Chapter 9
Improving Scheduling and Flow in Complex Outpatient Clinics

Craig M. Froehle and Michael J. Magazine

1 Outpatient Care: Workhorse of the Healthcare System

By nearly any measure, outpatient care is increasingly critical to healthcare systems' abilities to care for the masses of patients requesting service. In the USA, like many countries, outpatient services are growing the fastest of any mode of care delivery (Fig. 9.1). Three drivers of this trend are (a) the increasing ability of medicine to treat patients without admitting them to the hospital, (b) the financial advantages of providing care via outpatient means, and (c) the ongoing need to reserve inpatient capacity (which is not growing at the same rate as the population and, in some areas, is actually shrinking) for the most severely ill patients (Bazzoli et al. 2003).

One of the greatest concerns of outpatient facility managers is the rapidly rising cost of providing care, both overall and per capita (Peterson and Burton 2007). This trend reinforces the need to operate outpatient clinics in the most efficient ways possible. Human resources, such as physicians and nurses, and physical resources, like exam rooms and testing equipment, often represent significant fixed expenses, so maintaining a reasonable level of utilization is vital to maintaining the financial viability of a clinic or hospital.

C.M. Froehle (✉)
Department of Operations, Business Analytics, and Information Systems, Carl H. Lindner College of Business, University of Cincinnati, 2925 Campus Green Drive, Cincinnati, OH, 45221-0130 USA

James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA
e-mail: craig.froehle@uc.edu

M.J. Magazine
Department of Operations, Business Analytics, and Information Systems, Carl H. Lindner College of Business, University of Cincinnati, 2925 Campus Green Drive, Cincinnati, OH, 45221-0130 USA

B.T. Denton (ed.), *Handbook of Healthcare Operations Management: Methods and Applications*, International Series in Operations Research & Management Science 184, DOI 10.1007/978-1-4614-5885-2_9,

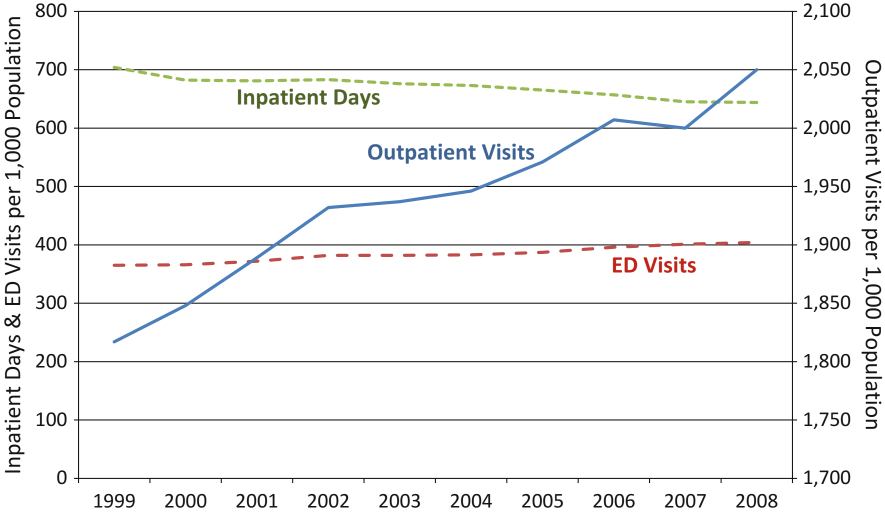

Fig. 9.1 US hospital visits per 1,000 population, 1999–2008. *Notes*: Data are for community hospitals (85 % of all US hospitals), while federal hospitals, long-term care hospitals, psychiatric hospitals, and other specialty institutions not included; Source data: StateHealthFacts.org (2011)

Another motivation for improving the ability to manage flow, or the ability of a healthcare organization to "serve patients quickly and efficiently as they move through stages of care" (Hall 2008), in outpatient environments is that many of the lessons we learn there are applicable to other healthcare settings. For example, scheduling of non-emergency surgeries has benefited from the extensive work on scheduling in outpatient care, as both involve elective visits and are appointment-driven operations (Cardoen et al. 2010; Gul et al. 2011). Additionally, inpatient and emergency care areas of a hospital often have common patient-contact points or share resources with outpatient functions; staff can provide care in more than one location, and the services can share common physical space, such as parking facilities. In some instances, emergency department (ED) flow can be modeled very similarly to outpatient care clinics, especially for that portion serving less urgent patients. In fact, scheduling of non-emergent ED visits, just as is typical of outpatient visits, is becoming increasingly common (Fiore 2010). Thus, the findings from research on outpatient operations can often be immediately useful in other care contexts, if not also to other service industries.

