto analysis of a total of 1198 types of surgical-specific procedure times in the year 2002. Ward's minimum variance method gives the number of clusters as 5. The following step is segmented into 5 groups, based on Ward's minimum variance method and then *K*-means clustering, to give the time expression shown in Table 1.

*5.2. Takt Time Mechanism.* Leveling is used to calculate the takt time for each surgery group. The surgical department organizes the working time according to a monthly time schedule. The monthly time available is 10,800 minutes, as there are 9 hours a day and 5 days in a week in this case. The monthly volume was measured, and the takt time for each group is shown in Table 2.

TABLE 1: The five groups.

Categories	1	2	3	4	5
Expression	$-0.001 + \text{ERLA} (28.7, 2)$	$-0.001 + \text{LOGN} (119, 226)$	$5 + \text{WEIB} (91, 0.856)$	$5 + \text{WEIB} (162, 1.2)$	$5 + \text{GAMM} (94.3, 1.51)$

TABLE 2: The monthly volume and takt time of each group.

Group	Monthly time available (minutes)	Monthly volume of surgeries (units)	Takt time (minutes)
1	10800	813	$\frac{10800}{813} \approx 13$
2	10800	159	$\frac{10800}{159} \approx 68$
3	10800	134	$\frac{10800}{134} \approx 81$
4	10800	346	$\frac{10800}{346} \approx 31$
5	10800	185	$\frac{10800}{185} \approx 58$

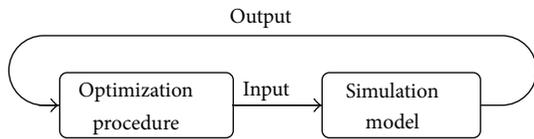


FIGURE 5: Relationship between simulation model and optimization.

5.3. *Simulation Model.* Rockwell Arena 13.51 was used to build the simulation model that represents the OR systems. The computer-based module logic design establishes an experimental platform that allows a decision maker to quickly understand the conditions of the system.

When the simulation model is constructed, we wanted to tighten precision cover on the population mean ( $\mu$ ); the smaller the confidence interval, the larger the number of required simulation replications. The length of one replication is set as one month. The coefficient of variation (CV), which is defined as the ratio of the sample standard deviation to the sample mean, is used as an indicator of the magnitude of the variance. The value of the CV stabilizes when the number of replications reaches 35, as shown in Figure 6 [43]. We generated the input values from probability distributions in Arena. The simulation model used the time expression with the run length of 1 month and 35 replications. Each replication starts with a both empty and idle system. The individual replication result is independent and identically distributed (IID); we could form a confidence interval for the true expected performance measure  $\mu$ . In this study, the mean daily cycle time ( $\mu$ ) and the 95% confidence interval are adopted as the system performance measure. We have an initial set of replications 35; we compute a sample average cycle time, 214.28 minutes, and then a confidence interval whose half width is 1.92 minutes. It is noted that the half width of this interval (1.92) is pretty small compared to the value of the center (214.28). The mathematical basis for the above discussion is that, in the 95% of the cases of making 35 simulation replications as we did, the interval formed like this will contain the true expected value of total population.

TABLE 3: The error between the real system and simulation.

Compare (average)	System	Simulation	Error (%)
Waiting time	46.14	43.10	7

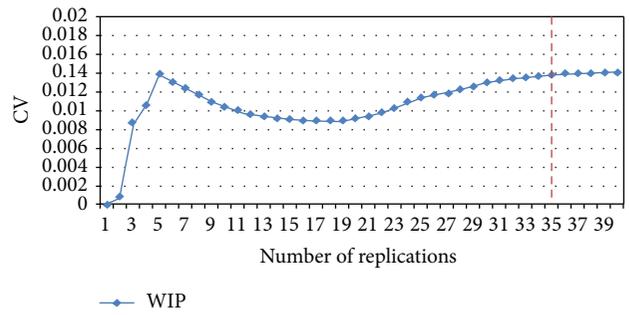


FIGURE 6: The CV chart.

In this study, simulation models for verification and validation are both used. Verification ensures that the model behaves as intended, and validation ensures that the model behaves like the real system. As shown in Table 3, the error between the simulation and the real system in terms of the daily waiting time in each OR is 7%.

5.4. *The Optimal Solution.* Identification of the optimal scenario uses one week in July, which in practice is usually 5 days. On each day, each group,  $i$ , is available and has an expression time. OptQuest is utilized in conjunction with Arena to provide the optimal solution. The required notations for the formulation are defined as follows.

