Scheduling Patients’ Appointments: Allocation of Healthcare Service Using Simulation Optimization

Ping-Shun Chen1*, Rex Aurelius C. Robielos2, Philline Kate Vera C. Palaña2, Pierre Lorenzo L. Valencia2 and Gary Yu-Hsin Chen1

1Department of Industrial and Systems Engineering, Chung Yuan Christian University, Chung Li District, Taoyuan City, 320, Taiwan, ROC
2School of Industrial Engineering and Engineering Management, Mapúa Institute of Technology, Intramuros, Manila 1002, Philippines

Submitted November 2014. Accepted for publication March 2015.

ABSTRACT

In the service industry, scheduling medical procedures causes difficulties for both patients and management. Factors such as fluctuations in customer demand and service time affect the appointment scheduling systems’ performance in terms of, for example, patients’ waiting time, idle time of resources, and total cost/profits. This research implements four appointment scheduling policies, i.e., constant arrival, mixed patient arrival, three-section pattern arrival, and irregular arrival, in an ultrasound department of a hospital in Taiwan. By simulating the four implemented policies’ optimization procedures, optimal or near-optimal solutions can be obtained for patients per arrival, patients’ inter-arrival time, and the number of the time slots for arrived patients. Furthermore, three objective functions are tested, and the results are discussed. The managerial implications and discussions are summarized to demonstrate how outcomes can be useful for hospital managers seeking to allocate their healthcare service capacities.

Keywords: appointment scheduling, healthcare services, scheduling policy, simulation optimization

1. INTRODUCTION

Scheduling appointments can be difficult for service provider industries that must meet the needs of both the customer and the service provider [1]. Customers want to avoid long waits whereas service providers need to minimize the idle time of their resources and the use of overtime [2]. Airline, legal, banking, and healthcare industries offer examples of the need to balance these conflicting interests. The objective of this study
is to develop an improved appointment scheduling system using combined mathematical programming and simulation optimization procedures to assist service providers in making their decisions. This research focuses on the healthcare industry, which has an ongoing need for improvement and is concerned with the high cost of equipment and staff. This study is designed to fulfill the needs of both customers and service providers.

The main challenge of current appointment scheduling approaches for healthcare providers is determining how to schedule the number of patient appointments using special time slots based on fluctuations in patients and stochastic patient treatment time [3]. The appointment scheduling problem can consist of one or more objective functions, such as minimizing patients’ average waiting time, machines’/doctors’ average idle time, overtime, and cost. Patient’s waiting time refers to the time between the patient’s arrival time and the actual service start time of the patient’s service/appointment. Patients’ waiting time is affected by (1) tardiness of earlier patients, causing a delay in the scheduled treatment time, and (2) patients having different treatment times, creating a stochastic environment. For example, if a patient finishes his/her treatment later than the scheduled time, subsequent patients would finish their treatments beyond the originally scheduled time. Consequently, patient must wait longer for their treatment. However, if the patient arrives punctually, condition (1) of patients’ waiting time will not be an issue. In this study, all patients are assumed to be punctual; thus only condition (2) is considered.

Determining the number of patients to assign to a time slot is an important research topic for healthcare providers. Idle time in appointment scheduling [4] refers to the time when staff, equipment, or resources are not being used. If a scheduler assigns too few patients to a time slot, medical resources will stand idle. However, if a scheduler assigns too many patients to a time slot, some patients will be forced to wait. Thus, a trade-off exists between patients’ waiting time and machines’ idle time. In addition, the length of each time slot can be fixed or variable. This study adopts the variable time slot. Therefore, the purpose of this study is to determine the optimal or near-optimal number of arrived patients for each time slot, patients’ inter-arrival time (the length of each time slot), and the number of the time slots for arrived patients based on the stochastic process time of patients, related constraints, and objective functions. The total number of patients served is the sum of the number of arrived patients for each time slot. As this study considers the minimal total patients served for the four implemented appointment scheduling policies, the total number of patients served will vary based on different policies.

The remainder of this article is organized as follows. Section 2 reviews the literature of the appointment scheduling problems and solution procedures. Section 3 introduces the background of the study hospital, the proposed mathematical model, and the simulation optimization procedures. Section 4 describes four appointment scheduling policies and related decision variables and constraints. Section 5 examines the four appointment scheduling policies to determine the optimal or near-optimal solutions based on three objective functions and compares their results; the results and discussions of the implemented appointment scheduling policies are also summarized. Finally, Section 6 concludes the paper and suggests future directions for research.
2. LITERATURE REVIEW
Appointment scheduling problems have received much attention in academic and practical discussions. The trends in appointment scheduling problems can be classified into two categories: integrating more characteristics (e.g., overtime, overbooking, no-shows, and tardy patients) or different objective functions into the appointment scheduling problems, and developing heuristics/metaheuristic algorithms or simulation optimization procedures to solve the appointment scheduling problem. For the former category, researchers extend their research problems to more reality-based situations by integrating more characteristics into the appointment scheduling problem, thereby resulting in more difficulties to solve. Edwards et al. [5] and Rising et al. [6] used performance indicators, such as patients’ waiting time and physicians’ utilization, to evaluate system performance. Their study showed the effects of changes to physicians’ scheduling on patients’ waiting time. Jacobson et al. [7] used the performance indicators of the volume patient throughput, patients’ waiting time, staff overtime, staff utilization, and low physicians’ idle time in order to commit to high quality and efficient patient flow. Klassen and Yoogalingam [8] studied appointment scheduling problems based on patients’ tardiness. They used the simulation optimization algorithm, which contained both scatter search and tabu search, to resolve the problem. They considered the patients’ and physicians’ lateness in their simulation scenarios. The simulation results showed that, if patients are punctual and physicians are late, the plateau-dome policy would be the best policy; on the other hand, if both patients and physicians are tardy, the increasing interval and clustering rule (IICR) would be recommended. Samorani and LaGanga [9] developed a heuristic algorithm to solve the predictive no-show patients and overbooking appointment scheduling problems, which they used to maximize profits as the objective function. In their appointment scheduling problem, they modeled a case clinic with a single server and the constant service time. Based on the quantitative results of a real outpatient clinic, the proposed heuristic algorithm earned only 2.9% less than the traditional column generation method, although it could be used to solve larger problems than the traditional column generation method.

