Omega xxx (xxxx) xxx

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice ☆

Alex Kuiper [a,*], Jeroen de Mast [b], Michel Mandjes [c]

[a] Department of Operations Management, IBIS UvA, University of Amsterdam, P.O. Box 15953, 1001 NL Amsterdam, the Netherlands
[b] Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
[c] Department of Mathematics, Korteweg-de Vries Institute, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, the Netherlands

ARTICLE INFO

Article history:
Received 11 October 2018
Accepted 17 September 2019
Available online xxx

Keywords:
Appointment scheduling
Healthcare operations management
Case-study research
Process variability

ABSTRACT

Operations research and management science have produced many algorithms or rules for appointment scheduling, approaching that task as a mathematical optimization problem. It is, however, not sufficiently clear to what extent such problem definitions capture the objectives and limitations of appointment scheduling in real healthcare applications. This paper aims to reconstruct the structure of the problem faced by outpatient clinics by applying a multiple-case-study approach, based on a research model developed from operations-management theory, followed up by workshops. This study therefore breaks new ground by linking the problem of appointment scheduling as rendered by the operations-research literature to theory in operations management and practice in the field.

The study shows that the context of appointment scheduling has changed substantially compared to the setting that the operations-research literature has largely assumed. Economic assumptions appear unwarranted in practice, and where the literature describes the service process as repetitive with only limited variety and customization, practice turns out to be excessively complex.

Especially in situations of high complexity and uncertainty, approaches based on process flexibility and variability reduction appear more promising than mathematical optimization. Reducing the number of service varieties and operating at a 'lower' level of utilization, i.e., looser schedules, are promising practical suggestions for improving the performance of outpatient clinics.

1. Introduction

Synchronizing capacity with demand is one of the central challenges in the management of healthcare operations, especially since both are subject to variability and uncertainty. Outpatient clinics generally manage demand by scheduling consultations and treatments as appointments. In such clinics, clinicians work in sessions of a few hours, which are subdivided into slots, and patients are allotted to a slot with a scheduled begin time (the *appointment time*). Scheduling demand evens out the arrivals of patients over a session and this reduces the variability in demand. Consequently, demand can be synchronized more efficiently with the availability of clinicians and other resources such as facilities and support staff.

Even in scheduled operations, however, there are sources of variability and uncertainty that make a perfect synchronization impossible. For example, the *service time*, that is the realized time needed for a consultation or a treatment, could be longer or shorter than the planned duration of the slot. Also, patients may show up late (*tardiness*) or not at all (a *no-show*). Consequently, it commonly occurs that patients see the clinician later than the appointment time, which results in *waiting time* for the patient. It also happens that the clinician waits for a next patient to arrive and thus faces *idle time*.

This paper studies the problem of designing a suitable policy for scheduling appointments. On the one hand, appointments should be set up such that excessive waiting times for patients are avoided, as these are an important determinant of the perceived service quality and satisfaction [3,32]. On the other hand, the scheduling approach should maximize the utilization of clinicians, staff and facilities by avoiding idle time. Utilization is an important factor in the unit-costs of delivered care, and in addition, it is a factor in the total capacity of the service in question. Higher utilization optimizes patient throughput, and thus improves admission times, appointment delays and availability of care [24,67]. Appointment scheduling, therefore, directly impacts the service quality, cost-efficiency and capacity of a substantial part of healthcare services.

☆ This manuscript was processed by Associate Editor AE J. Rosenberger.

* Corresponding author.
E-mail addresses: a.kuiper@uva.nl (A. Kuiper), jdemast@uwaterloo.ca (J. de Mast), m.r.h.mandjes@uva.nl (M. Mandjes).

https://doi.org/10.1016/j.omega.2019.102122
0305-0483/

Operations research and management science have produced many analytical studies that propose algorithms or rules for scheduling based on queueing theory or simulation. Comprehensive literature reviews are given in Cayirli and Veral [11], Mondschein and Weintraub [49], Gupta and Denton [25], and Ahmadi-Javid et al. [2]. Such studies generally treat the task of appointment scheduling as a mathematical optimization problem, for example along the following lines:

- The optimality of schedules is framed in terms of a simple objective function, which is typically a weighted average of expected idle times for clinicians and waiting times for patients.
- The main source of variability that the scheduling strategy should buffer against, is the variance of the service times.
- The permissible solutions are limited to determining optimal numbers of patients scheduled in each slot, and possibly, the begin and end times of the slots.

Such mathematical studies help in building insight into the complex dynamics of the behavior of appointment systems, and they have added refined mathematical machinery to approach such problems. They focus more on the development of mathematical theory, however, than on the empirical study of operations-management issues. Consequently, it is not sufficiently clear to what extent the problem rendering in the operations-research literature captures the objectives and limitations of appointment scheduling in real healthcare applications.

In this study we aim to reconstruct the structure of the problem of appointment scheduling in outpatient clinics — Reasoning from theory in operations management and the objectives and constraints in clinics, how should the problem be defined? We are also interested whether current practices in outpatient clinics are congruent with the rendering and the solutions offered in the operations-research literature and to discover the reasons for discrepancies. The questions build on the findings of De Snoo et al. [18], who conclude from an empirical study that there is more to scheduling than mathematically solving a well-defined problem. The questions echo the appeal in Ahmadi-Javid et al. [2] for a comparison of theory with practice by means of case-study research.

We contribute a multiple-case study, where we conducted interviews in ten outpatient clinics. Case-study research is a powerful approach for exploring an area, identifying the key issues and essential themes to be taken into account in more analytical studies [6,37,56,65]. The interviews focus on key elements needed to understand the problem of appointment scheduling. Our strategy for identifying these key elements is to follow Ackoff's model of the general structure of problems in operations research [1]. This model guides our systematic review of the literature, and it is the basis from which we designed the interviews. The interviews revealed major issues in the frameworks found in the literature, especially concerning the goals of scheduling in clinics. We followed up the interviews with workshops, conducted with clinicians and staff, to clarify these issues.

We present in the next section our review of the literature on appointment scheduling in healthcare. This review results in our research model, which postulates how current literature renders the problem of appointment scheduling. The section also presents our research design and the cases that we selected. Section 3 presents the analysis of the cases and discusses interesting findings. As the results revealed some important unclarities in the goals of scheduling, we followed up the interviews by workshops, which we describe in Section 4. Section 5 concludes by presenting our research findings and providing directions for developing the theory on appointment scheduling as well as practical implications by highlighting a number of paths for improving the performance of scheduling in outpatient clinics.

2. Theory and methods

The research questions address the structure of the problem of appointment scheduling. As point of departure, we take Ackoff's general conceptualization of the structure of problems in operations research, as presented in Ackoff and Vergara [1] and elsewhere. According to Ackoff, the elements of a problem are:

1. The courses of action available to the owner of the problem;
2. Uncontrollable variables in the environment;
3. The outcome, which is the result of the courses of action and also of the uncontrollable environmental variables;
4. Positive or negative value attached to possible outcomes by the problem owner;
5. Constraints, to which the courses of action are subject.

A solution to a problem, then, boils down to the problem owner choosing a course of action within the given constraints, which, despite the effects of uncontrollable variables, results in a positively valued outcome.

We studied the literature on appointment scheduling from the perspective of these five elements, identifying what potential courses of action are available in designing a scheduling approach, to what sort of constraints they are subject, and what uncontrollable variables are taken into consideration. In addition, we identified what outcome characteristics of scheduling approaches are considered important in the literature, and how the various outcomes are valued. We summarized the answers to these questions found in the appointment-scheduling literature in the elements *E1* through *E5* (see Fig. 1). These elements and their structure are our research model. It articulates as a conjecture the conceptualization of the problem of appointment scheduling in the literature, and it is the conjecture that we aim to study in the multiple-case study. Below, in Section 2.1, we discuss the literature that we studied and build the elements *E1* through *E5* of the research model. Section 2.2 explains how the research model guided the design of the multiple-case study.

