AMSI VACATION RESEARCH SCHOLARSHIPS 2018-2019

Including patient preference in outpatient appointment scheduling

Jamie Owen

Supervised by Belinda Spratt

Queensland University of Technology

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Australian Government
Department of Education and Training

Abstract

Hospital outpatient services are a key component in delivering healthcare to patients in the wider community. Efficient appointment scheduling is becoming necessary as the community grows whilst resources are fixed. In this paper, we develop a Mixed Integer Programming model for the outpatient appointment scheduling problem. The distinguishing feature of the model is the incorporation of patient preference, to improve satisfaction and reduce no-shows. Results indicate that the Australian government's Treat-In-Turn policy improves schedule fairness without reducing the number of patients who are treated in a preferred time block.

1. Introduction

Outpatient services are an integral part of Australian public health care. An outpatient appointment is an appointment that doesn't require the patient to be admitted to the hospital. Outpatient care is becoming a greater part of patient treatment, however there are finitely many resources available to service growing communities. This makes the efficiency of outpatient appointments of paramount importance.

As this demand grows, patient no-shows have a larger impact on hospital efficiency including patient waiting times. To help combat the rise in patient no-shows a more patient-centric scheduling system must be explored. One way in which satisfaction can be increased is by taking more consideration of the appointment blocks patients desire most rather than blocks they can possibly attend. This can be done by discriminating between patient availability and patient preference when formulating the model, and weighting the number of appointments patients prefer. An issue with this approach however is that if a patient desires or is available for a small number of time blocks they could be not scheduled over a more flexible patient.

To incorporate schedule fairness, the Australian government has devised a system called Treat-in-Turn. The Treat-in-Turn policy ensures that 60% of patients get seen in the order they submit a request for an appointment.

This work is innovative in a number of ways. Firstly, it helps separate patient availability and patient preference for specific blocks. Secondly it uses the Treat-In-Turn policy to help ensure fairness. In doing so, this model is able to balance fairness and patient preference, which has not been seen before.

The layout of this paper is as follows. A literature review of current research in patient preference in outpatient scheduling is provided in Section 2. This is followed by the model formulation (Section 3) and the solution techniques (Section 4). The results of the model are discussed in Section 5. Concluding remarks are made in Section 6. The mathematical model and analysis are the author's own work.

2. Literature Review

Outpatient scheduling plays an important role in serving the community better. It does this by both reducing costs and improving waiting time. Because of the important role it plays there has been much research done to help improve the field. Whilst outpatient scheduling has been addressed since the 1950s, popularity of the topic has increased rapidly in the more recent years (Ahmadi-Javid, Jalali, & Klassen, 2017).

Outpatient clinics are becoming more ingrained in the medical healthcare network due to shorter hospital stays and an emphasis on preventative medical care , necessitating improved outpatient scheduling techniques (Cayirli & Veral, 2003). With outpatient clinics being at the heart of the healthcare network, more emphasis must be placed on the efficiency of the appointment systems used. Better efficiency is of great importance as it is a way of reducing costs and increasing performance with existing resources.

There are many factors affecting the scheduling problem, such as the many stakeholders that healthcare affects and their differing priorities. The first set of stakeholders is the patients themselves. Their priorities include short waiting times, obtaining preferred appointment times, receiving the best care available and minimising costs. A second set of stakeholders is the staff whose priorities include spreading the workload evenly and limiting unexpected overtime. The third set of stakeholders is the upper management of the clinic or hospital whose main concerns are maximising profit and meeting key performance indicators set by regulatory bodies.

Another factor in outpatient scheduling is the large number of individual decisions to make when constructing the model. This makes the available literature highly diverse, bringing a large wealth of knowledge to draw from. The large number of decisions also means that it is hard to provide a general scheduling scheme for outpatient scheduling.

One gap in the literature is the area of patient preference. Few authors have considered patient preference whilst solving the outpatient appointment scheduling problem (Ahmadi-

Javid et al., 2017). Of particular note are the papers by Wang and Fung (2014) and Zhu, Hou, Wang, and Zhou (2012), this is because they are the only papers to use patient preference directly in their objective functions.

