

appointment scheduling



by **Christos Zacharias**, NYU Stern

About: We develop a user-friendly application that generates appointment schedules that achieve both high quality of service and high operational efficiency. **appointment scheduling** will enable health-care providers utilize their valuable resources efficiently, while their patients enjoy timely access and short waiting times.

1 Motivation

Appointment schedules play an important role in the health care delivery system. However, it is common for patients to not show up for their scheduled appointment.

Patients' no-show behavior and its negative impact on medical practice have been documented in many studies. To name a few: a) Perio and Niemeier (2010) report a 26% no-show rate for follow-up visits of tuberculin skin test, b) Defife et al. (2010) report a 21% appointment no-show rate in psychotherapy appointments, c) Dreiherr et al. (2008) report a 30% proportion of nonattendance in an outpatient obstetrics and gynecology clinic.

Unattended appointments result in under-utilization of a clinic's valuable resources. Various methods are being implemented to tackle the "no-show" phenomenon. A few, which do not completely resolve the issue, are: a) Sending appointment reminders (e-mail, postcards, cell phone), b) charging a no-show penalty, c) encouraging patients to cancel or reschedule in advance.

Another, more effective way to alleviate the negative impact of no-shows, is the practice of overbooking. However, overbooking potentially results in clinic's overcrowding, with increased patients' waiting times and physician's overtime. As argued in Krueger (2009): "Patient time is an important input in the health care system. Failing to take account of patient time leads us to exaggerate the productivity of the health care sector, and to understate the cost of health care". LaGanga and Lawrence (2007) show that a sensible use of appointment overbooking can significantly improve a clinic's performance by increasing patients' access to care and improving the clinic's productivity.

An optimal appointment schedule balances the trade-offs between the benefits of efficient resource utilization, and the costs of patients' waiting time and physician's overtime. Zacharias and Pinedo (2012) find that the no-show rate has a significant impact on a clinic's operational performance and should be taken under consideration in appointment scheduling.

We have designed an application for scheduling the arrivals of patients at a medical facility. Our application, by promoting a sensible use of overbooking to compensate for no-shows, generates appointment schedules that achieve both high quality of service and high operational efficiency. Quality is translated into shorter patients' waiting times, and efficiency is translated into better resource utilization.

"Aside from boredom and physical discomfort, the subtler misery of waiting is the knowledge that one's most precious resource, time, a fraction of one's life, is being stolen away, irrecoverably lost", Morrow (1984).

2 The app

Consider a primary care physician with a certain number of time slots available to see patients within one working day (for example 15 time slots per day). Patients' arrivals are driven by scheduled appointments: each patient is scheduled to arrive at one of the available time slots. We account for patients' no-show behavior: it is uncertain whether patients will show up for their scheduled appointment.

Our app operates in three regimes:

- **Efficiency Regime (ER):** Primary attention is paid to clinic's efficient utilization, while patients' waiting times are kept short.
- **Quality & Efficiency Regime (QER):** Equal attention is paid to clinic's efficient utilization and to quality of service.
- **Quality Regime (QR):** Primary attention is paid to providing short waiting times, while maintaining high clinic's efficiency.

For our proposed application we use the graphical representation of an appointment schedule as in Figure 1. The schedule in Figure 1 concerns a working day of 20 time slots. There is some overbooking in order to compensate for the no-show behavior. Strategically chosen, slots 1,5, and 11 have two patients assigned to them, whereas the rest of the slots have just one patient assigned to them.

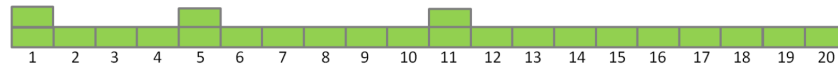


Figure 1: Graphical Representation of a Schedule

We have developed a graphical user interface in Java for generating optimal appointment schedules. The user (scheduler) is able to obtain the customized optimal appointment schedule for a medical facility, by following 4 simple steps:

1. Choose the number of appointment slots per working day.
2. Choose the observed patients' no-show rate.
3. Choose the operational regime.
4. Click the "Generate Schedule" button.

The attached .jar file is a demo of [appointment scheduling](#). Try it. It is easy to use, fast, and explicit.

3 The Underlying Model

Consider a primary care physician with n time slots available to see patients within one working day (for example $n = 20$ time slots per day). Patients' arrivals are driven by scheduled appointments. Each patient is scheduled to arrive at one of the available time slots. We account for patients' no-show behavior: each patient is expected to not show up with some probability p , which can easily be estimated from a clinic's records.