2.1. Theoretical development

Designing an appointment schedule would be straightforward if patients showed up on time, service times were constant or perfectly predictable, and no-shows, walk-ins, cancelations and other disruptions did not occur. The challenge is to design schedules to handle such variability as well as possible. Appointment scheduling is therefore an instance of the more general problem of dealing with process variability.

The literature on appointment scheduling takes into account various sources of process variability, either related to the environment (what needs to be scheduled?), such as number of clients and various types of patients, or to the execution of the schedule (variability in the service delivery process that prevents the schedule from being executed as planned; see [18]), such as absence of clinicians, breakdown of equipment and service-time variability. Fairly comprehensive lists can already be found in the earliest works, such as Welch and Bailey [62] and Fetter and Thomson [21].

One of the prominent sources is the variability in the service times. For simplicity, literature usually assumes that the service times are independent and identically distributed, which is an assumption that is contradicted by various empirical studies (e.g., [5,50]) as clinicians tend to increase their pace when patients are waiting. Many studies incorporate uncertainty in demand in the form of no-shows and walk-ins (e.g., [12,13,46,66,69]). Klassen and Yoogalingam [41] investigate the effects of tardiness of clinicians and interruptions, and also tardiness of patients is sometimes considered as a source of variability.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

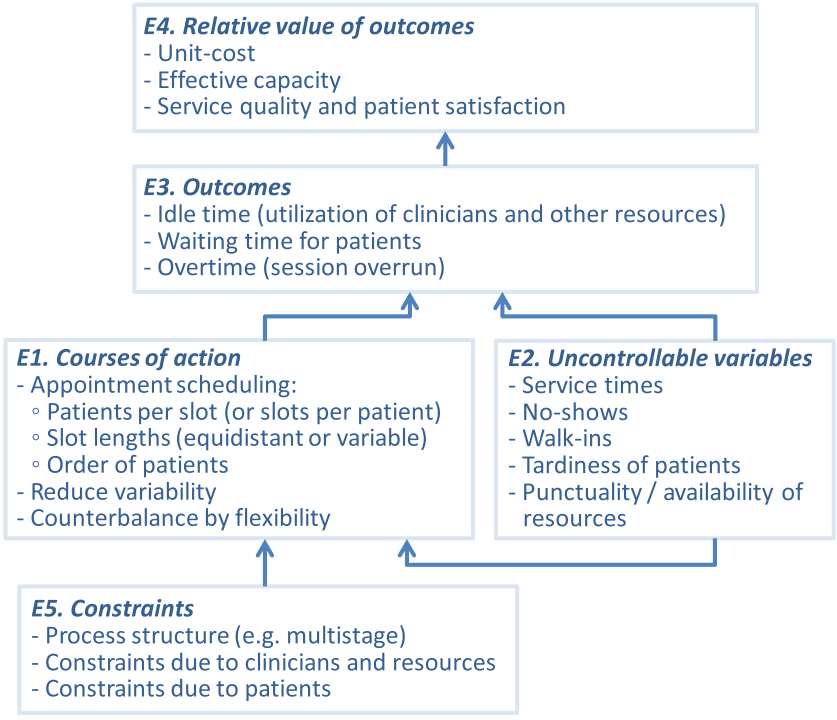

Fig. 1. Proposed research model.

In coping with process variability, Hopp and Spearman [31] suggest that a sensible first step is to try to reduce it. Appointment scheduling itself is a variability-reduction tactic, as it spreads demand more evenly over a session. In addition, facilities reduce uncertainty even further by bringing down no-shows and last-minute cancelations by employing reminders or sanctions [7,35], which is increasingly easy to employ by the growing availability of technology.

After variability has been reduced, Hopp and Spearman [31] suggest as a second step that variability be counterbalanced by flexibility of patients and resources (see also [16]). The idea is to reduce the negative effects of variability by exploiting that the type and timing of some demand and some tasks may be flexible, and furthermore that there may be flexibility in the availability of resources. The negative impact of peak loads, for example, is sometimes reduced as clinicians stretch their working day or shrink lunch time, and the potential loss of unanticipated idle time may be avoided as clinicians, rather than sitting idle, switch to administrative work or other pending tasks. The idea of flexible production was popularized following the success of Toyota, which systematically pursued it. Production flexibility has been studied systematically in manufacturing (see [15], and references therein) and services (e.g., [40]). Of particular interest for our study are volume flexibility [33], the ability to accommodate variability in demand, and process flexibility [16], the ability to accommodate disruptions and changes in the service process. This study of the literature motivates the first elements of our research model:

E1. Courses of action

Possible strategies for dealing with process variability in outpatient clinics are:

- Appointment scheduling;
- Exploitation of flexibility of resources and patients;
- Reduction of process variability by other means than appointment scheduling.

E2. Uncontrollable environmental variables

Sources of process variability that affect capacity and demand in outpatient clinics include:

- Variation in the service times
- Random no-shows and cancellations
- Random walk-ins
- Interruptions and tardiness of clinicians
- Unpunctuality of patients

After variability has been reduced by appointment scheduling, counterbalancing by flexibility, and other approaches, the *Variability Buffering Law* of Hopp and Spearman [31] predicts that the remaining variability will be absorbed by a combination of three buffers:

- A queue of patients waiting to get served (waiting time).
- Unutilized capacity of the clinicians (idle time).
- An inventory of finished products, built up in advance as a buffer to absorb peaks in demand.

The third is rarely an option, however, for the type of services that we consider, because products in our setting are treatments and consultations, and production usually cannot proceed until patient and clinician come together. Scheduling algorithms proposed in the operations-research literature are typically designed to minimize the first two buffers – the expected waiting and idle times. These dual objectives are partly a trade-off, as beyond some point one can only improve one at the expense of the other. Scheduling algorithms typically combine both objectives by optimizing the weighted average of the expected waiting and idle time, where the weight expresses the relative importance of waiting versus idle time. This notion that appointment scheduling aims to strike a balance between waiting time and idle time has pervaded the literature from the beginning [21,62], and is generally unchallenged. Especially in early works, minimizing idle time for the clinician has been the overriding objective — for example: “In practice, the requirement that the consultant be kept fully occupied is usually regarded as an over-riding consideration: large queues of patients

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

are often allowed to build up in order to avoid the possibility of the consultant ever having to wait for a patient." [5].

Many approaches, such as Fries and Marathe [22], White et al. [63], and Cayirli et al. [13], further consider *overtime*, the time that a session overruns the scheduled end time. Note that overtime is largely a by-product of waiting times building up over a session. To be exact, and taking the session-end time to be the end point of the last slot, it can be shown mathematically that a session's overtime is the sum of the last patient's waiting time and the duration of the last appointment minus its scheduled duration. Therefore, this is another way of capturing the trade-off between a heavily loaded and congested session, leading to waiting time for patients and overtime, versus an under-loaded session, implying idle time for clinicians.

Alternatively, some approaches minimize a weighted average of the expected waiting time and the expected duration of the session (e.g., [27,61]). The latter equals the sum of expected service times plus the sum of expected idle times, and minimizing it is equivalent to minimizing waiting and idle time [49]. This review leads to the third element of our research model:

E3. Outcomes

The performance of a schedule is a combination of waiting time for patients, idle time for clinicians, and resulting losses such as overtime. These outcomes are partly a matter of a trade-off, and the desired balance can be expressed in a weight.