Parameters:

- $i$  = an index for the groups of surgeries,  $i \in I$ ,  $I = \{1, 2, 3, 4, 5\}$ ,
- $j$  = an index for the number of operating rooms,  $j \in J$ ,  $J = \{1, 2, 3, \dots, 24\}$ .

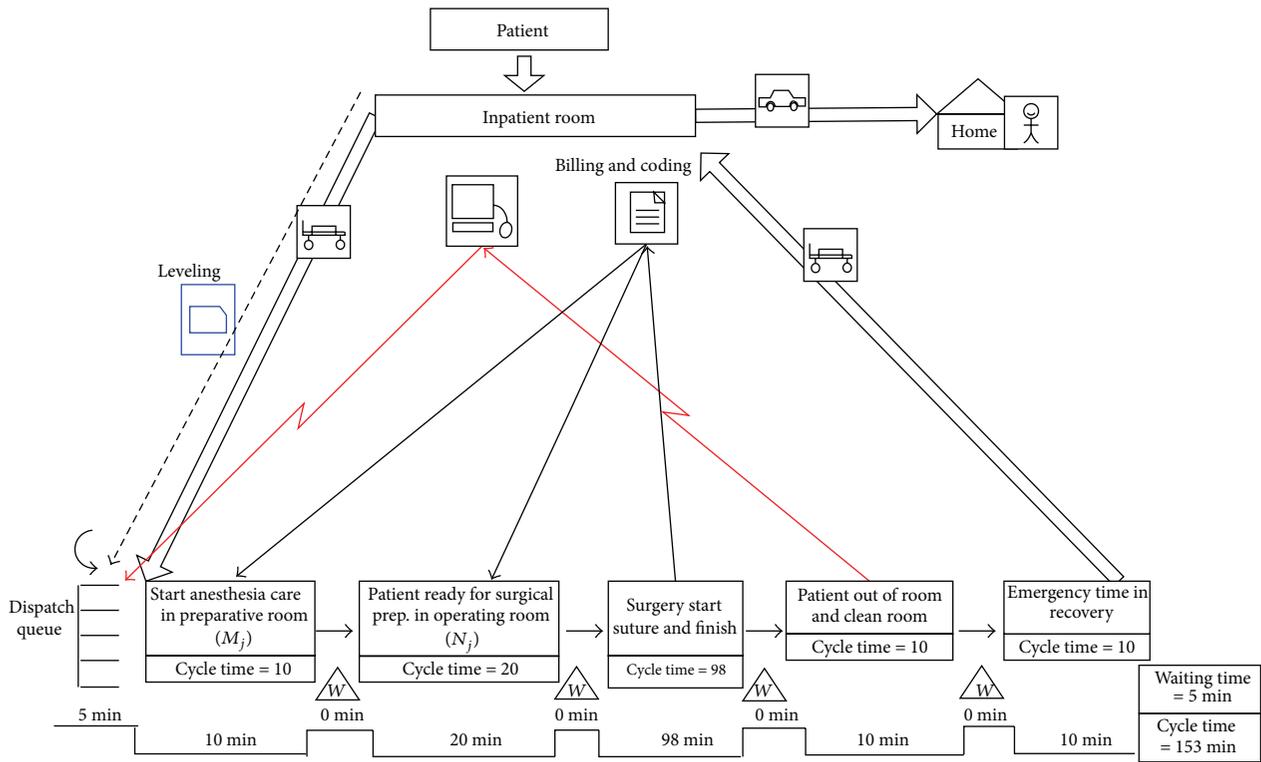


FIGURE 7: The future-state VSM.

Intermediate variables:

$O_j$  = the overtime associated with the ORs.

Decision variables:

$A_{ij}$  = a binary assignment, whether the surgery group,  $i$ , is assigned to operating room  $j$  ( $A_{ij} = 1$ ) or not ( $A_{ij} = 0$ );

$C_i$  = an index for the number of operating rooms that are allocated to the surgery group,  $i$ .