As the title suggests, the focus of this chapter is on "complex" outpatient clinics. We use that term here to refer primarily to those clinics caring for patients who need access to more than a few different types of care providers during their visits and whose paths through the system may be significantly heterogeneous. For example, Fig. 9.2 shows the actual paths (i.e., the sequence of care providers) through a pediatric, pulmonary clinic taken by a sample of 30 patients. This clinic provides outpatient care to children who suffer from pulmonary diseases, such as asthma and cystic fibrosis. Note that while there is a single path that represents a typical (i.e.,

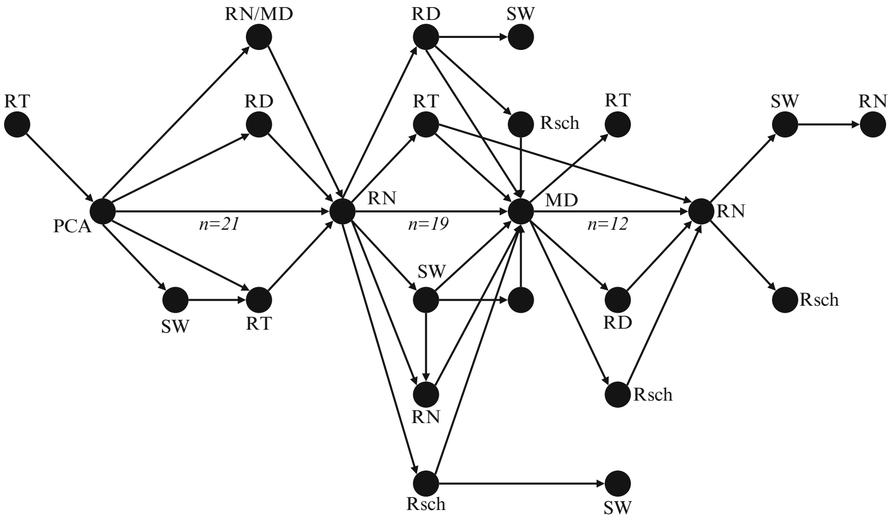

Fig. 9.2 Realized patient paths through a complex outpatient pulmonary clinic. *Notes*: *RT* respiratory therapist, *PCA* patient care assistant, *RD* registered dietician, *RN* registered nurse, *MD* medical doctor (physician), *SW* social worker, *Rsch* researcher, *n* number of patients on a path; Source: Cincinnati Children's Hospital Medical Center, Division of Pulmonary Medicine

most common) route through the system (i.e., PCA (patient care assistant) → RN (registered nurse) → MD (medical doctor) → RN), less than half the patients in the sample traveled along this path. Also note that this does not include the registration activity, nor does it include initial or mid-process waits, all of which can add complexity to the flow in a clinic.

Operational complexity is often a concern for specialty and multidisciplinary clinics (Gupta and Denton 2008), those caring for patients with comorbidities or who are chronically ill, and teaching institutions that employ fellows and residents. Large patient volume is a contributing factor to complexity, but large patient volume alone does not necessarily make a clinic complex. Similarly, even if different types (medically diverse) of patients are treated, it may not be an operationally complex environment if all patients progress through the same care process. Akin to a typical assembly line, homogeneous flow, even at high volumes, is generally more straightforward to plan for and manage.

Realistically, different outpatient clinics may choose rather different operational objectives. For example, the high-volume, low-acuity pediatric clinic may prefer to minimize patient waiting, while the for-profit, imaging clinic may wish to maximize facility utilization, and the orthopedic clinic may prefer to maximize provider (physician) utilization while keeping average patient waits below some chosen value. Or each of these may choose a different metric (or set of metrics), depending on their priorities, constraints, and stakeholder preferences. Devising desirable performance metrics, and balanced combinations of metrics, is a vital element of improving any clinic's operations (see Table 9.1 for examples).

Table 9.1 Some common operational metrics used in outpatient clinics

Metrices	Notes/description
Patient-centered metrics	
Time to access clinic	Also referred to as "indirect waiting" (Gupta and Denton 2008)
Flow time	Total duration a patient is in the clinic
Waiting time	Also referred to as "direct waiting" (Gupta and Denton 2008)
Touch time ratio	Time patient is being served divided by flow time
Provider-centered metrics	
Provider idle time	Total time a provider is in clinic, but not serving patients
Provider utilization	Time a provider is serving patients divided by available time
Provider touch time ratio	Time serving patients divided by available time
Organization-centered metrics	
Clinic overtime	Amount of time past scheduled end that patients are in clinic
Number of patients seen	Less useful in clinics with heterogeneous patient mixes
Facility (exam room) utilization	Can be helpful for space allocation decision-making
Revenue per clinic	Can depend greatly on reimbursement assumptions

In this chapter, we broadly examine some of the primary barriers and opportunities for operational performance and improvement in complex outpatient environments. That makes this work different but complementary to those focused on primary care. After reviewing the challenges, we propose a managerial and analytical framework that illustrates how solving various problems can contribute to a comprehensive operational approach to improved clinic flow. While this chapter is neither a comprehensive literature review nor an in-depth methodological treatise, it should help healthcare professionals understand how different studies and topics within operations research may be relevant to this context. It should also aid academics in identifying specific research opportunities for extending our knowledge of, and capability for improving, outpatient clinic operations.