The economic ramifications of idle time, for clinicians as well as other resources, are not made explicit in literature, but idle time is generally described as lost capacity. Its economic implications, then, would be its effect on unit-cost and the effective capacity of the service. Waiting times create dissatisfaction for patients and degrade the quality of service [3,32]. Congestion may also lead to scheduling conflicts in other processes, as resources and patients are held up longer than anticipated. The fourth element of the research model is:

E4. Relative value of possible outcomes to the problem owner

The economic implications of a schedule's performance are a combination of perceived service quality (affected by waiting time), disruptions in other processes (due to overtime) and unit-cost and effective capacity (affected by idle time).

Other operational restrictions and preferences could complicate the scheduling task, such as the structure of the service process. Literature on appointment scheduling predominantly considers a single-server, single-stage process [19,63]. Most clinics are run with multiple clinicians, but as long as specific patients are tied to a specific clinician (that is, clinicians are not interchangeable), each clinician essentially has her own schedule and the single-server model is appropriate. There are only a few studies that examine multi-server models, two of them are Zacharias and Pinedo [68] and Soltani et al. [54]. Also other resources, such as facilities and equipment, are often shared among clinicians [63], which could create scarcity and this is likely to create complications for scheduling.

A multi-stage process, or combination appointment, means that a patient receives more services than a single consultation or treatment, for example, an X-ray or blood sample followed by a consult. Rising et al. [50], Swisher et al. [57], White et al. [63], Salzarulo et al. [53] and Kuiper and Mandjes [43] study such multi-stage processes, where scheduling in one stage must be coordinated with demand management in the other stages.

The problem of appointment scheduling could further be complicated by heterogeneity in the patient population with respect to the expected duration of service time, such as the difference between new and return patients, or patients with different diagnoses [12,52,53,63]. Preferences of patients for certain days or certain slots could be taken into account, as well as preferences of clinicians [2]. We summarize these considerations in the fifth element of the research model.

E5. Constraints

The scheduling task may be complicated by operational restrictions:

- Restrictions brought about by the structure of the service process.
- Restrictions brought about by scarcity or preferences of clinicians and other resources.
- Restrictions brought about by characteristics and preferences of patients.

The five elements are summarized in the research model in Fig. 1. The arrows indicate the structure implied by Ackoff's model of the structure of problems, as explained above.

We make a final refinement of the research model. Where *E1. Courses of Action* enumerates the three general approaches for dealing with variability in outpatient clinics, the last part of this section elaborates in more detail the courses of action available in appointment scheduling.

A session divided in slots has always been an essential notion of appointment scheduling. The earliest systematic studies, by Bailey [5], Welch and Bailey [62], and Fetter and Thompson [21], established that the slot lengths should be based on the mean service time. These authors introduced the notion that the schedule should take process variability into account. The widely used *Bailey-Welch rule* works with slots of equal lengths, based on the mean service time, and with two or more patients scheduled in the first slot in order to build up a buffer of work that absorbs variability due to no-shows or shorter-than-average service times. Fries and Marathe [22], Liao et al. [45], Vanden Bosch et al. [59], and Zacharias and Pinedo [66] generalize such heuristics by determining the optimal number of patients to be scheduled for each slot (the so-called *block size*). The decision parameter in the scheduling task, therefore, is the number of patients to be scheduled in each of the slots. This approach could alternatively be framed as deciding how many slots to allot to a patient.

Charnetski [14] proposes to set slot lengths equal to the mean service time plus a multiple of the standard deviation in service times. Thus, the decision parameter is no longer the number of patients per slot, but the length of the slots. Ho and Lau [29,30] and Yang et al. [64] propose and compare similar scheduling rules, which are further enriched in Cayirli et al. [13] by implementing no-shows and walk-ins. A more flexible approach is to drop the constraint that all slots should be of equal length (which is an optimization in one parameter), and allow slots to be of variable length (an optimization in as many parameters as there are slots in a session). Studies such as Wang [61], Robinson and Chen [51], Kaandorp and Koole [36], Hassin and Mendel [26] and Kuiper et al. [42] find that it is often optimal to schedule shorter slots in the beginning and end of a session, and longer slots in the middle — so-called *dome rules* for scheduling.

Optimization problems with that many degrees of freedom are typically hard to solve, and efforts in the operations-research literature have concentrated on tactics to make the optimization problem manageable. Early attempts, such as Soriano [55] study the steady-state behavior of the queueing system, which is still used as an approach to provide further insight into appointment scheduling [44]. Also many papers analyze the problem assuming tractable service-time distributions, such as the exponential [26,27,36] or phase-type distributions [42,58,59,61]. Others study worst-case approximations [47] or discretized versions of the problem [69]. Even then, the optimization problem is

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

computationally hard [51], and recent approaches focus on approximation algorithms [8,13] and integer programming approaches [34].

Besides the appointment times, the order in which patients are scheduled could be optimized. Klassen and Rohleder [38], Cayirli et al. [12], Denton et al. [17] and White et al. [63] investigate the effects of sequencing policies based on the variance in service times of various types of patients, and generally find that scheduling low-variance patients early in the session minimizes waiting times, idle times and overtime. Cayirli et al. [12] found that sequencing decisions have more impact on performance than rules for the slot lengths.

The last addition to our research model, then, is a detailed elaboration of the courses of action available in appointment scheduling:

E1a. Courses of action in appointment scheduling

The possible courses of action in designing a schedule are combinations of:

- The number of patients scheduled in each slot (ranging from zero to a few). Examples: double-book the first slot by two patients to build up a buffer, or leave some slots empty to absorb overruns.
- The number of slots assigned to patients. Examples: allot three slots to new patients and one slot to return patients.
- The lengths of the slots, either under the constraint that all slot lengths are equal, or allowing variable slot lengths.
- The order in which patient (types) are scheduled.

2.2. Multiple case-study design

The above shows that substantial mathematical methodology has been developed to solve certain variants of the scheduling problem. Relatively little effort has been made to underpin or motivate that the assumptions made in defining the optimization problem are valid in practice. We address this objective by means of a multiple-case-study design. The stated objective is a combination of exploring the key issues in appointment scheduling and identifying critical variables and factors. Case-study research is a suitable approach for such questions [6,56,65].

The unit of analysis in our study is a single outpatient clinic or department, offering one or more health services, provided by multiple clinicians. Eisenhardt [20] and Barratt et al. [6] recommend four to ten cases for case studies. We selected 10 clinics, operating in 6 hospitals in The Netherlands, and covering a variety of specialties (see Table 1). In the final choice for the clinics we tried to find contrasting instances —that is, we aimed for theoretical replication rather than literal replication, in the terminology of [60,65]. The clinics represent contrasting variety in these aspects:

- Typical duration of a single slot (from very brief 5 minutes' slots to a clinic where a typical slot is 90 min).
- Modern operations driven by computerized workflow-management software, and more traditional operations.
- Large clinics (more than 20 clinicians) to small clinics (2 clinicians).
- Treatments and consultations that are relatively routine versus clinics that have to deal with substantial variability in service times.

In each of the cases we collected information by means of a structured interview following a set protocol. The questions in the interview address the five elements *E1* through *E5* of the research model, and elaborate them in more detail from two angles:

- *Descriptive angle*: How does the clinic *factually* structure the problem of appointment scheduling?
- *Prescriptive angle*: How *should* the clinic structure the problem of appointment scheduling?

Interviews typically lasted 60 to 75 min. Interviewees were medical secretaries, clinic managers, doctor's assistants, and team leaders. The first four interviews were done by two authors and, on the basis of the results, minor modifications were made in the questionnaire. The last six interviews were done by a single author. Where available, we collected quantitative data such as waiting times, no-show rates and services times. Before each interview, the interviewer always did a guided walk-through of the actual process.

