

OUTPATIENT APPOINTMENT SCHEDULING IN A MULTI FACILITY SYSTEM

Athula Wijewickrama

Department of Decision Sciences
Faculty of Management Studies & Commerce
University of Sri Jayewardenepura
Nugegoda, SRI LANKA

Soemon Takakuwa

Graduate School of Economics and
Business Administration
Furo-cho, Nagoya University
Nagoya, Aichi 464-8601, JAPAN

ABSTRACT

This study evaluates appointment systems used in hospitals by incorporating appointment rules and patient characteristics. Using an experiment unit at an internal medicine department of a large outpatient ward in Nagoya university hospital, a number of prevailing assumptions were relaxed, and twenty-five appointment systems were developed combining five appointment rules with five patient sequences. These appointment systems were evaluated under two different environments namely no-show and patient punctuality, with each of the two-levels totaling one-hundred different environments. A best appointment system is capable of identifying the problems in terms of both patient waiting time and doctor idle time.

1 INTRODUCTION

As the service sector now constitutes the largest economic sector in many postindustrial societies, both quality and productivity of the service have become growing concern. In general, the contemporary service economy has increased in size, but with slower growth in productivity. This is common to the healthcare industry, one of the largest industries in the service sector.

In healthcare, with greater emphasis on preventive medical practices, productivity improvement in outpatient department is become a central concern. For example, long wait times by patients seeking consultation has been a long term complaint. Today's customers seek quality service for the price they pay while fierce competition has intensified the industry further. Enhancing productivity while maintaining a high level of quality has become a challenge for healthcare managers. The major factor for patients in terms of quality concerns waiting time which has become a significant portion of determining the service quality.

Researchers have identified that the root of the problem in long waiting time for outpatients as being the result of an improper appointment system (AS) in place at a hospital. Over the last fifty-six years numerous researchers, from the

works of Bailey (1952) to Cayirli, Veral, and Rosen (2008), have investigated this issue extensively.

2 BRIEF LITERATURE REVIEW

Broadly speaking, outpatient appointment scheduling literature can be classified in to two groups: those that evaluated schedules and those that evaluated algorithms to uncover improve schedules (Kaandorp and Koole 2007). In case of the former type of study, simulations were used whereas the latter mostly applied analytical methods. A comprehensive review of the literature on AS can be found in Cayirli and Veral (2003).

The prevailing ASs range from single-block appointments on one extreme to individual appointments on the other side. Most of the appointment systems have concentrated on modification and combination of these two systems. Any combination in the appointment interval such as a block size or an initial block create an AS rule.

The *single-block* system assigns all patients to arrive in a block at the beginning of the clinic session, allocating a "date" rather than an exact appointment time (Babes and Sarma 1991). As expected, this system creates long waiting times for patients but shortens idle time for the doctors.

The *individual-block/fixe-d-interval* system calls patients individually at intervals equal to the mean consultation time (Klassen and Rohleder 1996; Rohleder and Klassen 2000; Cayirli, Veral, and Rosen 2006, 2008). The *individual-block/fixe-d-interval with an initial block* system is similar, but the number of patients assigned to the initial block is greater than one. Bailey (1952) introduced this rule to the AS literature, setting two patients at the beginning of the session. This simple Bailey rule was satisfactorily worked in many past studies. Ho and Lau (1992, 1999) evaluated and made additional amendments to the original Bailey rule. Klassen and Rohleder (1996), Yang, Lau, and Quek (1998), Cayirli, Veral, and Rosen (2006, 2008), Wijewickrama and Takakuwa (2005), and Kaandorp and Koole (2007) evaluated these rules in their studies. In addi-

tion, Brahim and Worthington (1991) suggested three patients for the initial block in their queuing model.

Dividing the clinic session into m blocks, the *multiple-block/fixed-interval* rule calls a fixed number of patients at the beginning of each block. Introducing this rule to the AS literature, Soriano (1966) advocated scheduling two-patients at a time with an interval of twice the consultation time. The two at a time rule was extensively studied with its original counterpart and its variations by Ho and Lau (1992, 1999) and Cayirli, Veral, and Rosen (2006, 2008). Cox, Birchall, and Wong (1985) investigated the *multiple-block/fixed-interval with an initial block* rule, introducing an initial block to the above rule studied by Soriano.

The *Variable-block/fixed-interval* rule assigns a different number of patients in a fixed appointment interval during the clinical session (Liu and Liu 1998). Using a dynamic programming approach, Lin (2000) optimized a quota AS.

The *Individual-block/variable-interval* rule calls patients individually with unequal appointment intervals. Ho and Lau (1992, 1999) introduced this rule by concluding that the variable-interval system, which sets an increasing appointment intervals towards the latter part of the session, works well in most environmental conditions. Many scholars have tested this rule using either analytical methods (e.g. Robinson and Chen 2003) or applying simulations (e.g. Yang, Lau and Quek 1998). Robinson and Chen (2003) identified the *dome* pattern of inter-arrival times which optimized the solution. This pattern made inter-arrival times shorter at the beginning and the ending part of a session, and longer in the middle, which represents a slight variation of variable interval rule. Cayirli, Veral, and Rosen (2006) modeled the *multiple-block/variable-interval* rule by combining the *dome* pattern of inter-arrival to “the two-at-a-time rule” and Bailey rules.