2 Pervasive Barriers to Outpatient Clinics Operating Effectively

Two phenomena that greatly inhibit a clinic's ability to manage itself in an efficient and effective manner are *uncertainty* and *complexity*. By uncertainty, we mean the inability to know in advance, with perfect accuracy, some operational aspect of the clinic, such as the duration of a consultation activity or the set of providers with whom a patient will need to interact during his visit. By complexity, again, we refer primarily to the quantity and diversity of medical staff and physical resources involved in providing care to the clinic's patients.

These phenomena affect two fundamental operational activities of a clinic—(a) planning and scheduling and (b) managing clinic flow—in different ways (see Table 9.2). Planning and scheduling activities for an outpatient clinic involve those functions that help ensure that clinics will be as accessible to patients, and executed

Table 9.2 Some implications of complexity and uncertainty for clinic operations

	Planning and scheduling (prior to clinic)	Managing clinic flow (during the clinic)
Complexity	1. Identifying all resources each patient needs to access takes more effort	5. Necessary to orchestrate multiple providers' and patients' activities
	2. Requires joint scheduling of multiple providers	6. Aggregating and analyzing complete and timely operational data difficult to do manually
Uncertainty	3. Patient/provider activity sequencing is generally suboptimal	7. Arrival of patients unknown 8. Timing of tasks unknown 9. Care needs (especially new patients) unknown
	4. Total patient resource needs unclear	

as smoothly, as possible. This encompasses a wide variety of tasks, such as initial scheduling of patient appointments, determining (or estimating) which services and providers each patient will need to see, reserving the necessary equipment and facilities, and so forth (Pinedo 2009). All of these planning and scheduling activities related to a specific clinic session occur before that particular clinic session starts.

In contrast, managing clinic flow is the daily challenge of implementing the clinic plan and directing the activities of staff and patients as the clinic occurs. During the clinic session, as patients arrive (early, on time, late, or not at all) and clinic staff meet with patients (for various amounts of time), deviations from the original schedule for that clinic nearly always occur. Reaction, in a timely manner, to these changes necessitates the constant reformulation of a new plan. That, combined with the responsibilities of guiding people and coordinating resources needed to deliver the appropriate care, contributes to the potential difficulty of managing clinic flow. Below, we discuss some of the particular challenges associated with both planning and scheduling activities and managing patient flow in complex and uncertain clinic environments.

2.1 Planning and Scheduling Prior to the Clinic Session

A great deal of planning for a clinic session generally takes place well in advance of the session actually getting underway. In complex and chronic care clinics, scheduling patients in advance is commonplace. Creating the ideal appointment templates (the arrangement of appointment times and durations reserved for different patient/visit types who need to see specific providers or provider types), and scheduling in general, has been the subject of extensive operations research in healthcare over the years. SeeCayirli and Veral (2003) and Gupta and Denton (2008) for excellent overviews of this area of study and Pinedo (2009) for a general review of scheduling work that potentially could be applied to this environment.

After the schedule has been filled, or is nearly full, a common practice for planning for a clinic is for staff to hold a meeting prior to the clinic session to review the medical needs of the slate of patients scheduled to receive care during the clinic session. This meeting, which some hospitals and clinics refer to as a "chart conference," is typically held to ensure that each patient's visit includes all care activities needed to address that patient's particular condition, and is often additional work for clinicstaff (#1 in Table 9.2). One result of a chart conference is the need to then schedule those care providers who may not already be scheduled to see that particular patient. Coordinating the schedules of the various providers can be challenging, especially if the providers are being drawn from a multitude of specialties or practices (#2 in Table 9.2).

Just as in manufacturing and other production environments, in complex outpatient clinics, uncertainty can make planning and scheduling more difficult. Patient arrivals to the clinic can be earlier than scheduled, on time, later than scheduled, or not at all (no-shows) (Salzarulo et al. 2011). Another source of uncertainty is the duration of the patient–provider consultation. While most clinic schedule templates have predetermined blocks of time roughly corresponding with the expected duration of the patient's consultation with the physician and/or other staff, the actual durations can be highly variable (Harper and Gamlin 2003; Chand et al. 2009; LaGanga 2011).