Wang and Fung (2014) consider the scarcity of medical resources in China, and construct two integer models: one to maximize patient satisfaction and another to maximize hospital revenue. They then developed an integrated model based of the two previous models to help analyze the payoff. They note that in the future there model could be improved with the addition of scheduling more than one work day at a time, and that there model in its current form lacks the certainty that it would be optimal over the long run as it has no way of checking if a person would be happier waiting a day later and getting an appointment block they want rather than getting a worse time block earlier. Lastly the model does not allow the hospital to overlap booking to help mitigate no-shows.

Zhu et al. (2012) considered both MRI scanning and CT scanning and developed a model to help schedule patients based of both their priority and there availabilities. Unusually, the patients are required to bid for their time slot.

There is a gap in the literature, as while both papers investigate the use of patient preference in scheduling they do not separate patient availability from patient preference. As the availability of a patient and their preference are fundamentally different, this area requires further study. We address that gap with this research.

3. Model

The model maximizes the number of blocks filled by patients that prefer that block, whilst ensuring fairness by employing the Treat-in-Turn policy. This is a new strategy as using blocks patient's prefer in conjunction with blocks they are available for has not been published at this time. It is important as patient no-shows have a large destabilizing effect on hospital throughput and by taking into account patient preference it is hoped that no-shows will be reduced.

3.1. Scalar Parameters

$\bar{B}$: the number of blocks in the planning horizon

$\bar{P}$: the number of patients to schedule

ϵ: the proportion of patients that must be treated in turn

M: the max possible days waiting

ω: the weighting associated with providing patients with preferred times

3.2. Index Sets

B: the set of blocks in the planning horizon. $B = \{1, \dots, \bar{B}\}$

P: the set of patients in the in planning horizon. P= $\{1, \dots, \bar{P}\}$

3.3. Indices

b: index for blocks in set B.

p: index for patients in set P.

3.4. Vector Parameters

U_{pb}: 1 if patient is available but does not prefer this block 0 otherwise, $\forall p \in P, \ \forall b \in B$.

V_{pb}: 1 if patient prefers this timeslot, 0 otherwise, $\forall p \in P, \ \forall b \in B$.

W_p: Number of days patient has been waiting in queue, $\forall p \in P$.

N_b: Number of doctors available during block , $\forall b \in B$.

3.5. Decision Variables

X_{pb}: 1 if patient p is scheduled in block b, 0 otherwise, $\forall p \in P, \ \forall b \in B$

θ_p: 1 if patient is in longest waiting cohort, $\forall p \in P$

η_p: 1 if patient p is treated and in the longest waiting cohort, $\forall p \in P$

E: an upper bound on the amount of time a patient not in the longest waiting cohort has been waiting.

F: a lower bound on the amount of time a patient in the longest waiting cohort has been waiting.

3.6. Objective Function

The objective function maximizes the total number of patients scheduled in their preferred timeslot.

$$\sum_p \sum_b (U_{pb} + \omega V_{pb}) X_{pb} \tag{1}$$

3.7. Constraints

Equation (2) calculates the length of the longest waiting cohort.

$$\sum_p \theta_p = \sum_p \sum_b X_{pb} \tag{2}$$

Equation (3) determines the longest a patient not in the longest waiting cohort has been waiting.

$$E \geq (1-\theta_p)W_p, \qquad \forall p \in P \tag{3}$$

Equation (4) determines the shortest amount of time that a patient in the longest waiting cohort has been waiting.

$$F \leq \theta_p W_p + (1-\theta_p)M, \qquad \forall p \in P \tag{4}$$

Equation (5) ensures that the longest waiting cohort only contains patients that have waited the longest

$$E \leq F \tag{5}$$

Equation (6) determines which patients are both in the longest waiting cohort and have been scheduled.

$$\eta_p \leq \frac{\theta_p + \sum_b X_{pb}}{2}, \forall p \in P \tag{6}$$

Equation (7) ensures that the proportion of patients needed to be treated in turn has been satisfied.

$$\sum_p \eta_p \geq \epsilon \sum_p \theta_p \tag{7}$$

Equation (8) only allows a patient to be scheduled if that block is available or preferred.