Regarding the algorithms developed in the current literature, Robinson and Chen [4] performed mathematical programming to formulate appointment scheduling with minimal patients’ waiting time and minimal doctors’ idle time. They derived the mathematical eqns. and developed a simple closed-form heuristic to obtain an optimal solution. Using quantitative examples, they found that the proposed heuristic performed on average within 2% of the optimal policy. Saremi et al. [10] constructed a mathematical programming for appointment scheduling of outpatient surgeries in a multistage operating room. In their model, the service time for patients was stochastic and there were multiple types of patients. The researchers developed three simulation optimization methods: simulation-based tabu search, integer programming enhanced tabu search, and binary programming enhanced tabu search (BPETS). The simulation results indicated that the BPETS performed better than the other methods.

Vink et al. [11] considered an appointment scheduling system that sought to minimize a linear combination of patients’ waiting time, machines’ idle time, and the lateness of the schedule. In their model, patients’ service time was assumed to be
stochastic. They developed a lag order approximation method to solve the proposed problem. The lag was the information of the predecessors. Based on a quantitative example of computed tomography (CT) scan patients, they found that if the hospital manager took two predecessors into account, the results would generate a near-optimal schedule within a reasonable solving time by using their proposed algorithm.

Chang et al. [12] published a two-part article describing a new simulation optimization method: the stochastic trust-region response-surface (STRONG) method. This new response-surface framework for simulation optimization combined the non-automated heuristic procedure without a convergence guarantee, called the response surface methodology (RSM) and the trust-region method, to eliminate the need for human intervention and to achieve desired convergence properties. Rau et al. [2] used a simulation model to examine strategic capacity planning of an outpatient physical therapy clinic in Taipei, Taiwan. A discrete event simulation model was constructed to study the dynamics of patient mixes to estimate the capacity of therapy rooms. The researcher performed sensitivity analyses to evaluate the maximum capacity of the therapy rooms.

Weerawat et al. [13] used the discrete-event simulation method to explore the performance of an outpatient clinics in a large Thai public hospital. They examined the detailed patient flow of the orthopedic ward and collected related data. After completing the modelling procedures, they validated the simulation model and performed “what-if” scenarios. Based on the quantitative results, they demonstrated that scheduling different appointments for patients would have a significant effect on the performance indicators, such as doctors’ idle time and overtime as well as patients’ waiting time.

Berg et al. [14] formulated an outpatient appointment scheduling system that incorporated a two-stage stochastic mixed integer approach. The first stage was to schedule the sequence of patients and the inter-arrival time for each patient. The second stage was to minimize patients’ total waiting time and machines’ idle time under each possible scenario. The researchers used three heuristic algorithms, i.e., the L-shaped method, hybrid multicut L-shaped method, and progressive hedging primal heuristic method, and defined the advantages and disadvantages of the three algorithms. They further studied the no-show patient effect and overbooking effect. They summarized useful managerial implications for hospital managers to make appointment scheduling decisions.

Different appointment scheduling policies have been tested on different hospital and facility cases, and the results have been promising. Many policies and strategies have already been implemented in the healthcare sector, and competition among policies has become more interesting with each new study. Although researchers are constantly refining their ideas, deeper analyses must be performed in order to develop more effective appointment scheduling policies.

Ho and Lau [15, 16] compared the performance of simple appointment scheduling policies in a single server system. They applied simulation in their research and modified their appointment scheduling policies by varying the interval time between
patients. Yokouchi et al. [17] focused on the appointment scheduling policies of patients treated in an outpatient chemotherapy department (OCD) at a medium-size general hospital in Kobe, Japan. Their study identified the best appointment scheduling policy based on the characteristics of the medical processes in OCDs using discrete event simulation.

Within outpatient departments’ appointment scheduling systems, many types of systems are modifications and/or combined systems; one example is the single-block appointment scheduling system and individual appointments [18]. Wijewickrama and Takakuwa [19] summarized major appointment scheduling policies or systems for outpatient departments. The appointment scheduling policies consisted of multiple block/fixed-interval policies, variable block/fixed-interval policies, and individual block/variable interval policies; the appointment scheduling system contained the single-block systems, the individual block/fixed-interval systems, and the variable interval system. These concepts were again applied and used with some modifications in their subsequent research studies [20, 21]. For the individual-block/variable-interval policies, Wang [22], Robinson and Chen [4], and Cayirli et al. [23] demonstrated that, for the independent and identically distributed (i.i.d.) service times for homogenous patients, the optimal appointment intervals would be close to a “dome” shape, indicating that the optimal appointment intervals could be classified into three sections: appointment intervals increasing in the first section, remaining constant in the second section, and then decreasing in the third section.