The optimization model solves

$$\text{Minimize } \sum_{j=1}^{24} O_j, \tag{13}$$

subject to the following constraints:

$$\sum_{i=1}^5 A_{ij} = 1 \quad \forall j, \tag{14}$$

$$C_i \geq 1 \quad \forall i, \tag{15}$$

$$\sum_{i=1}^5 C_i = 24, \tag{16}$$

$$A_{ij} \in \{0, 1\} \quad \forall ij. \tag{17}$$

The objective function minimizes the total amount of overtime. Constraint (14) specifies that each operating room must be assigned to one group at most. Constraint (15) ensures that each group is allocated at least in one operating room. Constraint (16) sets the limitation of operating rooms for all groups. Constraint (17) as a binary assignment, is whether the surgery group,  $i$ , is assigned to operating room  $j$ .

5.5. *The Result.* The results are plotted in Figure 7. The capacity buffers optimized against the remaining variability of each group are  $C_1 = 2$ ,  $C_2 = 2$ ,  $C_3 = 8$ ,  $C_4 = 9$ , and  $C_5 = 3$ . In the optimized solution, the computational results show that the waiting time and overtime for each operation room decrease from 46 minutes to 5 minutes and from 139 minutes to 75 minutes, respectively, which is a respective improvement of 89% and 46% as shown in Table 4.

5.6. *Conclusions and Further Research.* Maximizing the efficiency of the OR system is important because it impacts the profitability of the facility and the medical staff. OR scheduling must balance capacity utilization and demand so that the arrival rate,  $r_a$ , does not exceed the effective production rate,  $r_e$ . However, authorized overtime is increasing due to the randomness of patient arrivals and cycle times. This paper differs from the existing literature and makes a number of contributions. It focuses on shop floor control and uses a VUT algorithm that quantifies and explains flow variability. When the parameters are identified, the impact on the

TABLE 4: Optimal results.

	Overtime per operating room (minute)		Waiting time (minute)		Cycle time (minute)	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Original system	139	26	46	16	200	22
Optimal solution	75	2	5	1	153	2
Improvement (%)		46		89		24

surgery schedule using leveling based on group technology is illustrated. A more robust model of surgical processes is achieved by explicitly minimizing the flow variability. A simulation model is combined with the case study to optimize the capacity buffers against the remaining variability of each group. The computational result shows that overtime is reduced from 139 minutes to 75 minutes per operating room.

The most significant managerial implications can be summarized as follows.

- (i) To achieve a higher return on investment, high utilization and reasonable cycle times, which depend on the level of variability, are necessary. The identification and reduction of the main sources of variability are keys to optimizing the performance instead of utilization.
- (ii) This study solves OR scheduling using various heuristic methods and provides the anticipated start times for each case and each operating room. However, most real cases violate the assumptions (e.g., all cases are not ready at the start time, cycle times are stochastic, and resources do not break down, etc.). The schedule cannot be accurately predicted once the assumptions are violated.
- (iii) Sequencing patients using takt time based on group technology reduces the flow variability and waiting time by 89%.
- (iv) The empirical illustration shows that natural variability is prevented by optimizing the capacity buffers and reducing overtime by 46%.

In practice, there are additional constraints that affect the results, and these require further study.

- (i) Although the duration of surgery is analyzed for 263 types of surgical categories and for 340 surgeons, each hospital is different. For example, some hospitals have a higher proportion of complex surgeries and should make comparisons among institutions.
- (ii) The tests of model accuracy were performed using the year of 2002; they do account for diurnal variation. However, the year variation should be reflected.
- (iii) Additional constraints may arise due to the availability of surgeons or other staff. For example, surgeons may not be available when the case is moved from the morning to the afternoon because they have outpatient clinics or other obligations.

(iv) This study applies to facilities at which the surgeon and patient choose the day and the case is not allowed to be allocated to another day, even if performance may be increased by rescheduling.

(v) Additional constraints may arise due to the availability of the recovery room.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

The work described in this paper was substantially supported by a grant from The Hong Kong Polytechnic University Research Committee under the Joint Supervision Scheme with the Chinese Mainland and Taiwan and Macao Universities 2010/11 (Project no. G-U968). This work was also partially supported by the National Science Council of Taiwan, under Grant NSC-101-2221-E-006-137-MY3.