Nearly all the above studies concentrated only on appointment rule disrespecting the patient characteristics in designing the appointment systems. A number of studies considered the use of patient classification to sequence patients by classifying consultation time or type of procedure into groups based in order to design the AS (Cayirli, Veral, and Rosen 2006, 2008).

Klassen and Rohleder (1996) proposed a sequencing rule in scheduling based on the consultation time variance, whether a patient was considered “low variance” or “high variance.” The empirical research showed that setting low variance patients at the beginning of the session and high variance patients at the end of the session would minimize both patient waiting time and doctor idle time in most instances. Later Rohleder and Klassen (2000) identified that this rule is still effective when the scheduler is unable to identify the low variance and high variance patients proportionately.

Incorporating patient classification into AS rules, a number of ASs were designed under a wide range of clinical

environments by Cayirli, Veral, and Rosen (2006) by grouping patients as *new* and *returning*. They concluded that “sequencing decision have a more pronounced impact on performance than the choice of an appointment rule” (p. 57). More recently, by extending the scope of their previous research, Cayirli, Veral, and Rosen (2008) introduced well-designed ASs with adjusting appointment intervals to match the consultation time characteristics of different patient classes.

The contribution of our present study is to identify an AS in the real world by relaxing a number of unrealistic assumptions. First, it incorporated the various supporting facilities such as lab tests, x-ray, receptions and payment counters into the simulation model in order not to isolate the consultation room. Second, it considers patient as heterogeneous instead of patients as homogenous. Finally, the study took into account a second consultation time, and removed the popular assumption of “patients are served only once per visit.”

3 RESEARCH METHODOLOGY

3.1 Decision Factors

This study investigated two decision factors, namely the appointment rule (ARULE) and patient sequence (PSEQ), as originally explored by Cayirli, Veral, and Rosen (2006). The ARULE establishes the appointment interval and block size whereas the PSEQ determines the way patients are arranged or assigned to each block based on patient type.

The present study considers five ARULEs (Table 1). The rule Baily as the original Bailey rule (1952) which sets two-patients at the beginning of the session and then patient individually based on the mean consultation time. The rule 3Baily is a variation of the Baily setting three-patients for the initial block. The Ind is a *individual-block/fixed-interval* system which calls patients individually at intervals equal to the mean consultation time. The next rule, 2AtaTime, is the original Soriano rule (1966), which allocates two-patients at a time with an interval of twice the consultation time. The last rule is the variable-interval rule (V-I) introduced by Ho and Lau (1992) which adjusts appointment time so that the earlier patients setting earlier and the rest setting later compared to Ind. In this rule, k represented the medium patient allocated to each schedule, and it varied for each doctor in the current study, as a number of patients scheduled for each doctor was not identical unlike previous studies.

PSEQ is based on the types of patients. Three types of patients visit the clinic: appointment patients (APs), new patients (NPs), and walk-ins. As walk-ins have two standby doctors, they were excluded from the AS design. Hence, this study identified five sequencing rules based on APs and NPs.

Table 1: Appointment rules

Appointment Rule (ARULE)	Formulation
Baily: Two patients in the initial block and rest setting individually based on mean consultation time	$A_1 = A_2 = 0$ $A_i = A_{i-1} + \mu_i,$ for $i > 2$
3Baily: Three patients in the initial block and rest setting individually based on mean consultation time	$A_1 = A_2 = A_3 = 0$ $A_i = A_{i-1} + \mu_i,$ for $i > 3$
Ind: Each patient setting individually based on mean consultation time	$A_1 = 0$ $A_i = A_{i-1} + \mu_i,$ for $i > 1$
2AtaTime: Two patients at a time with an interval of twice the consultation time	$A_i = A_{i+1} = (i - 1)\mu_i,$ for $i = 1, 3, 5, \dots$
V-I: Number of k patients setting earlier and the rest ($n-k$) setting later, compared to Ind	$A_i = (i-1)\mu_i - 0.15(k-i)\sigma_i,$ for $i \leq k$ $A_i = (i-1)\mu_i + 0.30(i-k)\sigma_i,$ for $i > k$

NOTATION: A_i = Patient i th appointment time, μ_i = Mean consultation time, σ_i = Standard deviation of consultation time, n = Total number of patients in a given session, k = i th patient which divides n into two groups.