Klassen and Yoogalingam [24] studied how to minimize patients’ waiting time and doctors’ idle time in appointment scheduling systems. They developed a simulation optimization procedure based on the neural network, scatter search, and tabu search. The results of their research showed that the proposed method provided good results across key performance indicators (KPIs) in terms of the number of appointments, appointment length, and the total cost. The patterns of the simulation optimization procedure were close to the plateau-dome patterns.

Shylo et al. [25] presented a simulation model for a surgical suite in Pittsburgh, Pennsylvania, USA, that was used to evaluate and optimize different appointment scheduling policies. They proposed a dynamic appointment scheduling policy as an alternative that allowed for the reduction of the variance in the patients’ waiting time and backlogs.

In summary, researchers commonly conduct simulation-based studies for the healthcare industry due to stochastic processes involved, such as patients’ arrival time and service time. Whether addressing appointment scheduling, capacity planning, or facility and route planning, the simulation modelling approach has become a popular research methodology. Therefore, the current research uses simulation optimization procedures to obtain the optimal or near-optimal solution of the four implemented appointment scheduling policies, i.e., constant arrival, mixed patient arrival, three-section pattern arrival, and irregular arrival, based on the different objective functions in order to determine the ideal level of effectiveness and efficiency of the appointment scheduling systems [13, 15, 16, 19].
3. METHODOLOGY AND CASE STUDY
3.1. Mathematical Model
The appointment scheduling mathematical model was a revision of Robinson and Chen’s model [4], whose original model contained eqns. 1-4:

\[
\text{Minimize } \mathbb{E}\{S_N - \sum_{i=1}^{N-1} \xi_i + \sum_{i=2}^{N} \alpha_i W_i \} \quad (1)
\]

\[
\text{Minimize } \mathbb{E}\{S_N - \sum_{i=1}^{N} \xi_i \} \quad (1a)
\]

Subject to

\[
W_i = \max\{0, S_i - A_i\}, \quad \text{for } i = 1, 2, \ldots, N \quad (2)
\]

\[
S_i = \max\{A_i, S_{i-1} + \xi_{i-1}\}, \quad \text{for } i = 2, 3, \ldots, N,
\]

where \(S_1 = 0\) and \(A_1 = 0\). \quad (3)

\[
X_i = A_{i+1} - A_i \geq 0, \text{ and integer, for } i = 1, 2, \ldots, N-1. \quad (4)
\]

The revised model modified the objective function 1 to 1a. The revised model also added eqns. 5-7:

\[
\sum_{i=1}^{N} W_i / N \leq \text{the patients' average waiting time of the base model policy. } \quad (5)
\]

\[
N \geq \text{the total number of patients served by the base model policy. } \quad (6)
\]

\[
\text{Overtime} = \max\{0, (S_N + \xi_N) - \text{regular working time}\} \leq \text{the maximum allowable overtime. } \quad (7)
\]

The notations are introduced as follows: \(S_i\) is the real start time of the \(i^{th}\) patient’s treatment; \(\xi_i\) is the process/treatment time of the \(i^{th}\) patient; \(\alpha_i\) is the multiplier based on the value of patient’s time or doctor’s time for patient \(i\); \(W_i\) is the waiting time of the \(i^{th}\) patient; \(A_i\) is the appointment/arrival time of the \(i^{th}\) patient; \(X_i\) is the time allotted between patients \(i\) and \(i+1\); and \(N\) is the total number of patients.

Eqn. 1 minimizes the expected value of doctors’ total idle time and patients’ total waiting time. As this research only considered the resource utilization, eqn. (1) is revised to eqn. (1a) and patients’ average waiting time of eqn. (1) is revised to a new constraint eqn. 5. Eqn. (1a) minimizes the expected value of the total length of time during which the procedure rooms were empty. Eqn. (2) represents the waiting time of
patient $i$. Eqn. (3) represents the real start time of the following $i^{th}$ patient subsequent to the previous patient. Eqn. (4) shows the non-negativity and integer constraints for the inter-arrival time of patients. Eqn. (5) represents the patients’ average waiting time when the appointment schedule is less than or equal to the patients’ average waiting time of the base model policy. Eqn. (6) represents the total number of patients when it is greater than or equal to that of all patients served during the planning time period by the base model policy. Eqn. (7) represents overtime when it is less than or equal to the maximum allowable overtime.

To illustrate application of eqns. (2), (3), and (4), consider Figure 1 as an example. $A_i = \{A_1 = 0, A_2 = 8, A_3 = 20, \text{ and } A_4 = 30\}$ is a random variable. $\xi = \{\xi_1 = 10, \xi_2 = 8, \text{ and } \xi_3 = 10\}$ is a random variable. $S_1 = 0$ and $N = 4$. Therefore, the real start time of patients calculated using eqn. (3) is $S_2 = \max\{8, 0 + 10\} = 10$, $S_3 = \max\{20, 10 + 8\} = 20$, and $S_4 = \max\{30, 20 + 10\} = 30$. The total idle time of the procedure room calculated using eqn. (1a) is $S_4 - (\xi_1 + \xi_2 + \xi_3) = 30 - (10 + 8 + 10) = 2$. Patients’ waiting time, as calculated using eqn. (2), is $W_1 = \max\{0, 0 - 0\} = 0$, $W_2 = \max\{0, 10 - 8\} = 2$, $W_3 = \max\{0, 20 - 20\} = 0$, and $W_4 = \max\{0, 30 - 30\} = 0$. The length of time slots calculated using eqn. (4) is $X_1 = 8 - 0 = 8$, $X_2 = 20 - 8 = 12$, and $X_3 = 30 - 20 = 10$. The length of time slots is varied in this example.

