MATHEMATICAL MODELING AND COMPUTING, Vol. 9, No. 1, pp. 65–73 (2022)

Solving overbooking appointment scheduling problem under patient no-show condition using heuristics procedure and genetic algorithm

Chua W. Y., Rahmin N. A. A., Nawawi A.

Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

(Received 7 July 2021; Accepted 16 November 2021)

The existence of an efficient appointment schedule is important in the healthcare system since it can minimize patient waiting time, resource idle time, and resource overtime and, hence, optimize the utilization and productivity of healthcare organization. In this research, the overbooking technique is implemented to compensate for patient no-show behavior. The aims of this research are to identify the maximum number of patients that can be assigned to a time slot by examining the effects of multiple assignment and to construct a near-optimal overbooking appointment schedule. Heuristics procedure and genetic algorithm are used in this research. From the results obtained, the number of patients that can be assigned to a time slot is found to be at most three. This information can reduce the conflict which may occur when the patients arrive simultaneously. The results also show that the genetic algorithm has a better performance than the heuristics procedure in solving this problem.

Keywords: *overbooking, no-show, multiple assignment, heuristics procedure, genetic algorithm.*

2010 MSC: 90-50 **DOI:** 10.23939/mmc2022.01.065

1. Introduction

Appointment scheduling system is a vital and prevalent element in health care. A suitable and efficient appointment scheduling system should be designed according to the different conditions of healthcare organizations. In our research, we focus on outpatient appointment scheduling which is more common and significant to most of the people in this world. It is important for an outpatient clinic to have an efficient appointment scheduling system to optimize the utilization and productivity of resources.

Nowadays, the scenario of patient no-shows (scheduled patients who do not arrive for their appointments) becomes a severe problem and should not be neglected since it will bring a multitude of negative impacts to the healthcare system. Patient no-show behaviour had been documented in many previous studies which include the study by Dreiher et al. [4] that reported an overall 30.1% proportion of non-attendance in an obstetrics and gynecology clinic, the study by Green and Savin [5] that reported a 31% no-show rate for MRI screening, and the study by Defife et al. [3] that reported a 21% no-show rate in psychotherapy appointments.

Referring to the previous studies, patient no-show behavior is an important variability that should be taken into account when designing an appointment scheduling system. The negative impacts caused by patient no-shows will not only lead to low utilization and productivity of medical resources but also will limit the access to receiving medical treatments to other patients. This condition will then cause a decrease in the revenue and an increase in the total cost spent by the healthcare organization. An assumption is made in this study that the scheduled patient who showed up for appointments is punctual.

In this study, overbooking is a vital and fundamental technique that is used to compensate for patient no-shows. LaGanga and Lawrence [6] found that overbooking can significantly enhance the performance of the clinic by improving patient access and provider productivity.

In our research, we study a single-server overbooking model with fixed-length slot structures for scheduling the appointment under patient no-show behavior since the only uncertainty we considered is the presence of patients. There are two sessions in each day, the first session is from 0900 hours to 1300 hours whereas the second session is from 1400 hours to 1800 hours. The duration of each slot is fixed at 20 minutes hence there are 12 slots per session.

The main objectives of our research are to construct a near-optimal overbooking appointment schedule under patient no-show behavior and to identify the maximum number of patients that can be assigned to the same time slot by examining the effects of multiple assignments. We study an overbooking model to identify the best overbooking solutions which can minimize a weighted combination of the three performance measures which include patient waiting time, resource overtime and resource idle time. Multiple assignment is a condition where multiple patients are assigned to the same appointment time slot. The effect of multiple assignment is examined in order to observe the impact if the maximum number of patients assigned to the same time slot are limited by the service provider. By identifying and limiting the maximum number of patients to a time slot, the conflict which may occur when the patients arrive simultaneously can be reduced efficiently.

2. Background

The previous researchers, Chen et al. [2] developed two-stage stochastic mixed-integer programming models which consider patient no-show behavior and service duration to find the optimal overbooking solutions in order to minimize the total cost regarding patient waiting time, resource idle time and resource overtime. They also observed the efficiency of different types of time slot structures which include fixed-length slot structure, dome-pattern slot structure and optimal flexible slot structure in scheduling an overbooking appointment. They found that the flexible slot structure which allows the flexibility in appointment start times in a dome-dome-done pattern with alternate long and short time slots outperform among the three types of time slot structures.