The rule of first-come first-serve (FCFS) represented the standard existing system in hospitals whereby patients were ignored by patient type or category of problem. The Appointed Patients Beginning Rule (APBEG) arranges all appointed patients first and then allows for the arrival of new patients. The New Patients at the Beginning Rule (NPBEG) sequences all new patients first. The Alternate Rule (ALTER) assigns appointed and new patients in an alternating manner. After assigning the last new patient, the rest of the appointment patients are assigned as the order of their arrival. Finally, the Alternate-5 Rule (ALTER5) assigns a new patient after every five-appointment patients. This is repeated until finding the last new patient in the schedule. The rest of the appointment patients are assigned as the order of their arrival.

3.2 Environmental Factors

This study investigated the impact of operating environments on the performance of appointment systems. A number of environmental factors were explored in the literature on appointment scheduling and this study focused on two factors: no-show (NOSHOW) and punctuality (PUNCT) are considered.

(minutes early), and altogether 100 different environments (5 ARULE * 5 PSEQ * 2 NOSHOW * 2 PUNCT).

A no-show simply represented the missed appointment without notification to the staff. Although there are many reasons behind such a problem, such as tendency of youth, lower socioeconomics status, government-provided health benefits, psychosocial problems, logistical issues, and emotional barriers (Lacy et al. 2004), interestingly, the AS itself may be a cause of a no-show. It increases with increasing time between scheduling and actual appointment (Lacy et al. 2004).

A no-show creates slack time for waiting-patients so it leads to a reduction in the waiting time. This further facilitates the system which have a second time consultation. On the other hand, no-show keeps doctors idle with the unexpected cancellation. Therefore, a no-show creates conflicting impact on system performance and upsets the expected function of the AS. As a no-show is unavoidable, it should be incorporated into the AS. A credible AS must have the ability to absorb the expected number of no-show without sacrificing its original benefits.

The impact of no-shows on system performance has been investigated by many past studies that range from 5 to 30 percent (Ho and Lau 1992, 1999; Klassen and Rohleder 1996; Yang, Lau and Quek 1998; Cayirli, Veral, and Rosen 2006, 2008; Kaandorp and Koole 2007). Ho and Lau (1992) reported that no-shows were the most influential factor on performance of AS among three environmental factors reviewed.

Patient problems with punctuality can be defined as the difference between the time of appointment and the time of arrival, which could be the result of either lateness (+) or earliness (-), though earliness is general considered on-time. Empirical evidence suggests that patients arrive early much more often than late (Brahimi and Worthington 1991). Earliness builds excessive congestion which tends to increase waiting time and reduce doctor's idle time. Past studies revealed mix results on the impact of punctuality on performance measures as being either unrelated (Blanco White and Pike 1964) or statistically significant (Cayirli, Veral, and Rosen 2006).

The impact of punctuality on waiting time could differ with the underlined method of calculation. According to Blanco White and Pike (1964), waiting time is the time difference between the consultation starting time and the actual arrival time of the patient. Another method is the difference between the greater of appointment time or arrival time and the consultation start time (Cayirli and Veral 2003). The latter excludes waiting prior to appointment time, since it is not a consequence of the AS.

The current study considers two levels of no-show (0% and 5%) and two levels of punctuality (on-time and 10

3.3 Performance Measures

The objective of this study is to identify an AS which performs well under both environmental conditions. Hence, it considers two conflicting performance measures, namely, patient waiting time per patient (*WT*) and doctor idle time per patient (*IT*).

$$WT = \frac{1}{n} \sum_{i=1}^n WT_i \quad (1)$$

$$IT = \frac{1}{n} \sum_{j=1}^d IT_j \quad (2)$$

WT_i is the time difference between the consultation start time and the actual arrival time of patient i in both first and second consultations. In either situation, arrival time is considered as the time a patient joins the queue for a consultation. Dividing the total waiting time by the number of patients seen in both the first and second consultation (n), the waiting time per patient (*WT*) is calculated.

In equation (2), IT_j is the total idle time of doctor j , which is the summation of idle times that occurs between two consecutive patients on doctor j 'th day schedule. That is:

$$IT_j = \sum_{i=1}^m IT_i \quad (3)$$

where IT_i is the idle time for patient i and m is the total number of patients seen by the j th doctor in his or her day session. Hence, the total idle time of doctor 1 to d in (2) is divided by the total number of patients seen (n) in both consultation times in order to derive the doctor's idle time per patient (*IT*).

To investigate the unique impact of ARULE, PSEQ, NOSHOW and PUNCT on performance measures, the *WT* and *IT* were considered separately without adding them up. Since *WT* and *IT* are conflicting objectives, the uncontaminated impact would tend to offset in case of adding these two variables together.

3.4 Simulation Model

This study uses a simulation model to characterize the interaction of the multi-facility system. The experimental unit was represented by the internal medicine unit of the outpatient department at the Nagoya university hospital, which has about five hundred twenty-five patients visits per day. Eighty nine percent and eight percent of patients are appointment and new, respectively. In ten clinical departments, there existed a total of thirty-one doctors engaged in consulting these two types of patients. These clinical departments provide the following number of doctors: hematology (3), diabetology and endocrinology (5), nephrology (2), geriatrics (2), home medicine (1), neurology (3), respirology (3), gastroenterology (5), cardiology (5), and general medi-

cine (2). The number of patients visiting each clinical department differs and ranges from 19 patients for home medicine to 107 patients in diabetology and endocrinology. This tends to assign a varied number of patients for each departmental doctor, such as eight in hematology to twenty-one in diabetology and endocrinology.