### 3.2. Case Study

To develop and test the appointment scheduling policies, this study used the ultrasound department of a hospital for a case study. The study hospital provided eight types of ultrasound services. The hospital allocated five rooms for serving patients with appointments. Each type of ultrasound service was designated for a specific room or rooms, as shown in Table 1.
The ultrasound department was open six days a week and closed on Sundays. Although the facility opened at 8:30 am, the first patient was scheduled for 8:40 am. After 5:00 pm, ultrasound service operators were considered to be working overtime.

Two types of patients are included in the system: appointment patients and walk-in patients. The hospital provides two designated rooms for walk-in patients. In order to simplify appointment scheduling, only patients with appointments were considered in this research.

For the current appointment scheduling policy (i.e., a constant arrival policy), the ultrasound department allotted ten minutes for every patient’s appointment. The description and KPIs of the base model policy are discussed in Section 4.1.

3.2.1. Data Collection Procedure
The study hospital provided nine months of data including the process start and end times, process types of patients, and process times for treatment and imaging. Data of the current appointment scheduling system of the study ultrasound department were provided to the researchers. Data were collected throughout each working day. Outliers were removed from the raw data.

3.3. Simulation Model

3.3.1. Assumptions of the Simulation Model and Data Fitting
In order to create a simulation model, the following assumptions were made:

- Patients were assumed to arrive at the ultrasound department at their scheduled time and took their place in the waiting line.
- As the waiting line was in front of the ultrasound room, the time required for patients to walk from the waiting line to the ultrasound room was negligible. Therefore, no walking time was allotted for patients.
- There was no separate setup time for the ultrasound procedures. The process time in this study consisted of setup time and treatment time.
As maintenance can be scheduled on Sundays or after the regular working hours on weekdays, the ultrasound machines were assumed to be in good working order and not in need of maintenance during the regular working hours or overtime hours.

For the simulation model of the appointment scheduling system, patients were generated according to historical data of the eight ultrasound services. Each patient was scheduled for an available room in an available time slot. On the scheduled day, the patient would undergo his/her procedure. The length of time scheduled for each patient depended on the type of ultrasound procedures.

In order to determine the appropriate distribution of each ultrasound patient type, the chi square and Kolmogorov-Smirnov tests were performed. If the \( p \)-values for both tests were greater than or equal to 0.05, the distribution of the ultrasound patient types were accepted. According to the data-fitting procedures, the accepted process time distributions of the patients of the eight ultrasound services are shown in Table 2.

### 3.3.2. Parameters of the Simulation Model
All scenarios included in this study followed the same replication parameter settings. The selection of the number of replications was based on two criteria. The first criterion was about the half-width value of the utilization of the base model policy, wherein the half-width should fall within an acceptable range of 0.01 \( \mu \) with \( \mu \) being the average utilization of critical medical resources [26]. The second criterion was about the value of \( \sigma/\mu \), the standard deviation divided by the mean [27, 28], which should be lower than 0.02 [26]. The 70 replications would satisfy both criteria.

### 3.3.3. Verification and Validation of the Simulation Model
The verification procedure required checking the logic of the simulation model. To validate the ultrasound department’s simulation model, we examined the mean of the actual and simulated process times of the patients for the eight ultrasound services as well as the number of daily appointment patients. Table 3 indicates that both the simulated and actual means of the patients receiving the eight ultrasound services were not significantly different. Therefore, the simulation model was validated.

**Table 2. Process time distributions of the patients of the eight ultrasound services**

<table>
<thead>
<tr>
<th>Service Type of Ultrasound Patients</th>
<th>Process Time Distribution (Minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal Ultrasound (ABD)</td>
<td>10 + EXPO(3.94)</td>
</tr>
<tr>
<td>Kidney and Bladder Ultrasound (KB)</td>
<td>2 + GAMM(4.19, 1.50)</td>
</tr>
<tr>
<td>Penis, Prostate, and Scrotum Ultrasound (URO)</td>
<td>6 + 12*BETA(1.25, 2.04)</td>
</tr>
<tr>
<td>Lower Abdominal Ultrasound (LABD)</td>
<td>5 + GAMM(5.26, 1.51)</td>
</tr>
<tr>
<td>Breast Ultrasound (Breast)</td>
<td>3 + GAMM(4.33, 1.42)</td>
</tr>
<tr>
<td>Thyroid Ultrasound (Thyroid)</td>
<td>2 + ERLA(3.77, 2.00)</td>
</tr>
<tr>
<td>Duplex Ultrasound (Duplex)</td>
<td>2 + WEIB(10.4, 1.29)</td>
</tr>
<tr>
<td>Muscle Ultrasound (Muscle)</td>
<td>2 + WEIB(7.48, 1.92)</td>
</tr>
</tbody>
</table>
3.4. Simulation Optimization Procedures

This research used simulation optimization procedures (see Figure 2) to obtain an optimal or near-optimal solution of the four implemented appointment scheduling policies of the ultrasound department. The detailed simulation optimization procedures are described as follows:

Step 1. For each iteration, generate initial random values, which are between the lower bound and the upper bound, for three decision variables: the number of patients per arrival, patients’ inter-arrival time, and the number of the time slots for arrived patients. Go to Step 2.