$$X_{pb} \leq V_{pb} + U_{pb}, \qquad \forall p \in P, b \in B \tag{8}$$

Equation (9) allows patients to be scheduled at most once.

$$\sum_b X_{pb} \leq 1, \qquad \forall p \in P \tag{9}$$

Equation (10) limits the number of patients scheduled during a time block to not exceed the number of doctors available at that time block.

$$\sum_p X_{pb} \leq N_b, \qquad \forall b \in B \tag{10}$$

3.8. Model Assumptions

There are a number of assumptions made with the formulation of this model. One assumption is that appointment requests are put in a queue before being allocated. Another assumption is that appointment systems that comply with the Treat-in-Turn policy which guarantees a measure of fairness. Further assumptions were made about the patient physician interactions. Patients have no physician preference, all physicians can treat all patients, and all appointments are the same duration.

4. Solution Techniques

4.1. Data and Parameters

There is a lot of data that needs to be generated for the model to be run, and because of the novelty of the model much of this data is not readily available and must be estimated.

The number of blocks and patients in the model are important in the context of the ratio between them this being a major factor in how fully utilized the booking process will be. The absolute size of the blocks and patients also needs to be carefully chosen as too few blocks or patients will decrease the chance that a patient prefers or is available for any time block in the period, whereas too many blocks or patients will increase the complexity of the model and make it take longer to solve.

The proportion of patients that must be treated in turn has been set by the Australian government as 60% however the fraction of patients treated in turn is largely arbitrary and has been decided upon while more naïve systems of scheduling were commonly in use. This has a large impact as treat in turn is used as a solution for lack of patient throughput however it does this by reducing total patient throughput, so for better scheduling systems a more appropriate proportion could be justified.

Patient availability and preference were calculated using Bernoulli trials with a 20% chance for availability and 5% for preference per block. As no data is available for either of these two parameters an arbitrary distribution was used for the generation of the data. The accuracy of this model would be increased with better estimation of these parameters.

The doctor availability was generated using a binomial distribution with $p = 0.8$ and $n = 5$, however the p and n used can be very different depending on specific hospitals, although it will still follow this distribution. This corresponds to a system where there are five doctors in total, and each doctor has a probability of 80% of being available during a given time block.

The number of days a patient has been waiting in the queue were calculated using historically based waiting list distributions (Spratt, 2018). Here we considered the distribution of category one patients with a maximum waiting time of 30 days to approximate an outpatient scheduling system.

The patient preference weighting, ω, is any value of $\omega > \bar{P}$ such that both ω and ω^{-1} are above machine precision level. This is because even if there is a difference of the maximum number of patients between two solutions, the solution with more patients in preferred blocks has a greater value. This results in a schedule that maximizes the number of patients in their preferred time block and then maximizes the total number of patients.

4.2. CPLEX

CPLEX is a commercially available optimization software package, which is able to solve linear programming models and select nonlinear programming models. As CPLEX is better suited to solving linear constraints, as opposed to nonlinear constraints, equations (2) to (6) are used. A single non-linear constraint with fewer decision variables could be implemented in a MINLP. The CPLEX model formulation is provided in the appendix.

5. Results and Discussion

A number of tests were run on the model to find how the model reacted to different inputs. The computational experiments discussed in Section 5 are performed on a i7-4770K @ 3.50GHz with 16 GB of ram. Each computational experiment was run five times and the solution time was averaged.

The first computational experiment investigates how reducing the amount of blocks changes the total number of patients scheduled and the number of patients scheduled at their preferred time. The minimum proportion of patients to be treated in turn, ϵ, was set to 0.6 as this is in line with real-life implementation. The other parameters and patient data used were as discussed in the data section.

Table 1 shows the number of blocks used in the schedule, the number of patients in their preferred time blocks, the total number of patients scheduled and the proportion of scheduled patients in their preferred time blocks. The table also shows the number of ticks required to solve the problem using CPLEX and the average time in seconds.

As can be seen in

Table 1, the number of blocks available did not have a major impact on the effectiveness of the model until 20 or less blocks where used. Part of this sharp decline in effectiveness can be explained by the distributions involved in generating the patient and doctor data. As there were 20 blocks available and an expected number of four doctors per block, there were only 80 effective blocks available to service an expected 64 patients who have a preferred block in the time period.