This combination of uncertain patient arrivals and uncertain consultation durations makes creating the optimal schedule of all clinic activities difficult (#3 in Table 9.2) (Please see our approach to this in the research opportunities section of this chapter). Research has shown that scheduling low-variance patients earlier in the clinic, and high-variance patients later, can generate small reductions in patient waiting and overall clinic duration (White et al. 2011a). However, with few exceptions, these studies have generally considered only simple clinic environments (e.g., a single patient flow through a fixed sequence of providers), and there is as yet little evidence to guide scheduling of patients who need to see multiple providers and with different variability profiles.

Further complicating matters is that the exact resource needs of patients are not often fully known when planning the clinic session. For example, the specific providers a patient will need to see may not be entirely known until he has consulted with one or more care providers (#4 in Table 9.2). This is especially common with new patients and those with complex, chronic diseases. It creates a situation where the clinic must then react to changes in their assumptions about who needs to interact with that patient during his visit. An alternative that few prefer is to ask the patient to come back another day to meet with additional staff; this puts a hardship on the patient and creates another planning and scheduling responsibility for a future clinic. A different form of uncertainty regarding total patient resource needs (#4 in Table 9.2) is the fact that not all patients who will participate in the clinic may be known when planning and scheduling activities occur; add-ons (those added to the schedule after the chart conference has occurred) and walk-ins (those showing up without advance notice or an appointment) are two examples.

2.2 Managing the Flow of the Clinic

One of the keys to effective flow is proper planning and scheduling before the clinic session begins. As shown in Table 9.1, both complexity and uncertainty can limit the organization's ability to effectively manage clinic flow. As the complexity—the number and variety of care providers and patients—of the clinic increases, it becomes progressively more difficult to track and coordinate the actions of everyone involved (#5 in Table 9.2). The location and activity of each provider and patient, the time elapsed since they began those activities, and other details are essential to actively managing flow in a clinic. Keeping track of one or a few exam rooms with one or a few staff might be possible, even effective, using purely manual methods; however, any clinic more complex than that will likely require some form of information system to support the human decision-makers (#6 in Table 9.2).

Uncertainty surrounding patient arrivals creates a challenge for staff trying to manage the flow of the clinic (#7 in Table 9.2). Unless a patient contacts the clinic to let it know he will not be showing up on time, the clinic has no reliable information about when, or even if, the patient will ultimately arrive. This can make task sequencing decisions difficult. For example, a patient has arrived early (ahead of his appointment time) and another patient is late (has not yet arrived even though his appointment time has passed). If we place the present patient in an exam room (ahead of his appointment), it will delay serving the late patient should he arrive soon thereafter. While first come, first served (FCFS) may have been honored, FAFS (first appointment, first served) would be violated. See White et al. (2012) for work with an outpatient clinic that considered different queue disciplines.

Clear policies can help ensure consistent decision-making, but left unresolved is the question of whether or not the policy leads to the best decision at this particular moment with respect to the clinic's performance measures (e.g., average patient waiting, overtime, etc.). In the above example, the policy may be to make the early patient wait until his appointment time. Of course, that may be suboptimal if the late patient turns into a no-show; the waiting patient was delayed for no benefit to anyone. Some clinics have specific policies for this and other situations, but many do not, instead leaving it up to whoever may be managing the clinic at the time.

Similarly, if a patient and physician began their consultation 25 min ago and the scheduled duration is 30 min, there is little guarantee they will be done in less than 5 min, let alone exactly 5 min. This type of uncertainty (#8 in Table 9.2) creates a need for constant monitoring of all activities. Then, making decisions based on both actual events (the end of the consultation) and anticipated events (the fact that the physician and patient should both be freed up within the next few minutes) can easily overwhelm manual systems or those relying simply on a clinic manager acting as an "air traffic controller" (a common analogy in outpatient clinics).

Also, in complex clinics, the exact care needs of a particular patient may not be easily determined prior to physically meeting with the patient in the clinic (#9 in Table 9.2). Apart from the planning and scheduling challenges identified above, this complicates managing the flow of the clinic because additional activities may need to be inserted into the schedule while the clinic is happening. These real-time adjustments can pose significant problems for clinic managers for two reasons. First, the schedule was likely based on some estimates of how much staff time each patient would consume; if that time gets increased significantly, many other activities may need to be pushed back, potentially resulting in an overly long clinic, excessive patient waiting, and other issues. Second, if the added care component requires a specific type of medical staff, that resource may not be available at the time she is needed in the clinic. This can even further exacerbate an already chaotic and operationally compromised clinic.