Step 2. Perform 70 replications of the simulation of the current configurations, and record a summary of the simulation results.

Step 3. Check the feasibility of the solution. If the results of the solution satisfy all constraints, the solution will be a feasible solution; then go to Step 4. Otherwise, go to Step 6.

Step 4. If a feasible solution is better than the current best solution, go to Step 5. Otherwise, go to Step 6.

Step 5. Update the current best solution and go to step 6.

Step 6. The two common stopping rules of the simulation optimization procedures are maximum iterations and maximal programming running time. In order to ensure convergence of the simulation optimization procedures, the maximum iterations is used as the stopping rule. Through trial and error, the optimal solution of all four implemented appointment scheduling policies can be obtained with fewer than 35,000 iterations. Therefore, the maximum iteration is set to 35,000 iterations. If the maximum of iterations is reached, go to Step 7. Otherwise, go to Step 8.

Step 7. Obtain the optimal or near-optimal solution. The proposed procedures then stop.

### Table 3. Population Z-test summary

<table>
<thead>
<tr>
<th>Process Type</th>
<th>Average Process Time (min)</th>
<th>Standard Deviation (min)</th>
<th>Z_computed</th>
<th>Z_critical</th>
<th>Significant Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
<td>Simulated</td>
<td>Actual</td>
<td>Simulated</td>
<td></td>
</tr>
<tr>
<td>ABD Machines</td>
<td>13.90</td>
<td>13.84</td>
<td>3.64</td>
<td>0.38</td>
<td>0.97</td>
</tr>
<tr>
<td>Breast Machine</td>
<td>9.14</td>
<td>9.23</td>
<td>4.88</td>
<td>1.17</td>
<td>0.55</td>
</tr>
<tr>
<td>Duplex Machine</td>
<td>11.80</td>
<td>11.54</td>
<td>8.05</td>
<td>5.41</td>
<td>0.37</td>
</tr>
<tr>
<td>KB Machine</td>
<td>8.30</td>
<td>8.46</td>
<td>5.00</td>
<td>2.1</td>
<td>0.57</td>
</tr>
<tr>
<td>LABD Machine</td>
<td>12.90</td>
<td>11.30</td>
<td>6.77</td>
<td>6.96</td>
<td>1.79</td>
</tr>
<tr>
<td>Muscle Machine</td>
<td>8.64</td>
<td>7.86</td>
<td>3.63</td>
<td>3.31</td>
<td>1.84</td>
</tr>
<tr>
<td>Thyroid Machine</td>
<td>9.54</td>
<td>9.82</td>
<td>5.33</td>
<td>3.1</td>
<td>0.68</td>
</tr>
<tr>
<td>URO Machine</td>
<td>10.60</td>
<td>9.85</td>
<td>2.81</td>
<td>4.03</td>
<td>1.49</td>
</tr>
</tbody>
</table>
Step 8. Update the values of the decision variables for the next iteration. The procedures for generating a new solution are based on the scatter search and tabu search. The scatter search uses a weighted linear combination of reference points to generate new values for the decision variables. The tabu search is used to avoid revisiting the previous solutions in order to reduce the search time. Go to Step 2.
4. SCHEDULING POLICY
4.1. Base Model Policy
The base model policy was the representation of the current appointment scheduling system of the ultrasound department studied in this research. This policy, including a constant patient-arrival policy, was the basis of comparison for improving the other implemented policies in Section 4.2. However, the mathematical model of the base model policy consisted of eqns. 1a, 2, 3, and 4. According to the ultrasound department’s schedulers, the parameter settings were that the number of patients per arrival, \( n \), was equal to 5 patients, patients’ inter-arrival time, \( t \), was equal to 10 minutes, and the number of the batch (time slot) for arrived patients, \( B \), was equal to 30. Therefore, \( T = 30 \times 10 = 300 \) minutes, and the daily total number of patients served, \( N = 5 \times 30 = 150 \) patients.

Based on these parameters, the simulation results of this policy showed that the average total idle time of ultrasound rooms was 72.78 minutes, overtime was eliminated, the average room utilization was 95.12 percent, and patients’ average time in system was 50.75 minutes. Patients’ average waiting time was 38.17 minutes, which was the parameter used in eqn. 5. The total number of patients was 150, which was the parameter used in eqn. 6.

4.2. The Implemented Appointment Scheduling Policies
4.2.1. Constant Arrival Policy
The constant arrival policy was run by using simulation optimization procedures in order to obtain an optimal or near-optimal value. In this policy, a constant number of patients arrived at a constant time interval regardless of the service type. This policy was developed because it was easy to implement in hospitals. Compared to the base model policy, the constant arrival policy was used to determine the best values of \( (n, t) \) in order to minimize the average total idle time of each ultrasound room. The mathematical model of the constant arrival policy, mixed patient arrival policy model (M1 and M2), three-section pattern arrival policy, and irregular arrival policy consisted of eqns. 1a, 2 - 7.

The constraints of the constant arrival policy were as follows:
- \( n \) is a positive integer.
- \( t \) was assumed to be a positive integer.
- The time allowable for patient arrivals was limited to \( T \) minutes.
- Overtime was capped at \( O \) minutes, where \( O \) was assigned by managers of the ultrasound department.
- The total number of patients to be served per day should be greater than or equal to \( N \) (150 patients).
- Patients’ average waiting time should be less than or equal to that of the base model policy (i.e., 38.17 minutes).

