The outpatient appointment system : design of a simulation study

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

The outpatient appointment system: Design of a simulation study

J. VISSERS and J. WIJNGAARD

Department of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

This paper describes the model-construction of a simulation study. The purpose of this study was to produce a general method for determining a suitable appointment system for the clinics in the outpatient department of a hospital. The original model contained 11 variables. Investigation of the influence of each variable on patients' waiting-time and doctors' idle-time showed that a considerable reduction in the number of variables could be achieved. Only 5 variables were finally left in the simulation.

The use of the results of this study in a real-life clinic situation is discussed elsewhere.

1. Introduction

The appointment system of an outpatient department has been the subject of study many times, usually through the means of simulation techniques. First of all, Bailey [1,2] investigated a clinic assuming strict punctuality of patients. Blanco White and Pike [3] examined the effect of patients' unpunctuality. Fetter and Thompson [4] investigated the effect of a number of variables on waiting and idle-time. Soriano [5] followed an analytical approach assuming a steady-state distribution of the waiting-time. These studies, however, have not led to a generally applicable method of determining a suitable appointment system. The difficulty often lies in the large number of variables. In this paper it is shown how the number of variables can be reduced in such a way that the output is restricted to a one-page table or a few graphs with waiting and idle-time results. These results can easily be used in most clinics to design a suitable appointment system.

2. Problem formulation

The investigated problem can be stated as follows: which appointment system is suitable for a given clinic and given standards on permissible waiting and idle-time?

An appointment system is characterized by:

(1) the number of patients given an identical appointment-time at the beginning of the clinic session (beginblock: n_0),

(2) the number of patients given an identical appointment-time during the clinic session (blocksize: n),

(3) the interval between two successive appointment-times (appointment-interval: a).

Most appointment systems can be described by these three variables. The range of common appointment systems may vary from an individual system (all patients have different appointments) to an extreme block system (all patients are scheduled at the start of the clinic).

A clinic can be characterized by:

(1) the mean and standard deviation of the consultation-time (the time the doctor spends on a patient),

(2) the mean and standard deviation of patients' punctuality (the difference between his appointment-time and the time of arrival),

(3) the mean time the clinic session starts (the difference between the scheduled and the real start of the clinic session),

(4) the fraction of no-show (the number of patients that do not show up, divided by the number of appointments),

(5) the fraction of walk-ins (the number of patients that come without having an appointment, divided by the number of appointments),

(6) the priority rule (the order in which the patients are seen),

(7) the number of appointments made for the clinic session.

Most clinics have a utilization factor of at least 1, the utilization factor being determined by the ratio between the average consultation-time and the average interarrival-time. In this paper we shall assume a utilization factor of 1, unless stated otherwise. The service-

mechanism of a clinic session is complicate ', but normally patients are seen in appointment order; when a patient does not show up, the next waiting patient is in turn. This mechanism is also assumed in this paper. Further restrictions are that this paper investigates a single server system and that preliminary visits to facilities like x-ray and laboratory are not taken into account.

The performance of an appointment system is measured in patients' mean waiting-time (the mean time between his arrival at the clinic and his first being seen by the doctor) and doctors' idle-time (the total sum of the times during the clinic session when the doctor is not consulting because there are no patients waiting to be seen). The performance of the system, however, is better represented by subtracting the patient's own earliness from the patient's mean waiting-time. The system is not responsible if the patient shows up before his appointment-time.

3. Reduction of the number of variables

The variability in a given clinic situation can make it necessary to insure the doctor against too much idle-time. For example, a surgeon's clinic is more subject to variability and therefore less organizable than the clinic of a medical consultant. In a real clinic this insurance against the risk of running idle is reached in one or more of the following ways:

(1) patients' earliness,

(2) block-booking (using a larger blocksize than $n_0 \neq$),

(3) initial block-booking (using a larger beginblock than the normal blocksiz),

(4) doctor's lateness,

(5) utilization factor >1.