The paper written by Chen et al. [2] motivates our research to implement different methods which include heuristics procedure and genetic algorithm to solve the overbooking model with fixed-length slot structures under patient no-show behavior. We consider the model with fixed-length slot structures since the only uncertainty we considered in our research is the no-show behavior of patient. Other than that, this paper also motivates us to observe the effect of multiple assignment in an overbooking model in order to identify the maximum number of patients that can be assigned to the same time slot.

3. Mathematical modelling

3.1. Notation

We refer to the notation from the paper written by Chen et al. [2].

Table 1. Sets.

N	set of patients $\{1,2,\ldots,n\}$ to be scheduled for the session, indexed by i
J	set of time slots in the session, indexed by j
S	set of scenarios, indexed by s
N_s^1	set of patients who show up at the healthcare unit under scenario s
N_s^2	set of patients who do not show up at the healthcare unit under scenario s

The value of n represents the total number of patients scheduled for one session. It is fixed at $n = 14$ in the model since a sensible overbooking level of 2 patients is adopted as the baseline in the paper written by Chen et al. [2]. A finite number of scenarios which is denoted by S, is considered in the model to approximate those stochastic components. Each patient either shows up or does not, according to the no-show rate under each scenario.

Table 2. Parameters.

b_j	beginning time of time slot j
d_{is}	service duration for the i^{th} patient under scenario s
w^{ot}	penalty for each unit of resource overtime
w^{wait}	penalty for each unit of patient waiting time
w^{idle}	penalty for each unit of resource idle time
E	close time of the healthcare facility

Since the overbooking model we considered is with pre-defined time slots, the beginning time b_j of each slot j is fixed. The interval of each time slot is fixed at 20 minutes which is approximately the sum of the average setup time and the average examination time. Hence, $b_1 = 0$, $b_2 = 20$, $b_3 = 40$, and so on.

Table 3. Decision variables.

$x_{ij} = \begin{cases} 1, & \text{if the } i^{th} \text{ patient in the schedule is assigned to the } j^{th} \text{ time slot;} \\ 0, & \text{otherwise.} \end{cases}$	
a_i	appointment time of the i^{th} patient in the schedule
z_{is}^{start}	start time of the medical service provided to the i^{th} patient under scenario s
z_{is}^{end}	end time of the medical service provided to the i^{th} patient under scenario s
wait_{is}	waiting time of the i^{th} patient for receiving the medical service under scenario s
idle_{is}	idle time of the resource between the $(i)^{th}$ and the $(i+1)^{th}$ services under scenario s
ot_s	overtime of the resource under scenario s

Table 4. New decision variables.

x_j	number of patients assigned to the j^{th} time slot
m	maximum number of patients which can be assigned in one time slot

The decision variables in Table 4 are added by us to be used in examining the effect of multiple assignment.

3.2. Overbooking model

We refer to the overbooking model with pre-defined time slots from the paper written by Chen et al. [2]. Constraints (11), (12) and (13) are designed by us in order to examine the effect of multiple assignment.

The goals of our research are to determine the maximum number of patients which can be assigned in a time slot and to identify a near-optimal overbooking appointment schedule under patient no-show behavior.

The objective function of this model to minimize a weighted sum of the three performance measures which include resource overtime(OT), resource idle time(IT) and patient waiting time(WT) among all the scenarios is:

$$\min w^{ot} \sum_{s \in S} ot_s + w^{\text{idle}} \sum_{i \in N, s \in S} \text{idle}_{is} + w^{\text{wait}} \sum_{i \in N_s^1, s \in S} \frac{\text{wait}_{is}}{N_s^1}$$

subject to the constraints

$$a_i = \sum_{j \in J} b_j x_{ij} \quad \forall i \in N, \tag{1}$$

$$\sum_{j \in J} x_{ij} = 1 \quad \forall i \in N, \tag{2}$$

$$a_{i+1} \geqslant a_i \quad \forall i \in N \setminus \{n\}, \tag{3}$$

$$z_{1s}^{\text{start}} = 0 \quad \forall s \in S, \tag{4}$$

$$z_{is}^{\text{start}} = a_i + \text{wait}_{is} \quad \forall i \in N, s \in S, \tag{5}$$