A maximum of \( n \) patients per arrival was set, depending on the number desired by the manager, in order to allow for the consideration of overbooking patients per slot. The allowable overtime \( O \) was set to 90 minutes in eqn. 7.
4.2.2. Mixed Patient Arrival Policy

The mixed patient arrival policy represented the scenarios that did not have constant patient arrivals at different time intervals, contrary to the constant arrival policy. This study designed two recurring alternative values of \( n \). Two models of the mixed patient arrival policy, M1 and M2, were used to allow the schedule to be more flexible.

The first model, M1, scheduled more patients in the first slot than the second slot. The second model, M2, scheduled more patients in the second slot than the first slot. The decision variables for the M1 and M2 models were the first and second types of patients per arrival, time slots for the first and second types of patients, and maximal arrival of the first and second types of patients. The optimal or near-optimal solutions can be found based on the simulation optimization procedures described in Section 3.4.

4.2.3. Three-Section Pattern Arrival Policy

For the three-section pattern arrival policy, three decision variables were considered: one for the number of patients per arrival, another for the maximum number of arrivals, and the other for the time between patient arrivals (the length of the time slot). Time slots were arranged so that they coincided with the processing times. These could also be divided into three sections: early morning, midday, and late afternoon. Longer processing times were scheduled in longer time slots during the midday, while shorter time slots were used for the shorter processing times and scheduled for early mornings and late afternoons.

The constraints of the three-section pattern arrival policy were similar to those of the constant arrival policy. Each part of the day had its own group of constraints. For example, the time allowed between patient arrivals for the middle part of the day was assumed to be greater than or equal to 15 minutes. For the remaining parts, the time allowed between patient arrivals was assumed to be less than or equal to 15 minutes.

4.2.4. Irregular Arrival Policy

This study developed the irregular arrival policy to determine if it would generate better results than the regular arrival policies, including constant arrival and mixed arrivals. For the irregular arrival policy, every slot was allotted a specific time. The number of patients and the patient’s inter-arrival time were considered to be the decision variables for each slot.

For the ultrasound department, based on the simulation results from the other appointment scheduling policies, an allocation of greater than or equal to 50 slots was enough to hold an \( N \) of greater than or equal to 150 patients. Thus, the number of slots allocated, \( B \), was 50, and each slot has a value of \( n \). The creation time of the next batch of patients depended on the ending time of the previous batch of patients.

To make the policy more flexible, a single step size for the time interval constraints was applied to find an optimal or near-optimal solution for this policy. The constraints of this policy were similar to those of the constant arrival policy, but there were multiple time-between-arrival decision variables rather than one. Moreover, the number of
patients per arrival was constrained only to a range from 1 to \( n \) single-block patients whereas the time between patient arrivals ranged from 0 to \( t \) single-block minutes.

5. DATA ANALYSIS AND DISCUSSIONS

5.1. KPIs of the Implemented Policies

The appointment scheduling policies underwent simulation optimization procedures (Figure 2), which generated optimal or near-optimal results based on the simulation setting described in Section 4. According to the simulation optimization procedures, the best solution of the constant arrival policy was to schedule 12 slots of 35 minutes, during which time 14 patients would be scheduled. For the first model (M1), the best solution of the appointment schedule was to schedule 20 slots, in which 10 and 8 patients are alternately scheduled (see Figure 3); for the second model (M2), 24 slots were scheduled, allowing for 7 and 8 patients to be scheduled alternately. The time between slots alternated between 17 and 18 minutes, as shown in Figure 4.

For the three-section pattern arrival policy, after the simulation optimization procedures were executed, the best solution was that \((n, t, B)\) of early morning, midday, and late afternoon was \((9, 10, 7), (5, 15, 20),\) and \((7, 10, 7)\), respectively (see Figure 5).

Table 4 shows the best solution for the irregular arrival policy appointment schedule. The batch size of patients and time allotted to each patient varied. For slot 1 in Table 4, the number of appointment patients was 5 at \((8:30 + 0:10) = 8:40\); for slot 2, this number was 3 at \((8:40 + 0:03) = 8:43\). The remaining slots followed the same logic. Scheduled patients were assumed to arrive on time and join the waiting line. If one of the five ultrasound rooms became available, the patient would enter the available ultrasound room on a first-come, first-served basis.

The graph of the irregular arrival policy in Figure 6 showed features similar to the three-section pattern arrival policy graph in Figure 5. For example, in the early and late parts of the day, patients were allotted shorter times, while in the middle part of the day, patients were scheduled for longer times.

![Figure 3. Pattern of the M1 model](image-url)
The KPIs of the four implemented appointment scheduling policies are summarized in Table 5, which shows that with regard to the average total idle time of each ultrasound room, the mixed patient arrival policy model (M1 and M2) and the irregular arrival policy perform better, followed by the constant arrival policy, and the three-section pattern arrival policy, according to the paired sample $t$-test at the 95% confidence interval. Thus, the recommend appointment scheduling policies for hospital managers are the M1 and M2 models and the irregular arrival policy. In terms of ultrasound room utilization, a hate rate of utilization (i.e., over 93%) occurred for each policy. If each ultrasound service charges patients the same fee, the three-section pattern arrival policy serves significantly more patients during the 450 minutes than other policies. In other words, it earns more revenues for hospitals during the same service time. However, in terms of patients’ average waiting time, the constant arrival policy and the M2 model perform better than other policies. Therefore, for the different KPIs
Table 4. Appointment schedule for the irregular arrival policy