In order to move patients through their clinic visits with minimal delay while keeping staff and facilities adequately utilized, a comprehensive approach is needed that integrates decision-making associated with planning and scheduling and managing flow in real time. A proposed framework for such a system is described in the next section. The framework represents an assembly of existing policies and practices that can be found in many hospitals and clinics today as well as opportunities for overcoming the inadequacies of unassisted, manual decision-making in these dynamic, complex, and uncertain environments. Those opportunities are then further discussed in the final section of this chapter.

3 COMS: A Proposed Approach for Improving Flow in Outpatient Clinics

We refer to the Clinic Operations Management System (COMS) as a set of policies, practices, and analytical methods that may be embedded in a clinic's information system. It is a conceptual framework, used here to frame the discussion about the broader problem of planning and scheduling outpatient clinics in a comprehensive manner. To date, the literature has largely focused on planning and scheduling scenarios that consider only simplified models (e.g., a single server or no recourse over the clinic's duration) or address only subsets of the overall process. In contrast, COMS details one possible vision of an integrated system, with many opportunities for applying advanced analytical methods to improve the clinic's operating performance. A functional schematic is shown in Fig. 9.3, where COMS can be seen to assume four main responsibilities:

(a) It serves as the central clearinghouse for clinic schedule data.
(b) It assists planning by providing information about past patient visits and by recording decisions made about upcoming patient visits (e.g., which providers a patient must see during his visit).
(c) It tracks patients and providers as they move through the clinic (or works in conjunction with a separate real-time locating system, or RTLS).
(d) It supports real-time operational decision-making for coordinating the activities of staff and patients as the clinic progresses.

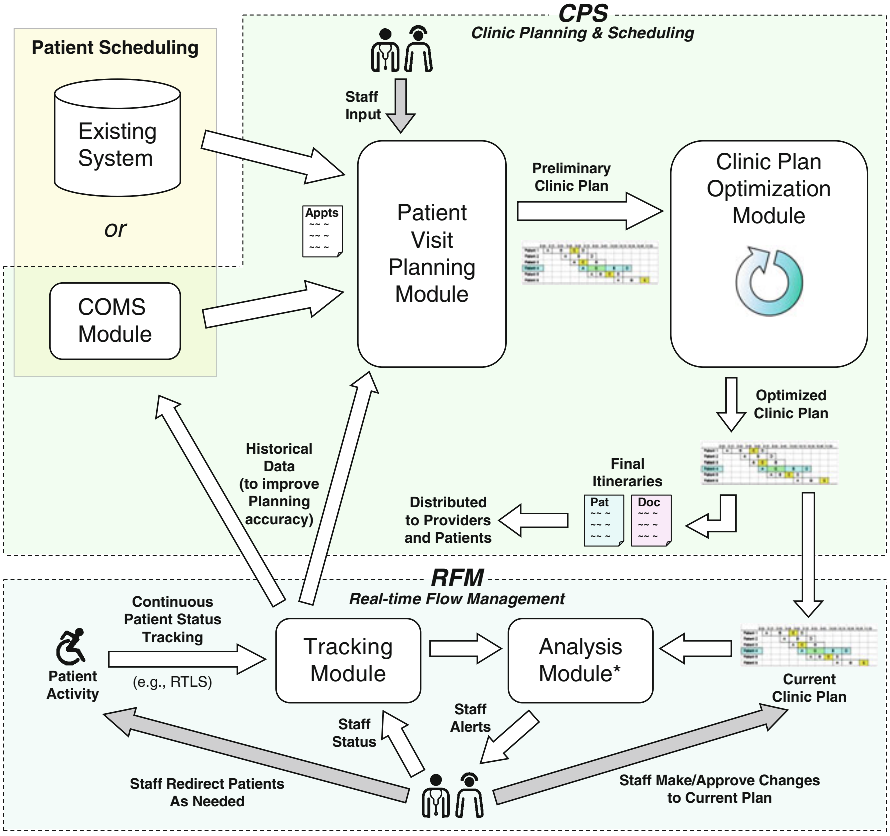

Fig. 9.3 COMS functional architecture. The Clinic Operations Management System (COMS) represents the major information flows, analytical components, and coordination mechanisms of a complete outpatient clinic planning, scheduling, and flow management system. *Asterisk* could employ analytical approaches similar to the Clinic Plan Optimization Module (see Sect. 4 for more details)

Responsibilities (a) and (b) are associated with improving the clinic's planning and scheduling activities, whereas responsibilities (c) and (d) have to do with improving the clinic's ability to do real-time flow management. Planning and scheduling activities occur prior to a clinic session, while flow management occurs during a specific clinic session.