$$z_{is}^{\text{start}} = z_{(i-1)s}^{\text{end}} + \text{idle}_{(i-1)s} \quad \forall i \in N \setminus \{1\}, s \in S, \tag{6}$$

$$z_{is}^{\text{start}} + d_{is} = z_{is}^{\text{end}} \quad \forall i \in N, s \in S, \tag{7}$$

$$z_{ns}^{\text{end}} + \text{idle}_{ns} - \text{ot}_s = E \quad \forall s \in S, \tag{8}$$

$$z_{is}^{\text{start}}, z_{is}^{\text{end}}, \text{wait}_{is}, \text{idle}_{is}, \text{ot}_s \geqslant 0 \quad \forall i \in N, s \in S, \tag{9}$$

$$x_{ij} \in \{0,1\} \quad \forall i \in N, j \in J, \tag{10}$$

$$x_j \leqslant m \quad \forall j \in J, \tag{11}$$

$$2 \leqslant m \leqslant 4, \quad m \text{ integer}, \tag{12}$$

$$\sum_{j \in J} x_j = n. \tag{13}$$

Constraint (1) acts to assign the appointment time for the i^{th} patient who is assigned to the j^{th} time slot in the schedule. Due to the assumption that the patients are punctual, the arrival time of each patient equals to their appointment time. Constraint (2) states that each patient can only be assigned to exactly one time slot. Constraint (3) ensures the correct sequence of the scheduled appointments. Constraint (4) initializes the start time of the medical service provided to the first patient as 0.

Constraint (5) and (7) are used to calculate the patient waiting time, service start time and service end time for each patient. Constraint (6) is used to compute the resource idle time between the $(i-1)^{th}$ and the i^{th} services. Constraint (8) is used to calculate the resource overtime. Constraint (9) enforces the non-negativity condition of the variables. Constraint (10) shows that the assignment of patients is binary.

Constraint (11) ensures that the number of patients assigned to the j^{th} time slot will not exceed the maximum number of patients which can be assigned in one time slot. Constraint (12) shows that the maximum number of patients can be set at $m = 2$, 3 or 4 before scheduling the appointment. The value of m is set in the interval $[2, 4]$ since 2 is the minimum value of patients to be assigned in one slot in an overbooking model and it is not reasonable to assign more than 4 patients in one slot. Constraint (13) ensures that the total number of patients assigned to all time slots equal to the total number of patients scheduled for one session, which means that $x_1 + x_2 + \cdots + x_{12} = n$.

3.3. Patient no-show model

We refer to the paper written by Chen et al. [2] for the procedure of modelling patient no-shows.
Notation: $\alpha =$ no-show probability of patient.
Procedure:

1. $\forall i \in N$ and $s \in S$, generate a random number U_{is} which is uniformly distributed in the interval $(0,1)$.
2. $N_s^1 = \{i \in N \colon U_{is} > \alpha\}$ and $N_s^2 = \{i \in N \colon U_{is} \leqslant \alpha\}$, i.e., (N_s^1, N_s^2) is a partition of N, where N_s^1 and N_s^2 are the sets of patients who show up for the appointment and does not, respectively, under scenario s.
3. For $s \in S$ and $i \in N_s^1$, $\{d_{is} \colon i \in N, s \in S\}$ are generated according to their empirical probability distributions. For $s \in S$ and $i \in N_s^2$, $d_{is} = 0$.

Taken the mean value of no-show probability from the empirical data in the paper written by Chen et al. [2], we obtain the value $\alpha = 0.176$. Therefore, we can conclude that if $U_{is} > 0.176$, the patient shows up for appointment whereas if $U_{is} \leqslant 0.176$, the patient does not show up for appointment.

4. Methodology

4.1. Data generation

For those patients who show up, their service durations which include the setup time and the examination time are generated according to their empirical probability distributions as shown in Table 5. For those patients who do not show up, their service duration is equal to zero.