We first analyzed each case separately, trying to understand the problem of appointment scheduling in each particular clinic (within-case analysis, [20]). These analyses were guided by the research model in Fig. 1, and resulted in case-specific characterizations in terms of the elements *E1* through *E5*. Early versions of the within-case analyses were presented to the respondents and their feedback was incorporated in subsequent versions of the analyses. Once we had obtained consensus among the authors and with the respondents about the within-case analyses, we compared and contrasted the per-case findings across clinics, looking for patterns across cases (cross-case analysis). Points of departure were again the research questions and their elaboration in the detailed questions for the interview. The findings are presented and discussed in the next section.

From the multiple-case studies we learned that the conceptualization of the economic implications of scheduling, as framed in the element *E4* of the research model, is debatable. Given the importance of this part of the scheduling problem, we conducted a follow-up study to clarify what performance aspects of scheduling are important to clinics. The design and execution of these follow-up studies are explained in Section 4.

3. Analysis and findings of the multiple-case study

We focus on the cross-case analysis, where we aim to identify patterns that hold across cases. The section is organized around the elements of the research model, which we discuss starting from the outcomes and objectives of scheduling (*E3* and *E4*) via the uncontrollable variables and constraints (*E2* and *E5*) to the appointment scheduling practices (*E1*).

3.1. Outcomes of scheduling practices and their valuation (E3 and E4)

The interviews explore the economic consequences of poor scheduling for the clinic. Also, interviewees are asked what they consider the relative importance of waiting time for patients versus idle time for clinicians and facilities and overtime for sessions.

Long waiting times occur generally, as do session overruns. The detrimental consequences of these for patients and for the clinic are obvious to respondents. Idle time for the clinician, however, is not seen as an important issue. One reason is that most clinics use tight schedules, and consequently, utilizations are high and idle time is rare. This comes, of course, at the expense of long waiting times for patients and frequent session overruns. The other reason is that clinicians have sufficient substitute work that they can do when they wait for a next patient, such as administrative work and management tasks. Of notable interest is that almost all clinics offer non-visit (e-consult) care, where the patient is at home and communicates with the clinician by telephone [2]. These non-visit consultations are not precisely scheduled, and clinicians are flexible when they do them, making them one of the options for putting idle time to productive use.

There is almost no awareness that there is a trade-off between waiting time for patients and idle time for clinicians. Respondents

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

Table 1
Overview of cases.

Case	Specialty	Patients per session	Sessions per week	Number of clinicians	Characteristics of the clinic		
					Appointment schedule	Single or multi-stage	Other details
1	Internal medicine	20	120	27	Fixed slot lengths of 10 min for returns; new patients are assigned 3 slots.	Half single and half multi-stage process	General hospital with modern IT system; it sends notifications based on waiting-time estimations.
2	Orthopedics	15	3	2	Fixed slot lengths of 8 min.	Often multi and sometimes single-stage process	Small outpost clinic of a general hospital, which only handles consultations.
3	Endoscopy	5	30	11	Fixed slots per type of treatment, 30, 45 or 60 min. Schedule is overbooked to absorb cancellations and no-shows.	Often single sometime multi-stage process	Clinic in a general hospital with a complex structure: expensive equipment to be shared and X-rays to be made.
4	Psychiatry	3	40	20	Slot lengths fixed at 1 h for psychiatry. For geriatrics 1.5 h (new) and 0.5 h (return patients).	Psychiatry: single stage process; geriatrics: often multi and sometimes single-stage process	A psychiatry clinic, combined with geriatry in a general hospital. Clinicians manage their own schedules.
5	Orthopedics	20	80	26	Fixed slots with some overbooking. Also empty slots to relax the schedule. Secretaries intervene if consults take too long.	Often multi and sometimes single-stage process	Large clinic in a general hospital. Complex structure due to constraints on available rooms and specialists. Quarterly evaluation of the schedule.
6	Otorhino-laryngology	25	40	10	Fixed slots of 5 min., a double slot is assigned to difficult cases. Some slots double-booked.	Often single sometime multi-stage process	Large clinic in a general hospital operating at multiple locations.
7	Ophthal-mology	22	20	7	Fixed slots of 10 min. and 30 mins for examinations. Manay slots are double-booked.	Often single sometime multi-stage process	Small clinic in a general hospital, unavailability of resources is a primary concern.
8	Ophthal-mology	12	44	30	Fixed slots of 15 min. (return) or 30 min. (new). Some slots are double-booked.	Often multi and sometimes single-stage process	Clinic in an academic hospital, most visits are multi-stage with a pre-examination by optometricians.
9	Neurology	16	5	8	Fixed slots are long: 90 min.	Often single sometime multi-stage process	Clinic in an academic hospital. Sessions devoted to specific sub-specialties, sometimes multi-stage.
10	Orthopedics	25	13	3	Fixed slots, 5 min. (return) or 10 min. (new). Difficult cases are assigned a double slot. Some slots double-booked.	Often single sometime multi-stage process	Small clinic in a general hospital. Sometimes an X-ray required making it multi-stage.

find the notion difficult to understand when we explain it, and the trade-off is no consideration in the way schedules are created in practice. Our overall impression is one of very limited insight into the goals of appointment scheduling and their dependencies. When asked, respondents tend to resort to politically correct answers — claiming that the patient's interests should come first and ignoring all other concerns.

3.2. Uncontrollable environmental variables (E2)

Variability in service times appears the most important source of variability in appointment scheduling. Respondents frequently mentioned that there are big differences between clinicians, where some clinicians habitually overrun the schedule, while other clinicians are quite consistent in keeping to the schedule. Besides the clinician's behavior, characteristics of patients (such as age and language difficulties) and findings in the examinations (for example, the number of polyps that are found) are given as causes of variability in service times.

No-shows, generally incorporated in scheduling algorithms, appear a minor problem in the clinics in our sample. The studied clinics do not take no-shows into account in scheduling, but instead, generally invest in prevention tactics such as reminders and sanctions, and these preventive measures effectively reduce the occurrence of no-shows to the range of 2–8%. This reflects efforts in recent years in the Dutch healthcare sector to reduce no-shows, and the situation in other countries may be different ([35], for instance, report figures as high as 42% no-shows in some specialties in the US). Studies such as Glowacka et al. [23] confirm that countermeasures may alleviate the no-show problem, but are unlikely to eliminate it completely.

Walk-ins are either handled by channeling them to a separate process, or by incorporating them in the scheduling approach (for example, by leaving several slots open). Tardiness of patients is seen as a minor issue, and when it happens, this is easy to handle as the clinics that we visited habitually overrun the schedule and sufficient patients are waiting so that another patient can change places with the belated patient.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

3.3. Constraints (E5)

Respondents were asked about the structure of the process and constraints that they take into account. A single-server model is realistic for almost all cases. All clinics have multiple clinicians, but patients are assigned to a single clinician to ensure continuity of care (except for emergency situations). The process is therefore operated as a single-server process. In about half of the clinics in our sample operational constraints are mild, and scheduling is relatively simple. In the other half, however, restrictions and dependencies create a complex puzzle for the schedulers to solve. Complications are related to the structure of the service process, limited capacity of resources that are diversified, service varieties, and characteristics of patients and clinicians. We discuss these factors below.

3.3.1. Process structure

Multiple-stage visits and combination appointments are more common than single-stage visits. Consultations and treatments are generally combined with examinations in other departments. All of these stages are generally scheduled separately, except for walk-in services such as blood samples. Schedulers make an effort to schedule all appointments for a patient on a single day, which creates dependencies to be taken into account.