3.1 *Clinic Planning and Scheduling Functions*

Clinic Planning and Scheduling (CPS) incorporates all functions needed by clinicians to collect, structure, and organize the operational information involved in planning a particular clinic session (e.g., morning clinic, Wednesday, August 14th), including:

1. Patient appointment requests—Ideally, when a patient contacts the clinic (or call center) for an appointment, the system (COMS or a separate scheduling system with information fields populated by COMS) would suggest appointments of appropriate durations based on the type of appointment being requested (e.g., new vs. follow-up) that permit that patient to see all appropriate providers in as compact a schedule (minimal patient waiting and provider idle time) as possible.
2. Patient visit planning—Before the patient arrives at the clinic, a list (or sequence, if task precedents are important) of service activities (providers, tests, etc.) for that patient's visit must be provided by clinic staff, such as, through the aforementioned chart conferences. These lists can be, and will likely initially start out as, generic templates, perhaps differentiated by patient category. However, the more accurate and detailed these patient visit plans are, the better able the system will be to optimize the operating plan for that clinic. This combination of patient arrivals (appointments) and the slate of providers each patient should see represents a preliminary *clinic plan* for that clinic session and is fed into the next step: optimization.
3. Clinic plan optimization—Once a clinic's available appointment slots have been filled, or the start of the clinic is a minimum planning period away (e.g., two days from now), an "optimization"[1] module will arrange patients, providers, and/or tasks so as to achieve one, or a combination, of several objectives (e.g., minimize patient waiting, minimize staff idle time, minimize total patient flow time, and maximize robustness to disruption). This can be achieved in several ways, which are discussed in the last section of this chapter.

The result of these three tasks is a comprehensive plan for a specific clinic session, including the roster of patients to be served, the (desired) order of their arrival, the sequence of service activities for each patient, and the planned start and end times for each activity. These sequences, plus their associated start and end times, create *itineraries* for both patients and providers as they travel through the clinic's overall operating plan. Figure 9.4 shows a small clinic plan covering one physician, a total of four different care providers/activities (e.g., medical assistant, nurse, physician, and phlebotomy), and six patients.

Patient 4's itinerary (shaded row) would consist of:

```
a. Arrive and see Provider A at 9:00
b. See Provider C at 9:15
c. See Provider B at 9:45
d. See Provider D at 10:15
e. Exit at 10:30
```

[1]The term "optimization" is used here to include both models/systems that produce truly optimal solutions as well as those that may rely on other methods, such as heuristics, to generate optimal or near-optimal solutions.

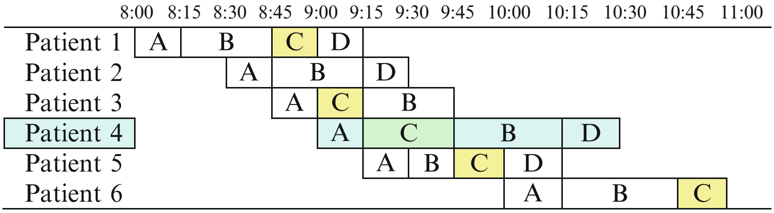

Fig. 9.4 Example of a clinic plan

Provider C's itinerary would consist of:

```
1. See Patient 1 at 8:45 for 15 minutes
2. See Patient 3 at 9:00 for 15 minutes
3. See Patient 4 at 9:15 for 30 minutes
4. See Patient 5 at 9:45 for 15 minutes
5.  Break  at 10:00 for 45 minutes
6. See Patient 6 at 10:45 for 15 minutes
```

The planned duration of each scheduled patient–provider task on the itinerary (e.g., a physician consulting with the patient) should be based on a combination of historical performance (i.e., data), ideal consultation times depending on the particular patient and reason for his visit, and the clinic's operational objectives. The system should know these estimated/planned task durations, which could be determined by clinic management, calculated from historical data, or some combination thereof.

3.2 Real-Time Flow Management Functions

All planning and scheduling activities described immediately above take place prior to the start of the clinic session. Then, with the optimized clinic plan in place, and after all patients and providers have been informed as to their individual itineraries, the clinic session begins. Ideally, all patients arrive at their designated start times and do not require significantly more time from any provider or in any task than was originally scheduled in the plan. If these conditions are met, it should be possible to come close to achieving the clinic's operational performance goals.

However, it will be rare that all patients show up when they are asked, that all providers spend the planned amount of time with each patient, and that there are no delays due to missed communication, misplaced files, or the myriad other sources of disruption typical of an outpatient clinic. When such disruptions happen, COMS must be able to advise the clinic manager as to the best recourse depending on the nature and magnitude of the deviation(s) from the plan, the current state of the system, and other policies and objectives that may influence the clinic's operations.