NPs are allocated to the available blocks of the schedule on the same day they visit where pre-APs are not filled. In measuring *WT* and *IT*, only these two types of patients with the thirty-one doctor panel is considered.

The remainder of the three percent of the total patients represented walk-ins and they were consulted by two stand-by doctors. As walk-ins did not disturb the AS and were excluded from the calculations of the *WT* and *IT*. Therefore, *WT* represents only appointment and new patients. Similarly, *IT* represented the 31 doctor panel which consulted only those two types of patients, APs and NPs.

During the consultation with the doctor, he/she decides whether the patient needs to be sent tests: blood, urinary, X-ray or endoscope or sent to the treatment room. The observations established that sixty-seven percent of the first-time consulted patients, after undergoing test/s or treatment, had to consult the same doctor (second time consultation) before leaving the hospital.

The principal data source of this study was the electronic medical records. This database consisted of the patient - appointment time, log-on and log-off time of the consultation service, the type of appointment (appointment and new), the category based on the required consultation service, and the treatment time of the treatment service, if any. The receptions and payment data were stored in another database. The test data were calculated via time studies with each sample size being 100. Other data important was collected via interviewing administrators, doctors, nurses, and other clerical personnel. A distance matrix was developed measuring the distance between every two facilities.

The appointment interval for any type of patients is ten minutes, while we assumed doctors were punctual. The ten minutes earliness followed a normal distribution with 2 minutes of standard deviation. The NOSHOW rate was the percentage of total number of patients, assuming identical to both types of patients. In order to calculate the effect of earliness (k_1) and the effect of lateness (k_2) in the V-I rule, best achieved parameters chosen for the study were those used by Ho and Lau (1992) and Cayirli, Veral, and Rosen (2006) and were 0.15 and 0.30, respectively.

Numerous sequence patterns based on a particular patient were developed in the simulation model. All these patterns and ASs were developed in VBA programs written in Excel. An animated simulation model was made using Arena shown in Figure 1. Statistics were collected after fifty replications for each of hundred different environments. In theoretical terms, the experimental unit could be identified as a multi-facility/phase, multiple-customer queuing system.

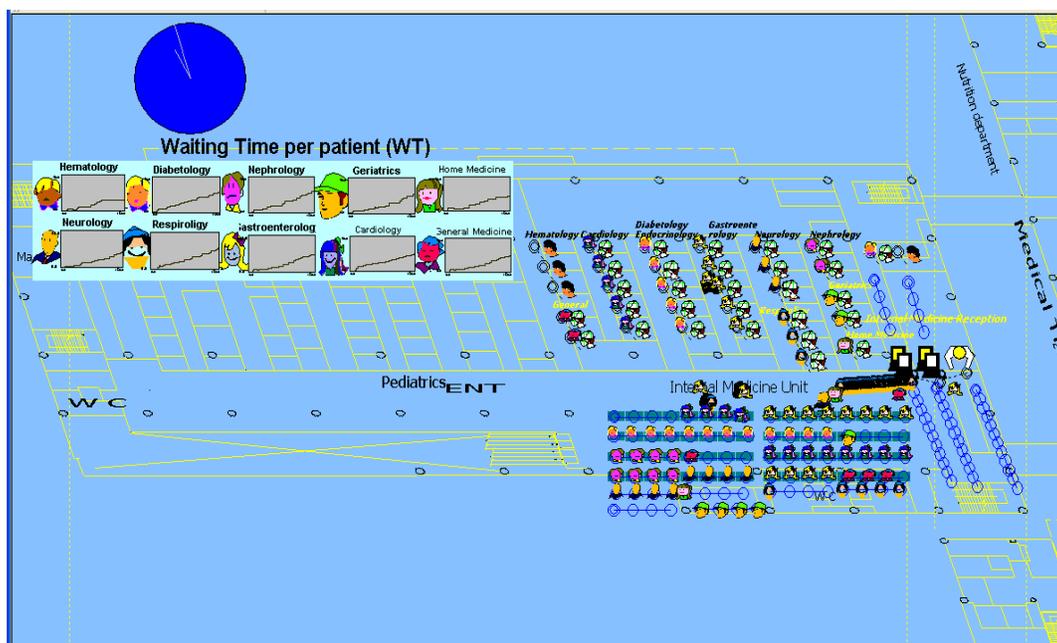


Figure 1: Part of the animation

The impact of independent variables (ARULE, PSEQ, NOSHOW, and PUNCT) on the dependent variable WT and IT require separate observation. As there are five different ARULEs and PSEQs, and two levels of NOSHOW and PUNCT, a full factorial 4-way design ANOVA was employed and statistical analysis was completed using SPSS.