3.3.2. Limited capacity of diversified resources

Some clinicians, rooms and equipment are generic, but we encountered many examples where clinicians have specialties within their field, certain rooms have facilities that other rooms do not have, and equipment is diversified. This diversification means that the interchangeability of resources is limited, which creates scarcities that must be taken into account in scheduling. For example, multiple patients who are likely to need the one room with X-ray equipment are not scheduled close to each other.

3.3.3. Service varieties

Consultations and treatments are offered in varieties. An omnipresent distinction is between a consultation for new patients versus return patients, and also based on their diagnoses and treatment plans, patients visit the clinic for different services. Schedulers take these varieties into account when they make appointments.

3.3.4. Patient and clinician characteristics

Besides service varieties, schedulers sometimes take foreseeable differentiators into account, such as specific clinicians who work slower than others, or certain patient characteristics (language difficulties for example) that make it likely that more than standard time is needed. Many clinics consider preferences of patients or clinicians for certain days or parts of the day.

3.4. Courses of action (E1)

The interview questions explore whether clinics make efforts to reduce variability, or to exploit flexibility to absorb it. Tactics for reducing process variability are exploited systematically for the problem of no-shows. Almost all clinics apply sanctions and reminders, and stimulate responsible behavior by giving patients ownership over making the appointment. These measures are so effective that no-shows are seen as a non-problem in all clinics in our sample.

Efforts to reduce other sources of variability are ad hoc, basic and crude. In one clinic, secretaries can intervene in long consultations and cut them short, thus reducing variability in service times. Some clinics reduce variability due to walk-ins by diverting these to a separate process. It is likely that scheduling performance could be improved by exploring possibilities for reducing variability in service times. Since many clinics observe substantial differences between clinicians, a credible line of approach would be to explore whether slower clinicians can be coached to learn from the faster clinicians.

Tactics for exploiting flexibility of resources and patients to counterbalance variability are used generally, albeit in an improvised manner, for example as staff works late when a session runs late. When faced with idle time, clinicians have a range of substitute tasks that they can flexibly switch to, thus preventing that idle time is lost capacity. Substitute tasks include administrative work, management tasks, and the before mentioned non-visit consultations done by telephone. For patients, waiting time is generally lost time, but some clinics make an effort to put it to use or at least make it pleasant by offering magazines, WIFI or shopping services.

3.5. Courses of action in appointment scheduling (E1a)

All clinics in our sample work with sessions divided in slots of equal length, often differentiating between new and return patients. These fixed lengths seem to have emerged through practice, and do not appear to be the result of a deliberate optimization attempt. The central idea in the operations-research literature, to optimize slot lengths in order to achieve a good balance between waiting and idle time, is not used in any of the clinics in our sample. As a matter of fact, in none of the clinics are the slot lengths determined based on data (such as measured service times, no-show rates, or walk-in occurrences). Furthermore, slots of variable lengths, as in the dome rule (shorter slots in the beginning and end, and longer slots in the middle of the session), are nowhere used.

The sequence or order of appointments is dominated by operational constraints (as discussed in Section 3.3), and the impact of order on the effect of variability is nowhere considered. One notable exception is Case 5, a clinic that evaluates the performance on a quarterly basis and then reconsiders the sequencing practices. Currently they employ the approach to schedule a new patient after two return patients to reduce variability and excessive congestion that it could result in otherwise. Appointment schedules in which more than one patient is assigned to a slot are used only rarely in the clinics that we visited, but many clinics habitually keep some slots in a session open to absorb variability. Walk-ins are often added to appointment slots on an ad hoc basis.

All in all, there appears to be a substantial disconnect between scheduling practices and their rationales in the field, and conceptualizations and approaches in the literature. Interviewees are unfamiliar with tactics proposed in the operations-research literature and found them difficult to understand.

3.6. Idiosyncratic problems

The problem of appointment scheduling is not the same in all outpatient clinics, and the prominent role of idiosyncratic conditions in many of the cases should be noted and creates complications for the formulation of general theory. Case 4, for example, is a combined psychiatry and geriatrics clinic, whose mode of operation stands apart from the other clinics, and seems irreconcilable with the theoretical framework in Fig. 1. The essential difference is that in this clinic service times cannot be conceived as autonomous random variables, where a service time is determined by how long the clinician needs to complete the tasks implied by the treatment or consultation. Instead, patients pay for and are entitled to a certain amount of time, typically an hour in this clinic, and the consultation ends when this time is spent.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

In other cases we encountered less essential idiosyncrasies affecting the appointment-scheduling problem. Case 2, for example, is an outpost clinic, where appointment scheduling is greatly hampered by the traffic situation, which occasionally results in substantial lateness of the clinician.

4. Follow-up studies elaborating the goals of appointment scheduling

The dual goals of minimal idle and waiting time are combined in the operations-research literature by minimizing a weighted average of the two, where the weight reflects the desired balance. This rationale has largely gone unchallenged, one notable exception being Mondschein and Weintraub [49], who investigate the economics underlying objective functions, criticizing the assumption that demand is exogenous. Also Millhiser et al. [48], criticize the traditional approach optimizing ratios of expectations, and instead, consider probabilities of excessive waiting and overtime and policy targets for the number of patients.

The multiple-case study reveals a substantial disconnect between this rationale, which has dominated theoretical contributions since the beginning, and practice, which does not recognize nor apply it. Practitioners almost universally see idle time as a minor issue, there is no awareness that there is a trade-off to be made, and interviewees have no coherent idea about a suitable balance. Given the importance of this matter for understanding the problem of appointment scheduling, we study the goals of appointment scheduling in more detail.

4.1. Theoretical development

The scheduling literature writes about the cost of idle time without making it explicit. Trying to understand the cost of idle time in outpatient clinics, we elaborate three alternative hypotheses that we can study in the field. The left side of Fig. 2 (*S0. Baseline scenario*) represents a day at a clinic operating with a tight schedule. The realized service times, represented by the widths of rectangles, are longer on average than the slot lengths, and consequently, there is only minor idle time (at the end of the second slot), and the session overruns the schedule. After the session, there are other, non-scheduled tasks for the clinician, such as administration and non-visit consultations. These non-scheduled tasks are represented by ovals.

The right-hand side of Fig. 2 explores the situation that the clinic would operate with a looser schedule. The first scenario (*S1. Idle time is lost capacity*) represents the argument that literature seems to assume implicitly. By making the slots longer, the clinician has more idle time. Fewer patients can be scheduled in the same session (6 instead of 8), and consequently, the effective capacity is lower. Also, the daily fixed cost of the clinician is spread over fewer patients, and therefore, the unit- (per patient) cost is higher. This argument is generally valid in a factory, where a machine sitting idle implies lost capacity. For clinicians in outpatient clinics, however, idle time is not lost capacity, as idle time is put to effective use.

That is the second scenario on the right-hand side of Fig. 2 (*S2. Idle time used for substitute tasks*). Here, clinicians, when faced with idle time, switch to some of the non-scheduled tasks instead. Consequently, after the session, less non-scheduled work remains, and the session can be scheduled to last longer. The number of patients and the number of non-scheduled tasks done in scenario *S2* is the same as in the baseline scenario *S0*, and consequently, the effective capacity and unit-cost are also the same. If this scenario is realistic, it would generally be preferable to the tighter schedule in *S0*, since waiting time and overtime are shorter, while unit-cost and capacity are equal. While the multiple-case study suggests that the scenario is realistic at least to some extent for clinicians, it is plausible, however, that it is not so for other resources such as facilities, support staff and equipment, and instead, that for such resources idle time indeed implies lost capacity as in scenario *S1*.