It should be noted that the first variable is at the patients' discretion, whereas the other variables are at the doctor's discretion. All these different methods have the same purpose, that is, letting patients come on average earlier than the expected moment of treatment. This underlying variable will be referred to as system earliness. All methods described can be translated into this one variable. For example, in a clinic with a size of 35 patients and a mean consultation-time of 5 minutes, the following appointment systems will create a system earliness of 5 minutes:

(1) patients' earliness of 5 minutes ($n_0 = n = 1$, $a = 5$),

(2) block-booking ($n_0 = n = 3, a = 15$),

(3) initial block-booking ($n_0 = 2, n = 1, a = 5$),

(4) doctor's lateness of 5 minutes ($n_0 = n = 1, a = 5$),

(5) utilization factor >1 ($n_0 = n = 1, a = 4.72$).

In the last case the eighteenth patient should arrive 5 minutes before his expected moment of treatment, which creates a utilization factor of 1.06. In Table 1 it is shown that waiting and idle-time for all these appointment systems are about the same, only the waiting-time for the last case is somewhat higher. This means that instead of these five variables one can use one variable, namely system earliness. Partial results of this type can also be found in [3].

In the same way it is possible to combine the variables standard deviation of consultation-time, fraction of no-show and fraction of walk-ins. When a patient does not show up, this can be interpreted approximately as a patient with consultation-time 0; an extra patient without appointment can be considered as an appointment needing an extra consultation. We shall demonstrate that this interpretation is adequate for the case of no-show, by calculating a revised mean and standard deviation for the consultation-time.

Let the mean consultation-time be m_c minutes, the standard deviation s_c and the fraction of no-show p.

The revised mean now becomes

$$m_c' = (1 - p)m_c + (p)0 = (1 - p)m_c \qquad (1)$$

Table 1

Comparison of different methods in obtaining a system earliness of 5 minutes for the example given (average results of 100 clinic sessions)

	Mean waiting-time with standard error	Idle-time with standard error
patients' earliness	10.57 min. ± 0.50	12.26 min. ± 0.89
block-booking	10.15 min. ± 0.45	12.12 min. ± 0.90
initial block-booking	10.49 min. ± 0.50	12.29 min. ± 0.88
doctor's lateness	10.61 min. ± 0.50	11.65 min. ± 0.85
utilization factor >1	11.88 min. ± 0.55	11.90 min. ± 0.78

and the revised variance

$$s_c'^2 = (1-p)\int (c - m_c')^2 f(c)\, dc + p(0 - m_c')^2$$

$$= (1-p)\int ((c - m_c) + (m_c - m_c'))^2 f(c)\, dc + pm_c'^2$$

$$= (1-p)\int (c - m_c)^2 f(c)\, dc + 2(1-p)(m_c - m_c')$$

$$\times \int (c - m_c) f(c)\, dc + (1-p)(m_c - m_c')^2$$

$$\times \int f(c)\, dc + pm_c'^2$$

$$= (1-p)s_c^2 + 0 + (1-p)p^2 m_c^2 + p(1-p)^2 m_c^2$$

$$= (1-p)s_c^2 + p(1-p)m_c^2 \tag{2}$$

where $f(c)$ stands for distribution of the consultation-time.

In Table 2 two clinics are compared. Clinic 1 is characterized by a mean consultation-time of 5 minutes, a coefficient of variation of the consultation-time (the standard deviation divided by the mean) of 0.50 and 20% no-show (which is compensated for by making the appointment-interval 20% shorter i.e. 4 minutes). Clinic 2 is therefore characterized by a mean consultation-time of 4 minutes (eq. (1)) and a coefficient of variation of 0.75 (using eqs. (1) and (2)). Comparison of the results shows that both clinics give about the same waiting and idle-time.

The effect of walk-ins can be found in a similar way by interpreting the fraction of walk-ins (p) as the probability that patients need a revised consultation-time equal to the sum of two consultation-times. In this case the following expressions can be derived for the revised mean and variance of the consultation-time:

$$m_c' = (1+p)m_c ,$$

$$s_c'^2 = (1+p)s_c^2 + p(1-p)m_c^2 .$$

This means that the effect of no-show and walk-ins can be found by means of their influence on mean consultation-time and coefficient of variation.