4 ANALYSIS AND DISCUSSION

4.1 Waiting Time (WT)

Table 2 shows the results from ANOVA on the dependent variable WT. It is clear that all the main effects in determining the WT as statistically significant ($p < .001$). The effect size of those is also significantly larger seeing Eta squared with Cohen's criterion. However, the ARULE effect accounted for 71 percent of the total variance, which is large, and the PSEQ effect only accounted for 0.01 percent. Though these two factors were statistically significant, the PSEQ effect had much less of an impact on the WT than the effect of ARULE.

All levels of interaction effects (two, three and four) were also significant. The strength of association in determining WT was considerably high for ARULE*PUNCT, and moderate for ARULE*PSEQ and ARULE*PSEQ*PUNCT. However, the strength of the interaction weakens in any interaction associated with NOSHOW.

A post-hoc comparison using Tukey HSD indicated that the mean values of all ARULEs were significantly dif-

ferent from each other. In PSEQ, the Homogeneous Subsets show that the mean for FCFS did not differ significantly from that for APBEG. Overall, V-I reduced WT substantially, by 13.07, 6.85, 3.83, and 1.48 minutes per patient, against 3Baily, Baily, 2AtaTime, and Ind, respectively. Therefore, the rule 3Baily was inferior in all cases. For PSEQ, ALTER5 worked well over the other four sequencing rules in reducing WT while FCFS and APBEG deteriorated under any ARULE combination.

The above findings can be visualized with the interaction effect of ARULE*PSEQ on WT as shown in Figure 2. The stable and parallel lines of (b) show the fluctuations of ARULE in (a) which is not a result of PSEQ but the result of ARULE itself. The five-lines of WT that were arranged from V-I to 3Baily in (b) reveal that the impact of PSEQ on WT was very little. However, the steep slope of the ALTER5 towards Ind in (a) shows a considerable improvement of WT were obtained with a joint effect of Ind and ALTER5.

Figure 3 illustrates a three-way interaction of ARULE/PSEQ*NOSHOW*PUNCT on WT. When the environment is more volatile as in both NOSHOW and PUNCT were high, then Ind outperforms V-I, with the joined effect of ALTER5 and Ind.

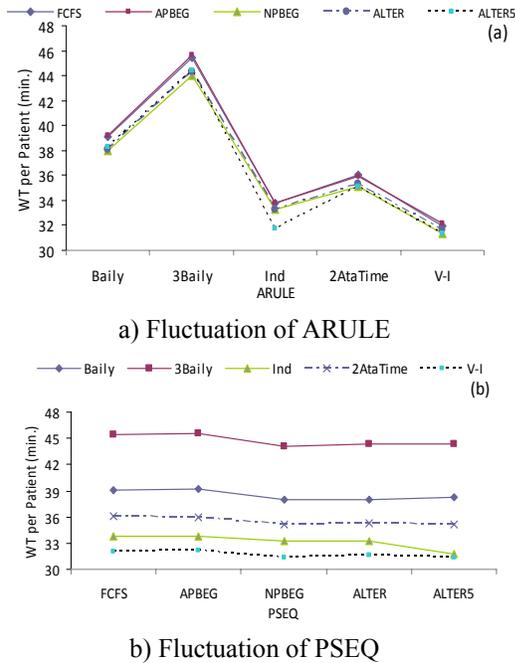


Figure 2: Interaction effect of ARULE and PSEQ on WT

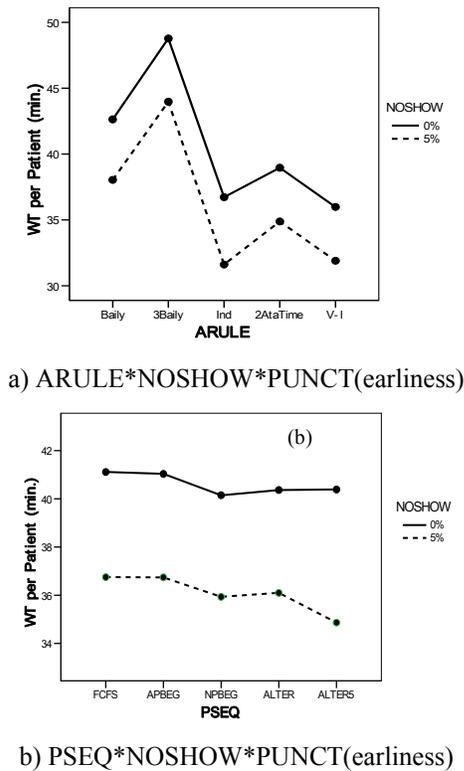


Figure 3: Three way interaction of the volatile environment

4.2 Idle Time (IT)

Similar to WT, all main and interactive effects were statistically significant as shown in Table 3. However, the practical significance of the strength of association in PSEQ on IT was very weak, considering the Eta squared. The effect of ARULE and PUNCT on the total variance accounted for 37 percent and 39 percent of variance, respectively. Except for ARULE*PUNCT interaction, all other interactions were not strong. The post-hoc comparison indicated that the mean differences were statistically different except for the 3Baily with 2AtaTime in ARULE, and APBEG with ALTER5 in PSEQ. It results, under any environmental level, Baily outperforms, conversely, V-I works poor in improving IT. In the case of PSEQ, the NPBEG performed best and ALTER5 deteriorated the performance of IT under any environmental condition.