Articulated by C.N. Parkinson in a humorous essay in 1955, Parkinson's law is the adage that "work expands so as to fill the time available for its completion." Inspired by this law, the last scenario in Fig. 2 (*S3. Idle time filled up*) hypothesizes that, when slots are scheduled looser, clinicians will adapt their pace accordingly. Thus, the consequence of longer slots is not that clinicians have idle time in which they switch to alternative tasks, but instead, that consultations are prolonged so as to fill up the available time, and service times go up on average. If scenario *S3* is realistic, longer slots increase unit-cost and decrease effective capacity of a clinic.

Almost all clinics in our sample work with a tight schedule, as in *S0*, with correspondingly long waiting times for patients and frequent session overruns. To examine what would happen if sessions were scheduled looser, we presented the three alternative scenarios *S1, S2* and *S3* to professionals working in outpatient clinics, including doctors, assistants, secretaries and clinical managers. We discuss these workshops below. Each workshop took approximately two hours in which we first made sure that they understood the main principles of appointment scheduling and

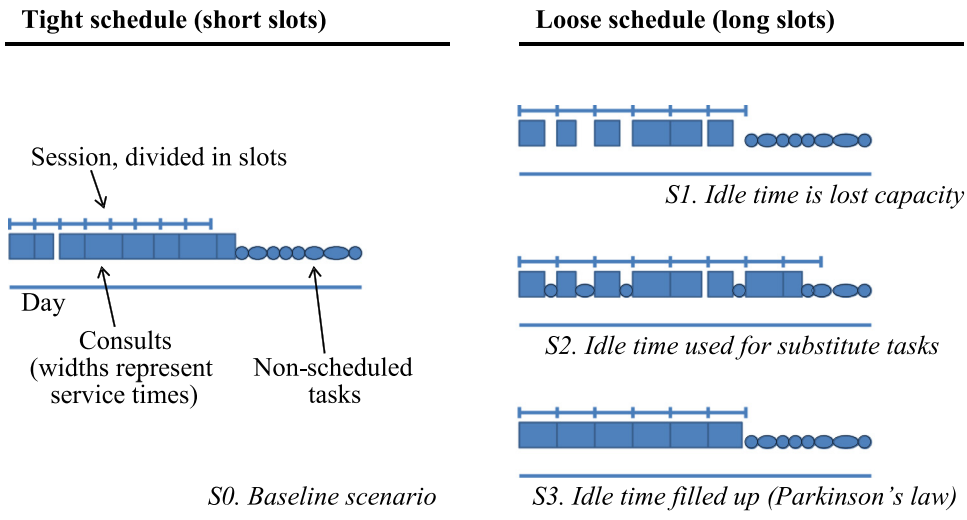

Fig. 2. Baseline scenario with a tight schedule (left) and three scenarios representing three alternative hypotheses of what would happen if the schedule were loosened.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

Table 2
Overview of the participants in the workshops.

Workshop	Participants
1	Two operations managers responsible for outpatient clinic policies
2	Doctor and head of an outpatient clinic, operations manager, and secretary
3	Doctor and head of an outpatient clinic, operations manager, two operations improvement specialists and secretary

then we discussed the different scenarios in detail for clinicians, support staff and facilities.

4.2. Discussion of the follow-up studies

Table 2 gives an overview of the workshops that we conducted and the professional positions of the participants. The protocol for the workshops starts with a brief explanation of a simplified version of the research model in Fig. 2 and the key findings from the interviews. Workshop participants were asked to criticize or corroborate our findings. In the second part of the workshops we explained the scenarios in Fig. 2. We asked participants to consider what would happen if a tight schedule (*S0* in Fig. 2) were loosened, and to what extent each of the scenarios *S1.*, *S2.* and *S3.* would be realistic in that case. We asked this question first focusing on clinicians, then focusing on facilities and equipment, and finally focusing on support staff. The workshops were conducted by two authors, with one author leading the discussion and the other author taking notes.

Participants convincingly confirmed the findings of the case study, and in particular that idle time of clinicians is not an important concern whereas session overruns create organizational challenges. For the management of hospitals, the perceived quality of service was mentioned as the most important consideration, as surveys repeatedly identified waiting time and the communication about it as major causes of dissatisfaction for patients. Participants emphasized that current practices are the result of historical factors, now well-entrenched in the organization.

Workshop participants were univocal in identifying that scenario *S1* (*Idle time is lost capacity*) is realistic for facilities and support staff, and that *S2* (*Idle time is used for substitute tasks*) and *S3* (*Idle time is filled up*) are realistic for clinicians. There is a strain between scenarios *S2* and *S3*, in that scheduling in longer slots could result in clinicians either using the extra time for substitute tasks, or to prolong the consultations as described by *S3*. There was disagreement among participants in assessing this strain. Some participants were worried that loosening the schedule would result in an overly relaxed attitude for clinicians. Other participants believed that this dilemma should be left to the professional discretion of clinicians, allowing them to prolong a consultation if warranted. Participants motivated this stance by noting that clinicians themselves are confronted with the consequences of poor judgment or an overly relaxed attitude.

From the workshops we concluded that indeed, idle time for clinicians is far less important than its prominent place in the scheduling literature suggests. We also noted that there is no strong reason to believe that current practices, historically grown, represent an economically good balance between the interests of the clinic and the patients.

5. Conclusions and managerial implications

5.1. General conclusions

Appointment scheduling in the ten clinics that we studied, is almost totally based on experience and practices that evolved over the years. None of the clinics had used any form of theory, data or formal method to design its scheduling practices. Moreover, there is no awareness of the concepts on which the literature on appointment scheduling is built, such as buffering to absorb variability and striking a balance between waiting and idle times. There is apparently an enormous gap between scheduling theory, which offers a rich edifice of formal mathematical optimization approaches, and practice, which deals with the problem based on experience and ad hoc improvisation. Although we recognize the finding of White et al. [63], that clinicians' intuition about managing capacity in clinics may differ substantially from best policies, we believe that there is more to the story than the simple conclusion that ignorance is the reason that formal approaches are not embraced.

In many clinics, formal approaches fail to satisfactorily capture the challenges of appointment scheduling. One reason is that oversimplified modeling of the process, as a repetition of identical and independent consultations, is defied by the complexity of many outpatient clinics, which face a large variety in service and patient characteristics and involve a multitude of constraints and demands to be considered, as discussed in Section 3.3. A second reason is that simple objective functions, with a precisely specified weight between waiting and idle time, fail to capture the ambiguity and messiness of the scheduling problem, where no-one has a clear and articulated answer to the question what the weight should be, and various stakeholders are likely to have different opinions about that issue. A third reason is that mathematical optimization approaches are limited to the static solution of calculating a schedule in advance, whereas it is likely that in cases of high uncertainty it is better to react flexibly to process variability when it happens during a session.

However, in some clinics, whose services have the characteristics of a high-volume and low-variety process, the structure of the problem may be reflected well in the assumptions of mathematical approaches. This touches on a conclusion of De Snoo et al. [18], who found that for problems such as scheduling, mathematical optimality is more important in situations of minor uncertainty and complexity, but that flexibility becomes more important when uncertainty and complexity are more substantial.

We find that the context of appointment scheduling has changed substantially compared to the setting that the operations-research literature largely assumes. Modern technology offers interesting opportunities to deal with variability by process flexibility. Idle time for the clinicians turns out to be less important than its role in the literature would lead us to suspect. This in turn puts the mature, traditional stream of research in this field in a different perspective. Our study suggests some directions for future research, and also some directions for immediate improvement in the field, which is described below.

5.2. Directions for future research

Our findings suggest two promising directions for future research.