### References

- [1] L. R. Farnworth, D. E. Lemay, T. Wooldridge et al., "A comparison of operative times in arthroscopic ACL reconstruction between orthopaedic faculty and residents: the financial impact of orthopaedic surgical training in the operating room," *The Iowa Orthopaedic Journal*, vol. 21, pp. 31–35, 2001.
- [2] J. Beliën, E. Demeulemeester, and B. Cardoen, "A decision support system for cyclic master surgery scheduling with multiple objectives," *Journal of Scheduling*, vol. 12, no. 2, pp. 147–161, 2009.
- [3] E. Litvak, M. C. Long, A. B. Copper, and M. L. McManus, "Emergency department diversion: causes and solutions," *Academic Emergency Medicine*, vol. 8, no. 11, pp. 1108–1110, 2001.
- [4] J. D. C. Little, "Little's Law as viewed on its 50th anniversary," *Operations Research*, vol. 59, no. 3, pp. 536–549, 2011.
- [5] W. J. Hopp and M. L. Spearman, *Factory Physics*, McGraw-Hill Education, Boston, Mass, USA, 3rd edition, 2011.
- [6] J. H. May, W. E. Spangler, D. P. Strum, and L. G. Vargas, "The surgical scheduling problem: current research and future opportunities," *Production and Operations Management*, vol. 20, no. 3, pp. 392–405, 2011.
- [7] B. Denton, J. Viapiano, and A. Vogl, "Optimization of surgery sequencing and scheduling decisions under uncertainty," *Health Care Management Science*, vol. 10, no. 1, pp. 13–24, 2007.
- [8] B. Cardoen, E. Demeulemeester, and J. Beliën, "Optimizing a multiple objective surgical case sequencing problem," *International Journal of Production Economics*, vol. 119, no. 2, pp. 354–366, 2009.