Through this reduction of variables the design of the simulation experiment could be restricted to the following variables:

(1) the mean consultation-time,

(2) the coefficient of variation of the consultation-time,

(3) the mean system earliness,

(4) the standard deviation of patients' punctuality, and

(5) the number of appointments.

4. Results

The relationship between the five variables mentioned in the foregoing section and the expected mean waiting-time and idle-time was investigated by means of a computer simulation model. In Table 3 and Fig. 1 some results are shown for a clinic-size of 20 patients. The mean consultation-time is taken as unit of time which makes the results independent of the mean consultation-time. Since the effect of the standard deviation of patients' punctuality appeared to be not so strong, this variable was incorporated by means of a correction-factor. The results in Table 3 and Fig. 1 are given for a standard deviation of patients' punctuality of 3 times the mean consultation-time.

Correction factors for the difference between the

Table 2
Comparison of two clinics (average results of 100 clinic sessions)

	clinic 1	clinic 2
mean consultation-time	5 min.	4 min.
coefficient of variation	0.50	0.75
number of appointments	50	50
beginblock = blocksize	1	1
fraction of no-show	0.20	–
appointment-interval	4 min.	4 min.
number of patients treated	40	50
mean-waiting-time and standard error	12.27 min. ± 0.67	12.01 min. ± 0.69
idle-time and standard error	18.77 min. ± 1.34	18.76 min. ± 1.23

Table 3

Summary of waiting and idle-time results (in units of mean consultation-time) for different system earliness (SE) and coefficient of variation (CV), with a clinic-size of 20 patients (average results of 25 sessions)

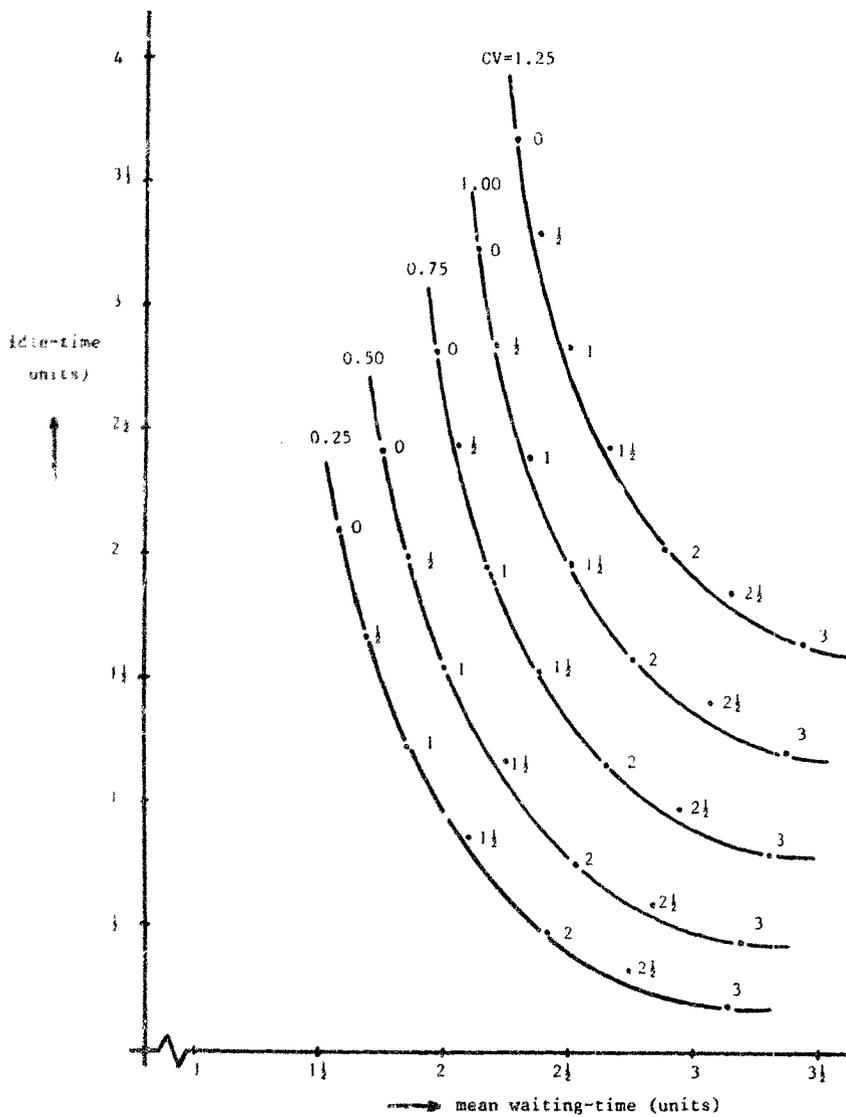| SE	CV	mean waiting-time (units)					idle-time (units)				
		0.25	0.50	0.75	1.00	1.25	0.25	0.50	0.75	1.00	1.25
0		1.6	1.7	2.0	2.1	2.3	2.2	2.5	2.9	3.3	3.7
½		1.7	1.8	2.0	2.2	2.4	1.7	2.0	2.5	2.9	3.3
1		1.8	2.0	2.2	2.3	2.5	1.3	1.6	2.0	2.4	2.9
1½		2.1	2.2	2.4	2.5	2.6	0.9	1.2	1.6	2.0	2.5
2		2.4	2.5	2.6	2.8	2.9	0.5	0.8	1.2	1.6	2.1
2½		2.7	2.8	2.9	3.1	3.1	0.4	0.6	1.0	1.4	1.9
3		3.1	3.2	3.3	3.4	3.3	0.2	0.5	0.8	1.2	1.7