5.2.1. Opportunities to exploit process flexibility

Where mathematical optimization is limited to determining slot lengths in advance such that the expected waiting and idle time are minimal, improving flexibility refers to enhancing the process's ability to deal with variability by reacting to it when it happens. In the clinics that we visited, process flexibility is

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

generally exploited in an informal and improvised manner, for example, when clinicians flexibly switch to other tasks when faced with idle time, or when a patient arriving early is swapped with a patient showing up late.

We propose that the performance can often be improved by elaborating such improvised behavior into deliberate policies, and testing and perfecting them in practice. Especially the further development of e-consults is likely to create opportunities for absorbing process variability by flexibility. Another direction is to extend slot lengths to deliberately include time to do patient related work directly, instead of postponing this work to after the session. In this way, increasing waiting times can be absorbed during the session by flexibly switching between consults and indirect patient-related tasks. Future research is needed to explore such opportunities systematically, and to formulate a framework and practical approaches. The literature on flexibility of service processes offers a promising starting point (see, for example, [28,39,40,70]).

5.2.2. Optimizing schedules by feedback adjustment

Mathematical optimization assumes a one-time calculation of a schedule's parameters, such as the lengths of the slots, which are then fixated and the basis for session schedules for a prolonged period of time. In feedback adjustment, to the contrary, the schedule's slot lengths are adjusted frequently — possibly even after every session — based on the discrepancy between the observed waiting and idle times, and their desired ratio. Feedback adjustment is used widely in process control (e.g., [10]) and control engineering (e.g., [4]).

The attractiveness of such approach is that it does not require complex process knowledge, nor the solution of an optimization problem. Instead, the process migrates automatically to an optimum in repeated adjustment cycles, and even when characteristics of the process change, the feedback adjustments steer the process automatically to a new optimum. To our knowledge, this sort of applications of feedback adjustment has not been studied before, and such research could bring such methodologies far beyond their traditional domains of applications.

5.3. Implications for practice

Our study suggests a number of directions for immediate improvement in practice, which we describe briefly here.

5.3.1. Simplify processes

The scheduling literature assumes minor service variety, that is, it is assumed that all appointments go through the same process, requiring similar types of resources and activities, and that clinicians and resources are interchangeable. In view of the findings discussed in Section 3.3, we conclude that some clinics indeed have such a high-volume and low-variety characteristic, but that other clinics have substantial job variety and a high level of customization. In combination with the many constraints taken into account, this makes scheduling complex.

This suggests a direction for improvement driven by the question whether these varieties and complexity are all needed. Similar to simplification approaches in Lean manufacturing and elsewhere, such improvement efforts could be driven by an analysis distinguishing between varieties that add substantial value and varieties that do not. When successful, process simplification makes the scheduling task simpler and the relaxation and elimination of boundary conditions and constraints could greatly enlarge the space of potential scheduling solutions.

5.3.2. Looser schedules

A phenomenon that catches our eye is that almost all clinics work with a tight schedule, resulting in much congestion and thus long waiting times for patients and frequent session overruns. Operating at such high levels of utilization is generally advised against in industrial-engineering textbooks, as waiting times and congestion increase exponentially as a function of utilization. Moreover, long waiting times and congestion tend to create even more work for staff and clinicians, thus creating a feedback mechanism that aggravates the congestion.

During our workshops, a few examples were given of such congestion effects: when patients have waited for a long time, clinicians tend to compensate by taking more time for them, thus increasing the delay for the next patients. Or when a session runs late, clinicians may postpone patient-related tasks such as updating the medical record or writing a letter to the GP; such postponements are highly inefficient and hamper rather than improve the process's capacity. Berry Jaeker and Tucker [9] reported similar examples in healthcare operations, where beyond a certain tipping point, increasing utilization becomes counterproductive.

We think that many clinics are now run at utilization levels which are beyond those that maximize productivity, and that creating looser schedules should be considered. Especially since idle time for clinicians is in general a minor problem, as discussed in Section 4, we propose that often, the performance can be improved by increasing the length of slots. Due to the longer slots the planned session time increases, however, both the session runtime and waiting times are reduced. Furthermore, in view of the high level of congestion in current schedules, looser schedules will likely not increase idle time substantially. Even if looser schedules result in more idle time for clinicians, then, following scenario *S2* in Section 4.1, idle time is likely not lost capacity but will be used for substitute tasks. Therefore, in many cases, a looser schedule will reduce congestion, but without significant impact on patient throughput or the capacity of the service.

References

[1] Ackoff RL, Vergara E. Creativity in problem solving and planning: a review. Eur J Oper Res 1981;7(No. 1):1–13.
[2] Ahmadi-Javid A, Jalali Z, Klassen KJ. Outpatient appointment systems in healthcare: a review of optimization studies. Eur J Oper Res 2016;258(No. 1):3–34.
[3] Anderson R, Camacho F, Balkrishnan R. Willing to wait?: the influence of patient wait time on satisfaction with primary care. BMC Health Serv Res 2007;7(No. 31):1–5.
[4] Åström KJ, Hägglund T. The future of pid control. Control Eng Pract 2001;9(No. 11):1163–75.
[5] Bailey NTJ. A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. J R Stat Soc. Ser B (Methodol) 1952;14(No. 2):185–99.
[6] Barratt M, Choi TY, Li M. Qualitative case studies in operations management: trends, research outcomes, and future research implications. J Oper Manag 2011;29(No. 4):329–42.
[7] Barron WM. Failed appointments. who misses them, why they are missed, and what can be done. Prim Care 1980;7(No. 4):563–74.
[8] Begen MA, Queyranne M. Appointment scheduling with discrete random durations. Math Operat Res 2011;36(No. 2):240–57.
[9] Berry Jaeker J, Tucker A. Past the point of speeding up: the negative effects of workload saturation on efficiency and quality. Manage Sci 2017;63(No. 4):1042–62.
[10] Box GEP, Kramer T. Statistical process monitoring and feedback adjustment: a discussion. Technometrics 1992;34(No. 3):251–67.
[11] Cayirli T, Veral E. Outpatient scheduling in health care: a review of literature. Prod Oper Manag 2003;12(No. 4):519–49.
[12] Cayirli T, Veral E, Rosen H. Designing appointment scheduling systems for ambulatory care services. Health Care Manag Sci 2006;9(No. 1):47–58.
[13] Cayirli T, Yang KK, Quek SA. A universal appointment rule in the presence of no-shows and walk-ins. Prod Oper Manag 2012;21(No. 4):682–97.
[14] Charnetski JR. Scheduling operating room surgical procedures with early and late completion penalty costs. J Oper Manag 1984;5(No. 1):91–102.
[15] Cousens A, Szwejczewski M, Sweeney M. A process for managing manufacturing flexibility. Int J Oper & Prod Mana 2006;29(No. 4):357–85.
[16] D'Souza DE, Williams FP. Toward a taxonomy of manufacturing flexibility dimensions. J Oper Manag 2000;18(No. 5):577–93.
[17] Denton B, Viapiano J, Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 2007;10(No. 1):13–24.
[18] De Snoo C, Van Wezel W, Jorna RJ. An empirical investigation of scheduling performance criteria. J Oper Manag 2011;29(No. 3):181–93.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122