4.3 Efficient Frontier

A well-designed AS must have the ability to reduce both the WT and IT simultaneously. Figure 4 shows the tradeoff between two performances under zero NOSHOW with on-time PUNCT. When the environment is certain, Baily records the least IT at the expense of high WT. Conversely, the rule V-I reduces WT while increasing IT. The Ind and 2AtaTime reduce both WT and IT considerably. Interestingly, Ind closes to the origin reducing both the idle times of patient and doctor simultaneously. Among the six-ASs recorded on the frontier, three of them were Ind. The other three are Baily, 2AtaTime, and V-I. All places for 3Baily were located far away from the origin due to increased WT.

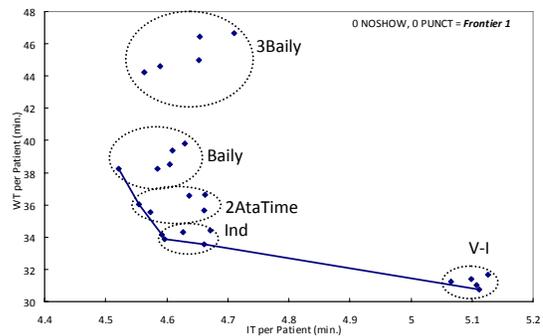


Figure 4: Efficient frontier

Being cautious, it appears that each group of ASs are scattered on ARULE than on PSEQ, justifying the findings of 4.1 and 4.2. Nevertheless, within each ARULE circle, the IT dispersion among PSEQs on 3Baily, Baily, and 2AtaTime groups were shrinking gradually when moving from upper-left to lower-right of the frontier. This implies that IT of ASs at the lower parts of the frontier such as Ind and V-I were more pronounced than the others.

Table 2: ANOVA results on dependent variable WT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	146724.785 ^(a)	99	1482.069	2125.950	.000	.977
Intercept	6746807.563	1	6746807.563	9677941.613	.000	.999
ARULE	107328.168	4	26832.042	38489.157	.000	.969
PSEQ	1161.888	4	290.472	416.667	.000	.254
NOSHOW	22751.824	1	22751.824	32636.298	.000	.869
PUNCT	12972.533	1	12972.533	18608.418	.000	.792
ARULE * PSEQ	375.005	16	23.438	33.620	.000	.099
ARULE * NOSHOW	111.139	4	27.785	39.856	.000	.032
PSEQ * NOSHOW	67.123	4	16.781	24.071	.000	.019
ARULE * PSEQ * NOSHOW	217.981	16	13.624	19.543	.000	.060
ARULE * PUNCT	866.838	4	216.710	310.858	.000	.202
PSEQ * PUNCT	53.178	4	13.294	19.070	.000	.015
ARULE * PSEQ * PUNCT	351.237	16	21.952	31.489	.000	.093
NOSHOW * PUNCT	90.670	1	90.670	130.062	.000	.026
ARULE * NOSHOW * PUNCT	55.563	4	13.891	19.926	.000	.016
PSEQ * NOSHOW * PUNCT	91.385	4	22.846	32.772	.000	.026
ARULE * PSEQ * NOSHOW * PUNCT	230.253	16	14.391	20.643	.000	.063
Error	3415.949	4900	.697			
Total	6896948.297	5000				
Corrected Total	150140.734	4999				

a. R Squared = .977 (Adjusted R Squared = .977)

Table 3: ANOVA results on dependent variable IT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	864.522 ^(a)	99	8.733	532.726	.000	.915
Intercept	130903.645	1	130903.645	7985733.338	.000	.999
ARULE	352.807	4	88.202	5380.719	.000	.815
PSEQ	7.890	4	1.973	120.336	.000	.089
NOSHOW	107.272	1	107.272	6544.110	.000	.572
PUNCT	364.819	1	364.819	22255.688	.000	.820
ARULE * PSEQ	3.019	16	.189	11.513	.000	.036
ARULE * NOSHOW	1.362	4	.341	20.777	.000	.017
PSEQ * NOSHOW	.846	4	.211	12.898	.000	.010
ARULE * PSEQ * NOSHOW	1.603	16	.100	6.113	.000	.020
ARULE * PUNCT	18.766	4	4.691	286.197	.000	.189
PSEQ * PUNCT	.946	4	.237	14.434	.000	.012
ARULE * PSEQ * PUNCT	1.482	16	.093	5.649	.000	.018
NOSHOW * PUNCT	1.063	1	1.063	64.839	.000	.013
ARULE * NOSHOW * PUNCT	.390	4	.098	5.953	.000	.005
PSEQ * NOSHOW * PUNCT	.872	4	.218	13.300	.000	.011
ARULE * PSEQ * NOSHOW * PUNCT	1.384	16	.086	5.276	.000	.017
Error	80.322	4900	.016			
Total	131848.489	5000				
Corrected Total	944.844	4999				

a. R Squared = .915 (Adjusted R Squared = .913)