Table 5. Summary statistics from empirical data (Adapted from Chen et al. [2]).

Random Variable	**Mean(μ)**	**Standard Deviation(σ)**	**Distribution**
Examination time per test (min)	12.70	8.09	0.5+87*BETA(2.3,12.7)
Setup time for each test (min)	6.40	5.17	-0.5+LOGN(7.01,6.43)

By referring to Chen et al. [1], the probability distributions as shown in Table 5 were estimated by them through the collection of data related to the appointments in a medical imaging center from January 2015 to March 2015.

In our research, we generate 3 different sets of data which include $D = 1$, $D = 20$ and $D = 100$ where D represents the total number of dataset. We solve the overbooking appointment scheduling problem with heuristics procedure and genetic algorithm by using these 3 different sets of data.

4.2. Heuristics procedure

Heuristics procedure is used to get the first initial solution by leaving the last two slots blank to prevent overtime and by assigning m patient to the first time slot to prevent idle time. The other slots are assigned to a random number of patients with the condition that 14 patients should be assigned for one session.

After getting the first initial solution by using heuristics procedure, the other initial solutions are generated by using local search algorithm. By holding the concepts of heuristic procedure which are assigning m patient to the first time slot and leaving the last two slots blank, local search algorithm is applied to the slots except for the first and the last two slots by exploring the neighborhood of the solution. A number of initial solutions is needed because genetic algorithm which is going to be used in obtaining the near-optimal solutions works on a set of solutions.

Heuristics procedure is used to find one best solution for each value of m and identify the maximum number of patients that can be assigned to the same time slot. This is to decide which sets of solutions to be used as the initial solutions to start the process of genetic algorithm.

4.3. Genetic algorithm

Genetic algorithm is used in obtaining the current solutions for the overbooking appointment scheduling problems to achieve a near-optimal overbooking solutions at the end. Five phases are considered in genetic algorithm which include initial population, fitness function, selection, crossover and mutation.

The solutions of heuristics procedure are taken as the initial population to start the process of genetic algorithm. The solutions are encoded by using integer representation by showing the number of slot assigned to each patient. A small population size will cause a high repetition rate in the selection process in large iteration. Therefore, to increase the population size, the initial solutions of heuristics procedure undergo the process of fitness evaluation, selection, crossover and mutation to produce more solutions to be added to the initial population.

Since the overbooking appointment scheduling problem is a minimization problem, the fitness value is evaluated by the formula:

$$\text{fitness value} = \frac{1}{(\text{objective value})}.$$

The higher the fitness value of the chromosome, the higher the chances of that chromosome for being selected for reproduction. Tournament selection is used to randomly select two parents from the current population.

Single point crossover operator is then implemented to produce two offspring from the pair of parents selected. A crossover point is selected at random in the chromosome and the genes before and after the selected point are exchanged among the two parents. After crossover operator, the two offspring produced will go through the mutation operator. Mutation occurs by resetting the value of gene randomly. A random value from the set of permissible values is chosen to mutate a randomly selected gene.

After crossover and mutation process, replacement happens by considering the fitness values of all parents and offspring in order to form a new generation. Since the population size is constant, every time a new offspring is produced, an old individual with the lowest fitness value is eliminated to provide space to the new offspring.

The process of fitness evaluation, selection, crossover, mutation and replacement are repeated for the new generation in the next iteration until the stopping criteria is satisfied. The stopping criteria of our research is a user-defined number of iteration.

5. Results and discussion

The results obtained from heuristics procedure and genetic algorithm for dataset $D = 1$, $D = 20$ and $D = 100$ are shown in Tables 6–9. The minimum value which corresponds to the best performance of each measure among heuristics procedure and genetic algorithm is highlighted.

5.1. Dataset $D = 1$

Table 6. Comparison table for dataset $D = 1$.

	Heuristics Procedure	Genetic Algorithm
Objective Value	21.9793	21.3960
Total Waiting Time	352.92	252.92
Total Idle Time	0.00	0.00
Overtime	31.62	31.62

5.2. Dataset $D = 20$

Table 7. Comparison table for dataset $D = 20$.