For example, patient D shows up 20 min late for his 10 a.m. appointment (10 a.m. being the time he was scheduled to first sign in with registration, *not* when he was to see the physician, as is the customary use of the "appointment time" concept). Should the clinic serve this patient (i.e., move him to the next part of his itinerary) next, should it prioritize another waiting patient who was not late, or should it take some other action? Few individuals running a clinic larger than a single physician and a few exam rooms will not have the awareness and real-time analytical capacity necessary to ensure that optimal decisions are consistently known. The system should assess the state of the clinic (e.g., patient A is in task 3 of his itinerary, patient B is in task 2, and patient C is in task 1; each has been in that task for a number of minutes and, therefore, is estimated to have a certain number of minutes remaining before moving onto the next task); consider patient D's itinerary, as well as any patients scheduled to show up soon; and determine what course of action will most likely achieve the clinic's desired operational results.

Rerunning the decision model every few minutes (or on demand, perhaps motivated by a significant event) would ensure that the clinic staff constantly have access to the best (or approximate best, if heuristics are relied on) course of action based on the latest information about the current state of the clinic. While powerful, this rescheduling capability, which also exists in many MRP systems (Ho 1989), causes disruptions to the system that may actually make things worse. These frequent shocks to the system, called "system nervousness," may not actually move the solution to a better place and can lead to staff and patient unrest. The challenge is to find the proper level of nervousness that balances the value of acting on new information with the cost of frequent adjustment.

One common scenario that such a decision model should be ready to address is the arrival of an unscheduled, or add-on, patient who was not part of the original plan developed for the clinic during the planning and scheduling process. Upon being informed that a patient will be arriving, or has arrived, for which the clinic plan was not prepared, COMS would ideally provide clinic management with the best course of action (e.g., serve immediately and at which task, serve at some later point in the clinic, or schedule for another day). It would also be necessary to input a new preliminary plan for the add-on patient. COMS would then generate an itinerary for him as well as update the staff itineraries for all those providers who will see the add-on patient during that clinic session.

3.3 Other Functions

In order to maximize its usefulness and overcome the barriers to clinic operating effectiveness identified earlier, COMS should handle several other responsibilities. Tracking of individuals and other resources will be requisite in order for the system to know the status of each patient, provider, and exam room at all times. Tracking should not rely on manual entry of one's status into the system, as that introduces data errors, delay, and unreliability, all of which reduce COMS' ability to make accurate recommendations. Therefore, the tracking system should require no action on part of providers or patients for it to know where they are at all times during the clinic. Some existing, commercially available RTLS solutions have this functionality, relying on RFID, ultrasound, and other technologies, but a constant data feed from the RTLS to COMS would be necessary.

An equally important function is the ability to record operating data as it is generated during the clinic session and synthesize those data into useful inputs for the planning and scheduling process. For example, if every consultation by a physician is accurately measured, then the system will be much better equipped to estimate the duration of a specific consultation involving that physician and a patient of a certain type (if not the specific patient). Or if patient arrivals are tracked accurately, then the clinic can have more confidence in certain decisions, such as determining which appointment slots should be overbooked with a second patient. Without these data, the clinic must resort to relying on averages and rough estimates at best, or simple guesses at worst. Unfortunately, in practice, the quality of operational data is often less than ideal, possibly because hospitals and clinics must often rely on manual data collection methods, which can be costly (or frustrating for staff and patients) to adopt as a standard practice. Automating the collection of highly detailed operational data, which many healthcare information systems do not do, represents a significant opportunity for improvement.

A system like COMS could also enhance the clinic staff's ability to participate during the planning and scheduling process. A browser-based interface for collaborative patient-visit planning by multiple providers would be helpful when face-to-face chart conferences are impossible. Automating the distribution of patient and provider itineraries (e.g., via email or post) would also be a natural function for the system.

The chaotic environment of a busy clinic can make coordinating, and communicating with, the patients, staff, rooms, etc., during the clinic session a daunting task. COMS should facilitate these activities by providing real-time status updates via a variety of media and means. For example, real-time display of the current clinic plan on large displays (e.g., wall-mounted displays) as well as portable devices (e.g., smartphones) would be invaluable to staff who are constantly circulating around the clinic. Also, flexible, multimode delivery of staff alerts, such as via text messaging, paging, email, and perhaps push notifications (such as through custom smartphone apps), would improve the timeliness of staff being alerted to changes to the clinic plan without being overly intrusive and disrupting patient–provider consultations. Or there may be a need to disrupt those meetings, such as in the case of a provider being unaware of the time (and patients queueing up in the waiting room), where an automated nudge would be helpful to indicate that it is time to wrap up the consultation.

COMS is a general concept, comprised of many elements, some of which are well established in practice, whereas others have yet to be developed. The next section explores in greater detail opportunities represented by some of the functions included in the COMS model, both in terms of innovation for practice and research for academics.