Figure 5 depicts all four-frontiers that corresponded to each environmental situation as applied in the simulation. The behavior of Frontier 2 was the result of 5% NOSHOW compared to Frontier 1 (0 NOSHOW, on-time PUNCT). As NOSHOW create slack time for waiting patients, their WT tends to be reduced while IT increases due to unexpected cancellation. Frontier 3 was the result of incorporating 10 minutes earliness into the frontier 1. Earliness increases congestions which in turn increases waiting time. At such a moment, IT was surprisingly increased along with the congestion. This becomes obvious when considering a second time consultation, since the doctor has to wait until the last scheduled patient arrives for a consultation. Finally, frontier 4 represents the most volatile environment with 5% NOSHOW and 10 minutes earliness. As the impact of NOSHOW on WT was more powerful than of PUNCT shown in Table 2, WT was becoming comparatively lower than frontier 1. However, the effect was reversed on IT as PUNCT became more powerful than NOSHOW in determining IT, as shown in Table 3.

Under any circumstance, Baily shows a low IT with increased WT. On the contrary, V-I shows a low WT with increased IT. These matched well with the previous findings of Ho and Lau (1992, 1999), Klassen and Rohleder (1996), and Cayirli, Veral, and Rosen (2006). However, in the more volatile environment, the popular Baily and V-I rules were dislodged from the frontier due to escalating WT or IT.

It was interesting to observe that slightly more than half of the designed ASs (twelve out of twenty-three) using simple Ind were placed *on* the frontiers. These placements were concentrated towards the origin, showing their power to reduce both idle times simultaneously.

In case of the PSEQ, the NPBEG contributed in designing twelve out of twenty-three ASs. Any ARULE that was combined with the FCFS performed poorly in both performances which signals the importance of sequencing patients in designing AS. In addition, the APBEG was also missing under any environmental condition. These three findings were in line with the finding of Cayirli, Veral, and Rosen (2006).

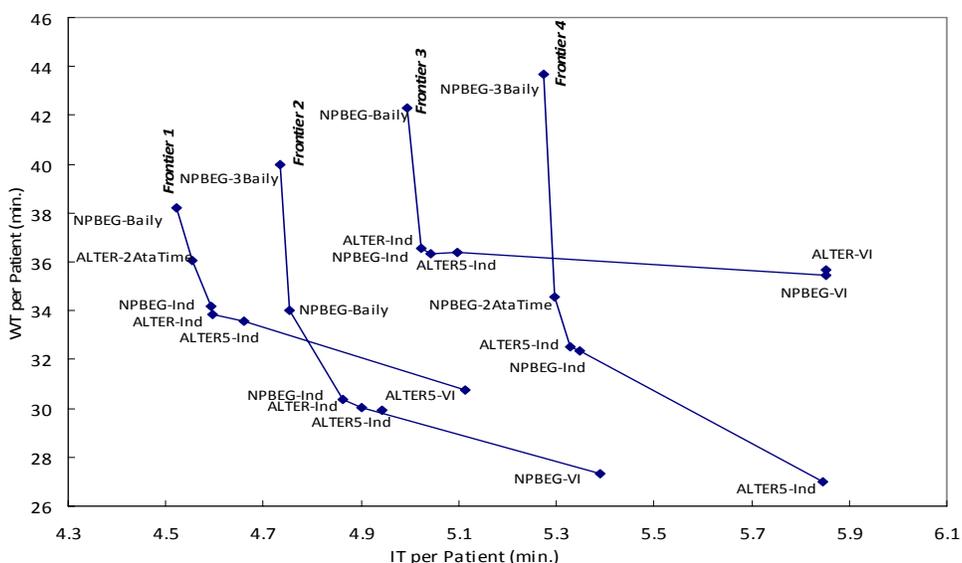


Figure 5: Frontiers on four environments

5 CONCLUSION

A simulation model was developed to evaluate ASs in a multi-facility system while relaxing a number of assumptions, especially considering second-time consultations. These systems were designed by incorporating both ARULE and PSEQ. The results were revealed that this collective effort leads to design most effectively than they work alone. As all ASs which combined the FCFS sequence deteriorated with the performances, the consideration of PSEQ in designing ASs is of vital concern. However, it was confirmed that in developing a well-designed AS, the impact of ARULE was categorically higher than that of PSEQ.

The simple Ind rule works well in reducing both WT and IT simultaneously, and it is robust under differing environmental situations. The rules Baily and V-I worked best in achieving a single-performance for IT and WT, respectively, but being inferior at the other extreme.