		Heuristics Procedure	Genetic Algorithm
Objective Value		24.3709	24.3709
Waiting Time	Mean	41.7450	41.7450
	Standard deviation	35.1200	35.1200
	Minimum	0.00	0.00
	Maximum	124.22	124.22
Idle Time	Mean	0.2574	0.2574
	Standard deviation	2.8475	2.8475
	Minimum	0.00	0.00
	Maximum	37.51	37.51
Overtime	Mean	32.3703	32.3703
	Standard deviation	26.1985	26.1985
	Minimum	0.00	0.00
	Maximum	85.12	85.12

5.3. Dataset $D = 100$

Table 8. Comparison table for dataset $D = 100$.

		Heuristics Procedure	Genetic Algorithm
Objective Value		25.6774	25.6421
Waiting Time	Mean	47.2904	44.6333
	Standard deviation	38.9688	38.0903
	Minimum	0.00	0.00
	Maximum	164.50	164.50
Idle Time	Mean	0.3830	0.3944
	Standard deviation	2.9507	2.8956
	Minimum	0.00	0.00
	Maximum	44.98	44.98
Overtime	Mean	33.0829	33.2418
	Standard deviation	30.7946	30.8490
	Minimum	0.00	0.00
	Maximum	134.99	134.99

5.3. Solutions for each dataset

Table 9. Solutions of heuristics procedure and genetic algorithm for dataset $D = 1$, $D = 20$ and $D = 100$.

$D = 1$	Heuristics Procedure												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	3	1	1	1	1	1	2	1	1	2	0	0
	Genetic Algorithm												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	3	1	1	1	0	1	1	1	2	3	0	0
$D = 20$	Heuristics Procedure												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	3	1	1	2	1	1	1	1	1	2	0	0
	Genetic Algorithm												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	3	1	1	2	1	1	1	1	1	2	0	0
$D = 100$	Heuristics Procedure												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	4	1	1	1	1	1	1	1	1	2	0	0
	Genetic Algorithm												
	SLOT	1	2	3	4	5	6	7	8	9	10	11	12
	NO. OF PATIENT	3	1	2	1	1	1	1	1	1	2	0	0

5.4. Discussion

Chen et al. [2] used the sample average method to solve the overbooking appointment scheduling problem. In this study, since the overbooking model in Chen et al. [2] had been modified by us, we only compare the results obtained among heuristics procedure and genetic algorithm.

Table 6 and Table 8 show that genetic algorithm improves the results of heuristics procedure. Genetic algorithm achieves a lower objective value than heuristics procedure which represents the lower cost is spent by the healthcare organization. On the other hand, Table 7 shows that genetic algorithm maintains the results of heuristics procedure. They have the same objective value which means that the same cost is spent by the healthcare organization. In general, we can conclude that genetic algorithm improves or at least maintains the results of heuristics procedure.

From part of heuristics procedure in Table 9, we see that the solutions of $m = 3$, $m = 3$ and $m = 4$ are used as the initial solutions to start the process of genetic algorithm for dataset $D = 1$, $D = 20$ and $D = 100$ respectively. However, from part of genetic algorithm, we observe that the near-optimal solution obtained for all dataset has the same maximum number of patients that can be assigned to each time slot, which is three. Therefore, in general, the multiple assignment condition of $m = 3$ achieves the best performance in solving overbooking appointment scheduling problem.

From other perspectives, we can see that the near-optimal overbooking solution for all dataset has a specific pattern. Three patients are assigned to the first slot to prevent idle time and no patient is assigned to the last two slots to prevent overtime. Moreover, compare to the other slots, more patients are assigned to the tenth slot to reduce patient waiting time. The other slots are mostly assigned to only one patient and sometimes are assigned to two patients.

6. Conclusion

This research focused on solving the overbooking appointment scheduling problem under the patient no-show condition in an outpatient department. Since the only uncertainty we considered is the no-show behavior of patients, an assumption is made which is the patients who show up will be punctual.

One of the objectives of our research is to identify the near-optimal overbooking appointment under patient no-show behavior in order to minimize the resource overtime, resource idle time and patient waiting time which can hence minimize the cost spent by the healthcare organization. Besides, the other objective of our research is to identify the maximum number of patients that can be assigned to the same time slot in order to reduce the conflict which may occur when the patients arrive simultaneously.