<table>
<thead>
<tr>
<th>Slot no.</th>
<th>Patient arrival time (min)</th>
<th>No. of patients scheduled</th>
<th>Time between the patient and the next patient (min)</th>
<th>Slot no.</th>
<th>Patient arrival time (min)</th>
<th>No. of patients scheduled</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8:30</td>
<td>0</td>
<td>10</td>
<td>26</td>
<td>13:42</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>8:40</td>
<td>5</td>
<td>3</td>
<td>27</td>
<td>13:50</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>8:43</td>
<td>3</td>
<td>4</td>
<td>28</td>
<td>14:08</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>8:47</td>
<td>5</td>
<td>2</td>
<td>29</td>
<td>14:11</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>8:49</td>
<td>4</td>
<td>1</td>
<td>30</td>
<td>14:23</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>8:50</td>
<td>5</td>
<td>1</td>
<td>31</td>
<td>14:34</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>8:51</td>
<td>4</td>
<td>1</td>
<td>32</td>
<td>14:38</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>8:52</td>
<td>4</td>
<td>1</td>
<td>33</td>
<td>14:49</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>8:53</td>
<td>4</td>
<td>0</td>
<td>34</td>
<td>14:59</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>8:53</td>
<td>3</td>
<td>20</td>
<td>35</td>
<td>15:09</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>9:13</td>
<td>2</td>
<td>16</td>
<td>36</td>
<td>15:22</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>9:29</td>
<td>3</td>
<td>8</td>
<td>37</td>
<td>15:29</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>9:37</td>
<td>3</td>
<td>19</td>
<td>38</td>
<td>15:45</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>9:56</td>
<td>4</td>
<td>15</td>
<td>39</td>
<td>15:49</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>10:11</td>
<td>4</td>
<td>12</td>
<td>40</td>
<td>16:07</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>10:23</td>
<td>3</td>
<td>17</td>
<td>41</td>
<td>16:07</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>10:40</td>
<td>3</td>
<td>20</td>
<td>42</td>
<td>16:07</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>11:00</td>
<td>4</td>
<td>10</td>
<td>43</td>
<td>16:11</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>11:10</td>
<td>5</td>
<td>1</td>
<td>44</td>
<td>16:26</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>11:11</td>
<td>3</td>
<td>2</td>
<td>45</td>
<td>16:31</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>11:13</td>
<td>5</td>
<td>19</td>
<td>46</td>
<td>16:43</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>11:32</td>
<td>3</td>
<td>20</td>
<td>47</td>
<td>16:43</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>11:52</td>
<td>5</td>
<td>5</td>
<td>48</td>
<td>16:43</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>12:57</td>
<td>5</td>
<td>18</td>
<td>49</td>
<td>16:43</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>13:15</td>
<td>4</td>
<td>18</td>
<td>50</td>
<td>16:48</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>13:33</td>
<td>4</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(see Table 5), the corresponding suitable appointment scheduling policies can be recommended. For hospital management, the director of the ultrasound department needs to consider the hospital’s KPIs before making decisions on a suitable appointment scheduling policy for patients.

In order to explore more managerial implications for hospital managers, we conducted two simulation optimization scenarios using the objective functions of minimizing patients’ average waiting time and maximizing the patients served during the 450 minutes. Table 6 shows that, based on patients’ average waiting time and patients’ average time in systems, the ranking order from small to large is the constant
arrival policy, the irregular arrival policy, the three-section pattern arrival policy, the M1 model, and the M2 model (paired sample t-test at the 95% confidence interval). Compared to Table 5, the number of patients served during 450 minutes for each policy

![Irregular arrival policy](image)

**Figure 6.** Pattern of the irregular arrival policy

**Table 5. Summary of results for all appointment scheduling policies for the objective function: Minimizing average total idle time of each ultrasound room**

<table>
<thead>
<tr>
<th>KPIs</th>
<th>Policy</th>
<th>Constant arrival policy</th>
<th>Mixed patient arrival policy models</th>
<th>Three-section pattern arrival policy</th>
<th>Irregular arrival policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average total idle time of each ultrasound room (min)</td>
<td>29.49&lt;sup&gt;b&lt;/sup&gt;</td>
<td>15.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.66&lt;sup&gt;a&lt;/sup&gt;</td>
<td>30.80&lt;sup&gt;b&lt;/sup&gt;</td>
<td>16.61&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Patients served during 450 minutes</td>
<td>166.91</td>
<td>171.17</td>
<td>171.26</td>
<td>184.69</td>
<td>170.49</td>
</tr>
<tr>
<td>Average utilization of ultrasound rooms</td>
<td>93.45%</td>
<td>95.49%</td>
<td>96.62%</td>
<td>93.65%</td>
<td>95.48%</td>
</tr>
<tr>
<td>Patients’ average waiting time (min)</td>
<td>17.15</td>
<td>31.05</td>
<td>21.24</td>
<td>30.75</td>
<td>24.12</td>
</tr>
<tr>
<td>Patients’ average time in system (min)</td>
<td>29.71</td>
<td>43.60</td>
<td>33.79</td>
<td>42.20</td>
<td>36.67</td>
</tr>
<tr>
<td>Number of patients served</td>
<td>168</td>
<td>180</td>
<td>180</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>Overtime (min)</td>
<td>0</td>
<td>23.29</td>
<td>22.61</td>
<td>33.87</td>
<td>71.52</td>
</tr>
<tr>
<td>Number of decision variables</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup> and <sup>b</sup> represent the ranking of the average total idle time of each ultrasound room by the paired sample t-test; <sup>a</sup> > <sup>b</sup>.
in Table 6 decreases by approximately 15 to 20 patients. Furthermore, the utilization of ultrasound rooms for each appointment scheduling policy also decreases by approximately 10 to 15 percent. The reason for these results is that fewer patients result in less waiting time for ultrasound rooms. Due to the constraints of minimum patients served (150 patients), each appointment scheduling policy serves approximately 150 patients during the entire planning period (450 regular working minutes plus 90 overtime minutes), except for the three-section pattern arrival policy.