Table 1 also indicates that, in the randomly generated instance, 40 time blocks is sufficient to treat all 100 patients. This is important as it shows this model can potentially handle a patient booking rate of 2.5 patients per block without a backlog forming.

Table 1: Reducing the scheduling horizon.

Blocks	Patients Preferred	Patients Scheduled	Proportion of preferred	Ticks	Time (s)
100	100	100	1	44.00	0.06
80	100	100	1	44.47	0.05
60	93	100	0.93	37.55	0.05
40	80	100	0.8	32.13	0.03
20	54	74	0.73	69.66	0.10
10	30	41	0.73	93.14	0.18

The Treat-In-Turn policy was implemented by the Australian Government to ensure patients did not become lost in the system and experience unreasonable waiting times. It does this by ensuring that 60% of patients get treated in the order in which they join the system. The proportion of patients being treated in turn is based on research on previous less efficiency scheduling schemes, therefore could be adjusted to better suit the new model. A major reason to change the Treat-In-Turn proportion is that the higher the percentage the less overall people will be scheduled therefore a balance must be struck between overall throughput and fairness to patients.

Table 2 shows how the model is affected by changing the Treat-In-Turn proportion, noting the total number of patients scheduled in one of their preferred time blocks, the overall number of patients scheduled, the proportion of patients scheduled into one of their preferred time blocks, the number of ticks the model took to solve using CPLEX, and the time taken averaged over 5 runs.

As can be seen in Table 2, the Treat-In-Turn proportion had a major impact on the number of patients scheduled in one of their preferred time blocks. When the proportion was at 1.0 (meaning that 100% of patients must be in the longest-waiting cohort) less than half the patients that could be schedule. With so few patients being scheduled any gains in ensuring patients are seen promptly are lost compared to the reduced effectiveness.

The Treat-In-Turn policy is effective at its current level, although it does make the model more difficult to solve. The policy does not reduce the number of people in blocks they prefer for the data generated whilst still ensuring fairness. This trade of between model complexity, patients scheduled, and fairness is unfortunately intrinsically linked meaning a compromise between all three areas must be reached. The harder a patient is to schedule, due to preferences in popular times or less preferences in total, the more the patient benefits from the attempts to make the model fairer. This is because they are a worse fit in the schedule and models will instead try to priorities easier to fit patients.

Table 2: Investigating the Treat-in-Turn Proportion.

Blocks	Treat-in-Turn Proportion	Patients Preferred	Patients Scheduled	Proportion of preferred	Ticks	Time (s)
10	1.0	12	35	0.34	57.40	0.11
	0.9	19	41	0.46	254.95	0.48
	0.8	23	41	0.56	246.02	0.49
	0.7	27	41	0.66	71.85	0.13
	0.6	30	41	0.73	93.14	0.18
	0.5	30	41	0.73	53.67	0.11

6. Conclusion

In this paper we developed a model that uses patient preference to schedule outpatient appointments with the overall aim of improving patient no-show rates. This work is valuable as little research has been done on the topic of patient preference in scheduling outpatient appointments and even less work has been done on separating patient preference from their availability.

Computational experiments show that the model is effective in scheduling a majority of patients into time blocks they prefer and that the Treat-In-Turn policies proportion is at an effective level. Future work includes the incorporation of historical data for patient preference and availability, and determining how this model can further reduce patient no-show rates. We also wish to expand the model to include more real-world constraints.

7. References

Ahmadi-Javid, A., Jalali, Z., & Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research, 258*(1), 3-34.

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: a review of literature. *Production and Operations Management, 12*(4), 519-549.

Spratt, B. (2018). *Reactive Operating Theatre Scheduling.* (PhD Thesis), Queensland University of Technology, Brisbane.

Wang, J., & Fung, Y. (2014). An integer programming formulation for outpatient scheduling with patient preference. *Industrial Engineering & Management Systems, 13*(2), 193-202.

Zhu, H., Hou, M., Wang, C., & Zhou, M. (2012). An efficient outpatient scheduling approach. *IEEE Transactions on Automation science and engineering, 9*(4), 701-709.