There are three costs associated with an appointment schedule: patients' waiting cost, and doctor's idle time and overtime costs. The objective is to minimize the weighted sum of the three costs. If there are no patients present during one of n time slots, the service provider remains idle and an idle time cost c_I is incurred. An overtime cost c_O is incurred for each overtime slot that the server has to remain present at the medical facility to see patients. The scheduler may overbook certain time slots and assign more than one customer to them in order to compensate for the no-show behavior. If several patients are present at the beginning of a time slot due to overbooking, then all but one of these patients have to wait. A waiting cost w is incurred for each time slot that a patient has to wait to see the physician.

Such overbooking models have been considered in the literature of appointment scheduling (see for example Robinson and Chen (2010), LaGanga and Lawrence (2012), Zacharias and Pinedo (2013)). Finding an optimal schedule is analytically intractable, and thus, most of the papers in the literature use enumeration, search algorithms, simulation, and/or heuristics.

Following the literature, see Robinson and Chen (2010,2011), we consider an overtime cost coefficient $c_O = 1.5 \times c_I$. The waiting cost coefficient w takes values between $0.1 \times c_I$ and $0.2 \times c_I$, depending on the operational regime of the clinic. In particular, we consider three regimes:

- **Efficiency Regime (ER):** Corresponds to $\frac{w}{c_I} = 0.1$.
- **Quality & Efficiency Regime (QER):** Corresponds to $\frac{w}{c_I} = 0.15$.
- **Quality Regime (QR):** Corresponds to $\frac{w}{c_I} = 0.2$.

Optimal schedules are presented in Table 1, for different operational regimes and different no-show probabilities p . For this particular example a working day consists of 18 time slots. It is evident that the overbooking level and the structure of the optimal schedule highly depend on the operational regime and the no-show rate.

Regime	$p = 0.2$	$p = 0.3$	$p = 0.4$
ER	2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1	2 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1	3 1 2 1 2 1 2 1 2 1 2 1 1 2 1 1 1 1
QER	2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1	2 2 1 2 1 2 1 1 2 1 2 1 1 2 1 1 1 1
QR	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1	2 1 2 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1

Table 1: Optimal Schedules for Different Operational Regimes and No-Show Rates.

References

- Defife, J.A., C.Z. Conklin, J.M. Smith, J. Poole. 2010. Psychotherapy appointment no-shows: Rates and reasons. *Psychotherapy: Theory, Research, Practice, Training*, **47**(3): 413-417.
- Dreiher, J., M. Froimovici, Y. Bibi, D.A. Vardy, A. Cicurel, A.D. Cohen. 2008. Nonattendance in Obstetrics and Gynecology Patients. *Gynecologic and Obstetric Investigations*. **66**: 40-43.
- Krueger, A.B. 2009. February 10. A Hidden Cost of Health Care: Patient Time. *New York Times*.
- LaGanga, L.R., S.R. Lawrence. 2007. Clinic overbooking to improve patient access and provider productivity. *Decision Sciences*. **38**(2): 251-276.
- LaGanga, L.R., S.R. Lawrence. 2012. Appointment overbooking in health care clinics to improve patient service and clinic Performance. *Production and Operations Management*. **21** (5): 874-888.
- Perio, M.A., R.T. Niemeier. 2010. Evaluation of Exposure to Tuberculosis Among Immigration Employees. *Health Hazard Evaluation (HHE) report, Chicago, Illinois*.
- Robinson, L. W., R. R. Chen. 2010. A comparison of traditional and open-access policies for appointment scheduling. *MSOM*. **12**(2): 330-346.
- Robinson, L. W., R. R. Chen. 2011. Estimating the implied value of the customer's waiting time. *MSOM*. **13**(1): 53-57.
- Zacharias, C., M. Pinedo. 2013. Appointment Scheduling with No-Shows and Overbooking. *Production and Operations Management*. To appear.

Appendix

The purpose of this appendix is to demonstrate the recursive expressions that completely describe the cost associated with a schedule, and to describe our computational procedure. In finding an optimal schedule, we use enumeration among the possible schedules. We find it necessary to introduce some notation.

Let a schedule be denoted by a vector $\bar{s} = (s_1, \dots, s_n)$, where s_t is the number of customers assigned to slot t , $1 \leq t \leq n$. Recall that p is the probability that a patient will not show up. Let $q = 1 - p$ be the probability that a patient will actually show up and attend the appointment.