- [9] B. T. Denton, A. S. Rahman, H. Nelson, and A. C. Bailey, "Simulation of a multiple operating room surgical suite," in *Proceedings of the Winter Simulation Conference*, pp. 414–424, Monterey, Calif, USA, December 2006.
- [10] M. Lamiri, X. Xie, and A. Dolgui, "A stochastic model for operating room planning with elective and emergency demand for surgery," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1026–1037, 2008.
- [11] J. Goldman, H. A. Knappenberger, and E. W. Moore Jr., "An evaluation of operating room scheduling policies," *Hospital Management*, vol. 107, no. 4, pp. 40–51, 1969.
- [12] E. Marcon, S. Kharraja, and G. Simonnet, "The operating theatre planning by the follow-up of the risk of no realization," *International Journal of Production Economics*, vol. 85, no. 1, pp. 83–90, 2003.
- [13] D. Gupta and B. Denton, "Appointment scheduling in health care: challenges and opportunities," *IIE Transactions*, vol. 40, no. 9, pp. 800–819, 2008.
- [14] Y.-J. Chiang and Y.-C. Ouyang, "Profit optimization in SLA-aware cloud services with a finite capacity queuing model," *Mathematical Problems in Engineering*, vol. 2014, Article ID 534510, 11 pages, 2014.
- [15] M. D. Basson, T. W. Butler, and H. Verma, "Predicting patient nonappearance for surgery as a scheduling strategy to optimize operating room utilization in a Veterans' Administration Hospital," *Anesthesiology*, vol. 104, no. 4, pp. 826–834, 2006.
- [16] M. L. McManus, M. C. Long, A. Cooper et al., "Variability in surgical caseload and access to intensive care services," *Anesthesiology*, vol. 98, no. 6, pp. 1491–1496, 2003.
- [17] E. Litvak, "Optimizing patient flow by managing its variability," in *Front Office to Front Line: Essential Issues for Health Care Leaders*, pp. 91–111, Joint Commission Resources, Oakbrook Terrace, Ill, USA, 2005.
- [18] E. Litvak, P. I. Buerhaus, F. Davidoff, M. C. Long, M. L. McManus, and D. M. Berwick, "Managing unnecessary variability in patient demand to reduce nursing stress and improve patient safety," *Joint Commission Journal on Quality and Patient Safety*, vol. 31, no. 6, pp. 330–338, 2005.
- [19] J. P. Womack, D. T. Jones, and D. Roos, *The Machine that Changed The World*, Free Press, New York, NY, USA, 1990.
- [20] M. Holweg, "The genealogy of lean production," *Journal of Operations Management*, vol. 25, no. 2, pp. 420–437, 2007.
- [21] T. Ohno, *Toyota Production System: Beyond Large-Scale Production*, Productivity Press, New York, NY, USA, 1988.
- [22] Y. Monden, *Toyota Production System: An Integrated Approach to Just-in-Time*, CRS Press, Florida, Fla, USA, 4th edition, 1998.
- [23] J. K. Liker, *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, McGraw-Hill Education, New York, NY, USA, 2004.
- [24] J.-C. Lu, T. Yang, and C.-Y. Wang, "A lean pull system design analysed by value stream mapping and multiple criteria decision-making method under demand uncertainty," *International Journal of Computer Integrated Manufacturing*, vol. 24, no. 3, pp. 211–228, 2011.
- [25] J. Miltenburg, "Level schedules for mixed-model assembly lines in just-in-time production systems," *Management Science*, vol. 35, no. 2, pp. 192–207, 1989.
- [26] N. Boysen, M. Fliedner, and A. Scholl, "The product rate variation problem and its relevance in real world mixed-model assembly lines," *European Journal of Operational Research*, vol. 197, no. 2, pp. 818–824, 2009.
- [27] P. R. McMullen, "The permutation flow shop problem with just in time production considerations," *Production Planning and Control*, vol. 13, no. 3, pp. 307–316, 2002.
- [28] M. A. Millstein and J. S. Martinich, "Takt Time Grouping: implementing kanban-flow manufacturing in an unbalanced, high variation cycle-time process with moving constraints," *International Journal of Production Research*, 2014.
- [29] P. T. Vanberkel and J. T. Blake, "A comprehensive simulation for wait time reduction and capacity planning applied in general surgery," *Health Care Management Science*, vol. 10, no. 4, pp. 373–385, 2007.
- [30] E. Hans, G. Wullink, M. van Houdenhoven, and G. Kazemier, "Robust surgery loading," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1038–1050, 2008.
- [31] Y. Yin, I. Kaku, J. Tang, and J. M. Zhu, *Data Mining: Concepts, Methods and Applications in Management and Engineering Design*, Springer, London, UK, 2011.
- [32] M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, NY, USA, 1973.
- [33] T. Yang and J.-C. Lu, "The use of a multiple attribute decision-making method and value stream mapping in solving the pace-maker location problem," *International Journal of Production Research*, vol. 49, no. 10, pp. 2793–2817, 2011.
- [34] M. Rother and J. Shook, *Learning to See: Value Stream Mapping to Add Value and Eliminate Muda*, Lean Enterprise Institute, Cambridge, Mass, USA, 2003.
- [35] T. Yang, C.-H. Hsieh, and B.-Y. Cheng, "Lean-pull strategy in a re-entrant manufacturing environment: a pilot study for TFT-LCD array manufacturing," *International Journal of Production Research*, vol. 49, no. 6, pp. 1511–1529, 2011.
- [36] J.-C. Lu, T. Yang, and C.-T. Su, "Analysing optimum push/pull junction point location using multiple criteria decision-making for multistage stochastic production system," *International Journal of Production Research*, vol. 50, no. 19, pp. 5523–5537, 2012.
- [37] T. Yang, Y. F. Wen, and F. F. Wang, "Evaluation of robustness of supply chain information-sharing strategies using a hybrid Taguchi and multiple criteria decision-making method," *International Journal of Production Economics*, vol. 134, no. 2, pp. 458–466, 2011.
- [38] R. B. Detty and J. C. Yingling, "Quantifying benefits of conversion to lean manufacturing with discrete event simulation: a case study," *International Journal of Production Research*, vol. 38, no. 2, pp. 429–445, 2000.
- [39] J. Banks, J. S. Carson, B. L. Nelson, and D. M. Nicol, *Discrete-Event System Simulation*, Prentice Hall, New Jersey, NJ, USA, 2000.
- [40] W. D. Kelton, R. P. Sadowski, and N. B. Swets, *Simulation with Arena*, McGraw-Hill Education, Boston, Mass, USA, 5th edition, 2010.
- [41] E. Erdem, X. Qu, and J. Shi, "Rescheduling of elective patients upon the arrival of emergency patients," *Decision Support Systems*, vol. 54, no. 1, pp. 551–563, 2012.
- [42] F. Glover, J. P. Kelly, and M. Laguna, "New advances and applications of combining simulation and optimization," in *Proceedings of the 28th Conference on Winter Simulation*, pp. 144–152, Coronado, Calif, USA, December 1996.
- [43] T. Yang, H.-P. Fu, and K.-Y. Yang, "An evolutionary-simulation approach for the optimization of multi-constant work-in-process strategy—a case study," *International Journal of Production Economics*, vol. 107, no. 1, pp. 104–114, 2007.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