8. Appendix

8.1. CPLEX

```
//scalars
int Bmax = 100;
int Pmax = 100;
int Omega = Pmax +1;
int Blocks = 10;
float epsilon = 0.0;
int M = 30;
//sets
{int} B = asSet(1..Bmax);
{int} P = asSet(1..Pmax);
//vector parameters
int U[P][B] = ...;
int V[P][B] = ...;
int W[P] = ...;
int N[B] = ...;
//decision variables
dvar boolean X[P][B];
```

```
dvar boolean theta[P];
dvar boolean eta[P];
dvar int E;
dvar int F;
//objective function
maximize
  sum (p in P, b in B)(U[p][b] + Omega*V[p][b])*X[p][b];
subject to
{
ctLongestLength:
sum(p in P) theta[p] == sum(p in P, b in B)X[p][b];

ctLongestShort:
forall(p in P)
  E >= (1-theta[p])*W[p];

ctShortestLong:
forall(p in P)
  F <= theta[p]*W[p]+(1-theta[p])*M;

ctLongestIsLongest:
E<=F;

ctLongestAndScheduled:
forall(p in P)
  eta[p] <= (theta[p]+sum(b in B)X[p][b])/2;

ctTreatInTurn:
sum(p in P)eta[p]>=epsilon*sum(p in P)theta[p];

ctCanBeScheduled:
forall(p in P, b in B)
  X[p][b]<=V[p][b]+U[p][b];

ctOneAppointment:
forall(p in P)
  sum(b in B)X[p][b]<=1;

ctDoctorLimit:
forall(b in B)
sum(p in P) X[p][b]<=N[b];

ctBlocksLimit:
forall(b in Blocks+1..Bmax)
  sum(p in P) X[p][b] ==0;
}
```

8.2. MatLab

```
%Generates the random data for the model
%Made by Jamie Owen 2019

%current day of the week, 1 = monday, 2 = tuesday ...
currentDay = 7;

%number of blocks per day
blocks = 20;
```

```
%number of days in planning horizon
days = 20;

% fraction of blocks available
available = 0.2;

% fraction of blocks preffered
preffered = 0.05;

%patient not available, available, preffered PDF
patientPDF = [1-available-preffered,available,preffered];
patientCDF = cumsum(patientPDF);

%number of patients
patients = 100;

%linear regression constants for initial waiting time
M = 4;
a = -214.60;
b = 887.60;
T = 2580;

%probability of being on the waiting list since a parrticular day
Pd=[0.12,0.24,0.16,0.20,0.22,0.03,0.02];
Pd = circshift(Pd,currentDay);
%sets for pdf construction of initial waiting time
w = 0:min(floor(-b/a),M)-1;
d = 1:7;

%construct waiting list pdf and cdf
waitingPDF = zeros(length(w)*length(d)+1,1);
for i = w
    for j = d
        waitingPDF(7*i+j+1)=Pd(j)*(a*i+b)/T;
    end
end
waitingPDF(1)=1-min(floor(-b/a),M)*(a*min(floor(-b/a),M)-a+2*b)/(2*T);
waitingCDF = cumsum(waitingPDF);
waitingCDF(end)=1;

%output data block declaration
patient = zeros(patients, blocks * days);
daysWait = zeros(patients, 1);
doctor = zeros(1, blocks * days);

%generate random uniform(0,1) data
randomPatient = rand(patients, blocks * days);
randomDaysWait = rand(patients, 1);

%generate the patient preference and availability
for i = 1:patients
    for j = 1: blocks * days
        patient(i,j) = find(randomPatient(i,j)<patientCDF,1);
    end
end
```

```
%generate days waiting
for i = 1:patients
    daysWait(i) =find(randomDaysWait(i)<waitingCDF,1);
end

%assign how many doctors are available in any time block
for i = 1: blocks * days
    doctor(end,i) = binornd(5,0.8);
end

%write data to file
xlswrite('testdata.xlsx',int8(patient==1),'unavailable');
xlswrite('testdata.xlsx',int8(patient==2),'available');
xlswrite('testdata.xlsx',int8(patient==3),'preffered');
xlswrite('testdata.xlsx',doctor,'doctorAvailability');
xlswrite('testdata.xlsx',daysWait,'daysWaited');
```