Let $b(m, q, k)$ be the probability that a binomial (m, q) random variable takes a value equal to k , i.e., $b(m, q, k) = \binom{m}{k} q^k (1 - q)^{m-k}$. Let $B^j(\bar{s}_t) = B^j(s_1, s_2, \dots, s_t)$ denote the probability of a backlog of j customers at the end of slot t , given that s_1, s_2, \dots, s_t customers have been assigned to slot $1, 2, \dots, t$ respectively, $1 \leq t \leq n$. As a convention, let $\bar{s}_0 = 0$ and

$B^0(\bar{s}_0) = 1$. Then $B^j(\bar{s}_t)$ can be expressed recursively as follows:

$$B^j(\bar{s}_t) = \begin{cases} B^0(\bar{s}_{t-1})[b(s_t, q, 0) + b(s_t, q, 1)] + B^1(\bar{s}_{t-1})b(s_t, q, 0) & \text{for } j = 0 \\ \sum_{i=0}^{l(\bar{s}_{t-1})} B^i(\bar{s}_{t-1})b(s_t, q, j - i + 1) & \text{for } 1 \leq j \leq l(\bar{s}_t) \end{cases}$$

where $l(\bar{s}_t) = \sum_{i=1}^t s_i - t$ is the maximum possible backlog at the end of slot t , $1 \leq t \leq n$.

Let $W(s, k)$ denote the total expected waiting time of s customers, who are scheduled to arrive in the same given time slot, assuming that there is already a backlog of k customers at the beginning of that slot. Then

$$W(s, k) = q \sum_{i=1}^s \sum_{j=0}^{i-1} (k + j)b(i - 1, q, j)$$

and the aggregated expected customers' waiting time under schedule \bar{s} is

$$W(\bar{s}) = \sum_{t=1}^n \sum_{j=0}^{l(\bar{s}_{t-1})} B^j(\bar{s}_{t-1})W(s_t, j).$$

If $I(\bar{s})$ denotes the total expected number of idle slots among slots $1, \dots, n$, then

$$I(\bar{s}) = \sum_{t=1}^n B^0(\bar{s}_{t-1})b(s_t, q, 0).$$

Let $O(\bar{s})$ denote the expected number of overtime slots, then

$$O(\bar{s}) = \sum_{j=0}^{l(\bar{s}_n)} j B^j(\bar{s}_n).$$

The objective is to find an optimal schedule $\bar{s}^* = (s_1^*, \dots, s_n^*)$ that minimizes the total expected cost $V^* = \min_{\bar{s}} \{V(\bar{s}) = c_I I(\bar{s}) + wW(\bar{s}) + c_O O(\bar{s})\}$. Equivalently, $\bar{s}^* = \arg \min_{\bar{s}} \{V(\bar{s}) = I(\bar{s}) + wW(\bar{s}) + c_O O(\bar{s})\}$, by normalizing the objective function with respect to c_I , i.e., $c_I = 1$.

Lemma 1. *In an optimal schedule every slot is assigned with at least one customer.*

Algorithm 1 Optimal Schedule

```
procedure OPTIMAL( $n, p, c_I, c_O, w$ )
  stop  $\leftarrow$  0
   $m \leftarrow n$   $\triangleright m =$  number of customers to schedule
  (schedule, cost)  $\leftarrow$  BEST( $m, n, p, c_I, c_O, w$ )  $\triangleright$  Optimal schedule for fixed  $m$ 
  repeat
     $m \leftarrow m + 1$ 
    (schedule1, cost1)  $\leftarrow$  BEST( $m, n, p, c_I, c_O, w$ )
    if cost1 < cost then
      schedule  $\leftarrow$  schedule1
      cost  $\leftarrow$  cost1
    else
      stop  $\leftarrow$  1  $\triangleright$  the cost is convex in  $m$ 
    end if
  until stop = 1
  return schedule
end procedure
```

Algorithm 2 Optimal Schedule for a fixed number of customers to be scheduled

```
procedure BEST( $m, n, p, c_I, c_O, w$ )
  all_schedules  $\leftarrow$  combinations( $n, m$ )  $\triangleright$  All possible ways to allocate  $m$  patients in  $n$ 
 $\triangleright$  slots, with each slot being assigned with at least one patient
  cost  $\leftarrow$   $10^{10}$ 
  for every schedule1 in all_schedules do
    cost1  $\leftarrow$  COST(schedule1,  $p, c_I, c_O, w$ )
    if cost1 < cost then
      schedule  $\leftarrow$  schedule1
      cost  $\leftarrow$  cost1
    end if
  end for
  return (schedule, cost)
end procedure
```