Table 7 shows that the three-section pattern arrival policy serves the most patients during the 450 minutes, followed by the irregular arrival policy, the M1 model, the constant arrival policy, and the M2 model based on the paired sample $t$-test at the 95% confidence interval.

In summary, in order to minimize the average total idle time of each ultrasound room, the recommended appointment scheduling policies are the mixed patient arrival policy models (M1 and M2) and the irregular arrival policy. In order to
minimize patients’ average waiting time, the recommended appointment scheduling policies are the constant arrival policy, the irregular arrival policy, and the three-section pattern arrival policy. To maximize the patients served during 450 minutes, the recommended appointment scheduling policy is the three-section pattern arrival policy. According to the results of the scenarios for three objective functions, the common recommended appointment scheduling policy is the irregular arrival policy, which is more feasible than the others, while the others can all be viewed as simplified forms of the irregular arrival policy. However, it is too difficult for hospitals to implement this policy due to the 100 decision variables that must be considered. Hence, the irregular arrival policy may not be the best option for hospital managers in practice.

5.2. Sensitivity Analysis

Sensitivity analysis was conducted for the second model (M2) of the mixed patient arrival policy. The M2 model was chosen to be analyzed because its implementation procedures could be easily applied to the ultrasound department. The sensitivity analysis helped the ultrasound department director determine how many patients the department could accommodate each day based on the KPIs.

Three observations were considered for the sensitivity analysis (see Figure 7). First, there was a trade-off between the number of patients served and the average total idle time of ultrasound rooms. The interactions of the four lines in Figure 7 indicated that 180 to 220 patients served by the M2 model had good results among these KPIs. Similarly, Table 5 showed that the number of patients served by the M1 model, the three-section pattern arrival policy, and the irregular arrival policy arrival fall within this special range.

Second, the scenario of allocating fewer than 160 patient appointments daily would have a smaller effect on patients’ average waiting time and on overtime. Serving fewer than 110 patients daily would be undesirable due to the longer average total idle time (i.e., almost 300 minutes) of ultrasound rooms. Finally, exceeding the range of patient appointments (more than 230 patients served daily) would be impractical as there was a limit on overtime (less than or equal to 90 minutes) for the study hospital for this case study.

![Figure 7. Performance varies with the number of patients served.](image-url)
5.3. Limitation of Research
The limitations of this research are as follows. First, this research assumes that all patients arrive at their appointments on time. In practice, approximately 2 to 10 percent of scheduled patients are no-shows, according to the study hospital managers. Second, hospital managers primarily considered one critical KPI in this study: the total expected empty time of the procedure rooms. If each procedure room is assumed to be operated by one doctor and the associated medical staff, the total expected empty time of the procedure rooms can indicate the total expected idle time of physicians. However, different KPIs might have significant effects on the outcome of scheduling patient appointments. Therefore, two extra KPIs—patients’ average waiting time and the patients served during the 450 minutes—were selected as the objective functions to perform the sensitivity analysis. The results were summarized in Section 5.1. Determining which KPIs to consider will be important and challenging decisions for hospital managers.

In light of these limitations, there are two potential avenues for future research. First, appointment scheduling can incorporate new characteristics into the proposed policies, such as walk-ins, late patients, no-shows, early patients, or patients with direct waiting times [3, 26]. Second, multiple objective functions should be applied, for instance, simultaneously considering both the cost of patients’ average waiting time and average total idle time of ultrasound rooms. Based on the two objective functions, the Pareto curve can be found, thereby enabling hospital managers to use the best solution to make decisions.

6. CONCLUSION
This study determined the best solutions of different appointment scheduling policies using mathematical programming and simulation optimization procedures, which could be used by service provider managers for decision-making purposes. From the perspective of hospital management, hospital managers seeking to improve KPIs on the patient appointment scheduling system should focus more on searching for optimal solutions or developing better appointment scheduling policies. To evaluate the difference between the targeted and current values of KPIs of the base model policy (the corresponding status of the study hospital), hospital managers should decide the most desirable KPI for improvement. Once a hospital manager decides which KPI to use as the objective function, the current methodology could help him/her determine how many patients per day his/her facility can accommodate, thereby identifying an appropriate appointment scheduling policy. Hospitals could therefore benefit from the improved KPIs. The present study suggests useful guidelines to hospital managers seeking to implement appointment scheduling policies in the future.

ACKNOWLEDGEMENTS
The authors are grateful to five anonymous referees for their very valuable comments, which have significantly enhanced this work. This research is supported by the Ministry of Science and Technology under contract no. MOST 103-2221-E-033 -057 -.
CONFLICT OF INTEREST
The authors indicated no potential conflicts of interest.

REFERENCES