We solved the overbooking appointment scheduling problem by using heuristics procedure and genetic algorithm. From the results obtained, we found that the multiple assignment condition of $m = 3$ achieves the best performance in minimizing the cost of the healthcare organization. Furthermore, for the near-optimal overbooking solution, we obtained an overbooking appointment schedule with the pattern of three patients are assigned to the first slot, no patient is assigned to the last two slots, more patients are assigned to the tenth slot and the other slots are mostly assigned to only one patient and sometimes are assigned to two patients.

Lastly, we can conclude that the genetic algorithm has a better performance in solving overbooking appointment scheduling problem compared to the heuristics procedure. Genetic algorithm will improve or at least maintain the best solution obtained from the heuristics procedure.

For future research, other metaheuristics methods such as simulated annealing and tabu search can be used to solve the overbooking appointment scheduling problem and consequently compare them to observe which method will produce a solution with a higher optimal level. Other than that, since the only uncertainty we considered is the no-show behavior of patients, other sources of variability such as patient punctuality and lateness of physicians when solving overbooking appointment scheduling problems can also be considered. Last but not least, other branches of mathematics such as graph theory are also an active field in proposing the optimal solution of the scheduling problem. There are a vast number of research papers applying the concept of graph coloring to solve the scheduling problems such as doctors' or nurses' duty roster, class or examination schedule, etc.

Acknowledgement

The research is supported by Ministry of Education under Fundamental Research Grant Scheme (01-01-20-2306FR).

[1] Chen Y., Kuo Y. H., Balasubramanian H., Wen C. Using simulation to examine appointment overbooking schemes for a medical imaging center. Winter Simulation Conference (WSC) (2015).

[2] Chen Y., Kuo Y. H., Fan P., Balasubramanian H. Appointment overbooking with different time slot structures. Computers & Industrial Engineering. **124**, 237–248 (2018).

[3] Defife J. A., Conklin C. Z., Smith J. M., Poole J. Psychotherapy appointment no-shows: Rates and reasons. Psychotherapy: Theory, Research, Practice, Training. **47** (3), 413–417 (2010).

[4] Dreiher J., Froimovici M., Bibi Y., Vardy D. A., Cicurel A., Cohen A. D. Nonattendance in obstetrics and gynecology patients. Gynecologic and Obstetric Investigation. **66** (1), 40–43 (2008).

[5] Green L. V., Savin S. Reducing delays for medical appointments: A queueing approach. Operations Research. **56** (6), 1526–1538 (2008).

[6] LaGanga L. R., Lawrence S. R. Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences. **38** (2), 251–276 (2007).

Вирішення проблеми планування прийому з надмірним бронюванням за умови неявки пацієнта за допомогою евристичної процедури та генетичного алгоритму

Чуа В. Ю., Рахмін Н. А., Нававі А.

Кафедра математики та статистики,
Факультет природничих наук,
Університет Путра Малайзія,
43400 UPM Серданг, Селангор, Малайзія

Ефективний графік прийому пацієнтів є важливим для системи охорони здоров'я, оскільки він може мінімізувати час очікування пацієнтів, час простою і понаднормовий час ресурсів, а отже, оптимізувати використання та продуктивність організації охорони здоров'я. У цьому дослідженні застосовано техніку надмірного бронювання (овербукінг), щоб компенсувати неявку пацієнтів. Цілі цього дослідження полягають у визначенні максимальної кількості пацієнтів, які можуть бути призначені на певний інтервал часу, шляхом вивчення ефектів множинного призначення та побудови майже оптимального графіка прийомів із надмірним бронюванням. У цьому дослідженні використовуються евристична процедура та генетичний алгоритм. З отриманих результатів виявляється, що кількість пацієнтів, які можуть бути віднесені до одного інтервалу часу, становить не більше трьох. Ця інформація може послабити конфлікт, який може виникнути, коли пацієнти прибувають одночасно. Результати також показують, що генетичний алгоритм має кращу продуктивність, ніж евристична процедура у вирішенні цієї проблеми.

Ключові слова: *овербукінг, неявка, множинне призначення, евристична процедура, генетичний алгоритм.*