Fig. 1. The waiting and idle-time results for a clinic of 20 patients and different coefficients of variation (average results of 25 sessions). The figures along the curve refer to the system earliness. Waiting-time, idle-time and system earliness are expressed in units, the mean consultation-time taken as one unit of time.

Table 4

Correction factors for the difference between the simulated standard deviation of patients' punctuality of 3 units and the true standard deviation (average results of 25 clinic sessions)

number of appointments	correction-factor (units)	
	waiting-time	idle-time
10	0.3 *	0.4
30	0.2	0.3
50	0.1	0.3

* A correction-factor of 0.3 for the waiting-time means that, in case of a real standard deviation of $3 + x$, the waiting-time should be increased by $0.3\ x$.

simulated standard deviation of patients' punctuality of 3 units and the real standard deviation are given in Table 4.

These results can be used in the design of a suitable appointment system. First the investigated clinic should be defined by specifying the variables mentioned in the first section, expressing the scale-dependent variables in units of the mean consultation-time. As far as necessary, variables are combined in the way described in the foregoing section. Next, with use of the simulated results, waiting and idle-time for the investigated clinic can be determined and a suitable appointment system can be found which meets given standards on permissible waiting and idle-time. Finally, a correction is possible for the difference between the simulated standard deviation of patients' punctuality and the true standard deviation. The use of the results of this study in a real-life clinic is discussed in more detail in [6].

References

[1] N.T.J. Bailey, A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting times, J. Roy. Statist. Soc., B. 14 (1952) 185–199.

[2] N.T.J. Bailey, A note on equalising the mean waiting time of successive customers in a finite queue, J. Roy. Statist. Soc., B.17 (1955) 262–263.

[3] M.J. Blanco White and M.C. Pike, Appointment systems in out-patients' clinics and the effect of patients' unpunctuality, Medical Care 2 (1964) 133–145.

[4] R.B. Fetter and J.D. Thompson, Patients' waiting time and doctors' idle time in the outpatient setting, Health Services Research 1 (1966) 66–90.

[5] A. Soriano, Comparison of two scheduling systems, Operations Research 14, 388–397.

[6] J.M.H. Vissers, Selecting a suitable appointment system in an outpatient setting, Medical Care, to appear.