Considering two NOSHOW rates and two PUNCT times, in all four-situations, the Ind scattered towards the origin of the efficient frontier, showing its power to improve both performances simultaneously. In the most volatile environment, the popular Baily and V-I disappeared from the frontier due to increased WT or IT.

The findings of the current study are not easy to generalize as the data source is restricted to a single outpatient

center. Therefore, it would be worthwhile to examine the power of PSEQ and ARULE in designing AS further by using more empirical data.

ACKNOWLEDGMENTS

This research was supported by “Grant-in-Aid for Scientific Research” of Japan Society for the Promotion of Science (JSPS).

REFERENCES

- Bailey, N. T. J. 1952. A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society* 14:185-199.
- Blanco White, M. J., and M. C. Pike. 1964. Appointment systems in out-patients’ clinics and the effect of patients’ unpunctuality. *Medical Care* 2:133-145.
- Brahimi, M., and D. Worthington. 1991. Queuing models for out-patient appointment systems – a case study. *Journal of the Operational Research Society* 5:91-102.
- Babes, M., and G. V. Sarma. 1991. Out-patient queues at the Ibn-Rochd health center. *Journal of the Operational Research Society* 42:845-855.
- Cayirli, T., and E. Veral. 2003. Outpatient scheduling in health care: A review of literature. *Production and Operations Management* 12:519-549.
- Cayirli, T., E. Veral, and H. Rosen. 2006. Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science* 9:47-58.
- Cayirli, T., E. Veral, and H. Rosen. 2008. Assessment of patient classification in appointment system design. *Production and Operations Management* 17:338-353.
- Cox, T. F., J. F. Birchall, and H. Wong. 1985. Optimizing the queuing system for an ear, nose and throat outpatient clinic. *Journal of Applied Statistics* 12:113-126.
- Ho, C., and H. Lau. 1992. Minimizing total cost in scheduling outpatient appointments. *Management Science* 38:1750-1764.
- Ho, C., and H. Lau. 1999. Evaluating the impact of operating conditions on the performance of appointment scheduling rules in service systems. *European Journal of Operational Research* 112:542-553.
- Kaandorp, G. C., and G. Koole. 2007. Optimal outpatient appointment scheduling. *Health Care Management Science* 10:217-229.
- Klassen, K. J., and T. R. Rohleder. 1996. Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management* 14:83-101.
- Lacy, N. L., A. Paulman, M. Reuter, and B. Lovejoy. 2004. Why we don’t come: Patient perceptions on no-shows. *Annals of Family Medicine* 2:541-545.
- Lin, C. K. Y. 2000. Analysis of quota assignment and time block duration in multi-block appointment systems. *IMA Journal of Mathematics Applied in Business and Industry* 11:85-99.
- Liu, L., and X. Liu. 1998. Block appointment systems for outpatient clinics with multiple doctors. *Journal of the Operational Research Society* 49:1254-1259.
- Robinson, L. W., and R. R. Chen. 2003. Scheduling doctors’ appointments: Optimal and empirically-based heuristic policies. *IIE Transactions* 35:295-307.
- Rohleder, T. R., and K. J. Klassen. 2000. Using client-variance information to improve dynamic appointment scheduling performance. *Omega* 28:293-302.
- Soriano, A. 1966. Comparison of two scheduling systems. *Operations Research* 14:388-397.
- Wijewickrama, A., and S. Takakuwa. 2005. Simulation analysis of appointment scheduling in an outpatient department of internal medicine. In *Proceedings of the 2005 Winter Simulation Conference*, ed. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 2264-2273. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Yang, K. K., M. L. Lau, and S. A. Quek. 1998. A new appointment rule for a single-server, multiple-customer service system. *Naval Research Logistics* 45:313-326.

AUTHOR BIOGRAPHIES

ATHULA WIJEWICKRAMA is a senior lecturer in the department of Decision Sciences at the University of Sri Jayewardenepura, Sri Lanka. He pursued a B.Sc. Degree in business administration from Sri Jayewardenepura University and an MBA Degree from Colombo University, Sri Lanka. He received a Master’s and PhD in industrial management systems at Nagoya University. He has a diploma in computer systems design from the National Institute of Business Management, Sri Lanka. His research and teaching interests are in simulation of service systems and operations management. Currently he is working as a postdoctoral research fellow at Nagoya University.

SOEMON TAKAKUWA is a Professor in the Graduate School of Economics and Business Administration at Nagoya University in Japan. He received his B.Sc. and M.Sc. degrees in industrial engineering from Nagoya Institute of Technology in 1975 and from the Tokyo Institute of Technology in 1977 respectively. His Ph.D. is in industrial engineering from The Pennsylvania State University. He holds a Doctorate of Economics from Nagoya University. His research interests include optimization of manufacturing and logistics systems, management information system and system simulation. He has prepared the Japanese editions of both the introductory to simulation using SIMAN and the Simulation with ARENA. He serves concurrently as the senior staff of Department of Hospital Management Strategy and Planning at Nagoya University Hospital.