[19] Drupsteen J, Van der Vaart T, Van Donk DP. Integrative practices in hospitals and their impact on patient flow. Int J Oper & Prod Manag 2013;33(No. 7):912–33.
[20] Eisenhardt KM. Building theories from case study research. AcadManag Rev 1989;14(No. 4):532–50.
[21] Fetter RB, Thompson JD. Patients' waiting time and doctors' idle time in the outpatient setting. Health Serv Res 1966;1(No. 1):66–90.
[22] Fries BE, Marathe VP. Determination of optimal variable-sized multiple-block appointment systems. Oper Res 1981;29(No. 2):324–45.
[23] Glowacka KJ, Henry RM, May JH. A hybrid data mining/simulation approach for modeling outpatient no-shows in clinic scheduling. J Oper Res Soc 2009;60(No. 8):1056–68.
[24] Green LV, Savin S. Reducing delays for medical appointments: a queueing approach. Oper Res 2008;56(No. 6):1526–38.
[25] Gupta D, Denton B. Appointment scheduling in health care: challenges and opportunities. IIE Trans 2008;40(No. 9):800–19.
[26] Hassin R, Mendel S. Scheduling arrivals to queues: a single-server model with no-shows. Manage Sci 2008;54(No. 3):565–72.
[27] Healy KJ. Scheduling arrivals to a stochastic service mechanism. Queueing Syst 1992;12(No. 3):257–72.
[28] Heskett JA, Sasser WE, Hart CW. Service breakthroughs. New York, NY: The Free Press; 1990.
[29] Ho CJ, Lau HS. Minimizing total cost in scheduling outpatient appointments. Manage Sci 1992;38(No. 12):1750–64.
[30] Ho CJ, Lau HS. Evaluating the impact of operating conditions on the performance of appointment scheduling rules in service systems. Eur J Oper Res 1999;112(No. 3):542–53.
[31] Hopp WJ, Spearman ML. Factory physics. 3rd ed. Boston, MA, USA: McGraw-Hill; 2008.
[32] Huang X. Patient attitude towards waiting in an outpatient clinic and its applications. Health Serv Manag Res 1994;7(No. 1):2–8.
[33] Jack E, Raturi A. Sources of volume flexibility and their impact on performance. J Oper Manag 2002;20(No. 5):519–48.
[34] Jiang R, Shen Y, Zhang Y. Integer programming approaches for appointment scheduling with random no-shows and service durations. Oper Res 2017;65(No. 6):1638–56.
[35] Johnson BJ, Mold JW, Pontious JM. Reduction and management of no-shows by family medicine residency practice exemplars. Annal Family Med 2007;5(No. 6):534–9.
[36] Kaandorp GC, Koole G. Optimal outpatient appointment scheduling. Health Care Manag Sci 2007;10(No. 3):217–29.
[37] Ketokivi M, Choi T. Renaissance of case research as a scientific method. J Oper Manag 2014;32(No. 5):232–40.
[38] Klassen KJ, Rohleder TR. Scheduling outpatient appointments in a dynamic environment. J Oper Manag 1996;14(No. 2):83–101.
[39] Klassen KJ, Rohleder TR. Combining operations and marketing to manage capacity and demand in services. Serv Ind J 2001;21(No. 2):1–30.
[40] Klassen KJ, Rohleder TR. Demand and capacity management decisions in services: how they impact on one another. Int J Oper & Prod Manag 2002;22(No. 5):527–48.
[41] Klassen KJ, Yoogalingam R. Appointment system design with interruptions and physician lateness. Int J Oper & Prod Manag Econ 2013;33(No. 4):394–414.
[42] Kuiper A, Kemper B, Mandjes M. A computational approach to optimized appointment scheduling. Queueing Syst 2015;79(No. 1):5–36.
[43] Kuiper A, Mandjes M. Appointment scheduling in tandem-type service systems. Omega (Westport) 2015;57(No. B):145–56.
[44] Kuiper A, Mandjes M, De Mast J. Optimal stationary appointment schedules. Oper Res Lett 2017;45(No. 6):549–55.
[45] Liao CY, Pegden DC, Rosenshine C. Planning timely arrivals to a stochastic production or service system. IIE Trans 1993;25(No. 5):63–73.
[46] Luo J, Kulkarni V, Ziya S. Appointment scheduling under patient no-shows and service interruptions. Manuf Serv OperManag 2012;14(No. 4):670–84.
[47] Mak HY, Rong Y, Zhang J. Appointment scheduling with limited distributional information. Manage Sci 2015;61(No. 2):316–34.
[48] Millhiser WP, Veral EA, Valenti BC. Assessing appointment systems' operational performance with policy targets. IIE Trans Healthc Syst Eng 2012;2(No. 4):274–89.
[49] Mondschein SV, Weintraub GY. Appointment policies in service operations: a critical analysis of the economic framework. Prod Oper Manag 2003;12(No. 2):266–86.
[50] Rising EJ, Baron R, Averill B. A systems analysis of a university-health-service outpatient clinic. Oper Res 1973;21(No. 5):1030–47.
[51] Robinson LW, Chen RR. Scheduling doctors' appointments: optimal and empirically-based heuristic policies. IIE Trans 2003;35(No. 3):295–307.
[52] Rohleder TR, Klassen K. Using client-variance information to improve dynamic appointment scheduling performance. Omega (Westport) 2000;28(No. 3):293–302.
[53] Salzarulo PA, Bretthauer KM, Côté MJ, Schultz KL. The impact of variability and patient information on health care system performance. Prod Oper Manag 2011;20(No. 6):848–59.
[54] Soltani M, Samorani M, Kolfal B. Appointment scheduling with multiple providers and stochastic service times. Eur J Oper Res 2019;277(No. 2):667–83.
[55] Soriano A. Comparison of two scheduling systems. Oper Res 1966;14(No. 3):388–97.
[56] Stuart I, McCutcheon D, Handfield R, McLachlin R, Samson D. Effective case research in operations management: a process perspective. J Oper Manag 2002;20(No. 5):419–33.
[57] Swisher J, Jacobson S, Jun J, Balci O. Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation. Comput Oper Res 2001;28(No. 2):105–25.
[58] Vanden Bosch PM, Dietz DC. Minimizing expected waiting in a medical appointment system. IIE Trans 2000;32(No. 9):841–8.
[59] Vanden Bosch PM, Dietz DC, Simeoni JR. Scheduling customer arrivals to a stochastic service system. Naval Res Logis 1999;46(No. 5):549–59.
[60] Voss C, Tsikriktsis N, Frohlich M. Case research in operations management. Int J Oper & Prod Manag 2002;22(No. 2):195–219.
[61] Wang PP. Optimally scheduling n customer arrival times for a single-server system. Comput Oper Res 1997;24(No. 8):703–16.
[62] Welch JD, Bailey NTJ. Appointment systems in hospital outpatient departments. The Lancet 1952;259(No. 6718):1105–8.
[63] White DL, Froehle CM, Klassen KJ. The effect of integrated scheduling and capacity policies on clinical efficiency. Prod Oper Manag 2011;20(No. 3):442–55.
[64] Yang KK, Lau ML, Quek SA. A new appointment rule for a single-server, multiple-customer service system. Naval Res Logis 1998;45(No. 3):313–26.
[65] Yin RK. Case study research: design and methods. 6th ed. London, United Kingdom: Sage Publications; 2013.
[66] Zacharias C, Pinedo M. Appointment scheduling with no-shows and overbooking. Prod Oper Manag 2014;23(No. 5):788–801.
[67] Zacharias C, Armony M. Joint panel sizing and appointment scheduling in outpatient care. Manage Sci 2017;63(No. 11):3978–97.
[68] Zacharias C, Pinedo M. Managing customer arrivals in service systems with multiple identical servers. Manuf Serv Oper Manag 2017;19(No. 4):639–56.
[69] Zacharias C, Yunes T. Multimodularity in the stochastic appointment scheduling problem with discrete arrival epochs. Manage Sci 2018 Forthcoming.
[70] Zeithaml VA, Parasuraman A, Berry LL. Problems and strategies in services marketing. J Mark 1985;49(No. 2):33–46.

Please cite this article as: A. Kuiper, J. de Mast and M. Mandjes, The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice, Omega, https://doi.org/10.1016/j.omega.2019.102122