4 Optimizing the Clinic Plan

Perhaps, the primary opportunity for innovative improvement to outpatient clinic management is the clinic plan optimization component required by COMS. Multiple approaches could be taken in constructing a model (or system of models) to optimize the clinic plan. While solving a flow shop-style scheduling problem has been widely addressed in the literature (e.g., Minella et al. 2008; Li and Ierapetritou 2008; Pinedo 2009; Baker and Trietsch 2009), the fact that some tasks may have appointment times (versus complete jobs having due dates, as is typical of job-shop scenarios) makes this variation on the traditional machine-scheduling problem more complicated.

One approach could be to treat the set of patient arrival times—the order and approximate times of patient arrivals—as exogenous and fixed, unable to be modified by clinic plan optimization. Taking that approach would limit the model to manipulating only the sequence of activities for each patient within his visit. This can become a challenging problem as more activities, other than arrival times, are associated with appointment times: If there is only one scheduled activity (e.g., a physician's consultation) for each patient, there may be greater opportunity to resequence all of the patient's other activities in order to avoid provider conflicts, reduce patient idle time, etc. However, some care providers in addition to the primary physician may also have their appointments scheduled. This is common for hospital-based providers who may see patients in several different specialties' clinics throughout the day. If several of a patients' tasks are associated with appointment times (e.g., the physician consultation, the dietician consultation, and the physical therapist consultation), the opportunities for improving the patient's itinerary may be very limited. If meeting all of these patient–provider appointment times is considered desirable, but not mandatory (e.g., deviation from the desired appointment time is included in the objective function instead of making task start times constrained to be equal to appointment times), the model may produce better results while still staying within reasonable windows around the original appointment times. This approach also lends itself very well to application both as the initial planning optimizer and the real-time clinic management *re*-optimizer, since reasonably short solution times should be possible, especially if deterministic consultation times are assumed (Froehle et al. 2011).

An alternative approach to planning the clinic could be to treat the patient scheduled arrival times as flexible, which should permit better solutions through manipulation of patient arrival order in addition to the provider/task sequence within each patient described above. In the absence of a scheduling policy where all patients arrive at the beginning of the clinic, taking this approach may require a two-stage patient appointment-setting process; the patient would first be assigned to a specific clinic (e.g., the morning of Wednesday, August 14th), and then, once the roster is full and the planning problem is solved, the patient would be contacted with a specific arrival time (e.g., 10:30 a.m.). A similar two-stage approach is relatively common in scheduling elective surgeries, where "first, advance scheduling gives patients some future date for surgery [and] second, allocation scheduling determines the sequence and resource assignment of the cases in a given day" (Pham and Klinkert 2008). Of course, in an outpatient setting, patients' unwillingness to learn the specific time of their appointment so close to the visit date may limit the utility of this approach.

Another method commonly used is "open access" (also known as "advanced access") scheduling, where some capacity is reserved for patients who need appointments on short notice (Murray and Berwick 2003). While not typically used for the complex and/or multidisciplinary clinics discussed in this chapter, investigating its use in those contexts may be fruitful. See Dobson et al. (2011)and for discussion of this topic.

Many, if not most, clinic scheduling optimization models take a single-period approach; they determine the best plan now and assume future disruptions, which could drive the clinic away from the optimal plan, will not happen. Immediately below is an example of a mixed-integer programming model that provides an initial plan for a clinic using predetermined patient arrival times and focusing on arranging the sequence of patient–provider tasks within each visit.

4.1 Example: Sequencing of Tasks in a Clinic

Considering the clinical environment, an optimization model that would determine the sequence of tasks to be performed on a set of patients with scheduled arrivals could be useful. Here, a mixed-integer programming model has the aim of sequencing all the given tasks for all patients such that the total time in the system for the patients, weighted by their respective processing time, would be minimized. As a first step, we assume processing time for each task to be deterministic and therefore minimizing total time in system amounts to minimizing patient wait time. Weighting by total processing time results in taking a relative perspective on patient wait time, where patients who necessitate longer processing time can wait longer than patients who only need short services.

Consistent with clinical workflow, the model uses the following inputs: a set of patients scheduled for the given block schedule; a set of tasks to be performed on each patient; a set of resources that will perform the tasks; and a duration for each task. In addition, the tasks are assigned to one of four categories (not counting the registration task, which must occur first): (a) tasks that must be performed before the patient sees the physician; (b) tasks that must be performed after the patient sees the physician; or (c) tasks that can be performed either before or after the patient sees the physician. For example, taking the patient's vitals may be a task that has to occur before the patient sees the physician. Similarly, a patient might have to get a blood test after seeing the physician. Other tasks, such as seeing a dietician, could take place either before or after the encounter between the patient and physician. The fourth category of tasks is (d) the encounter between the patient and the physician. This categorization constitutes a constraint on the sequencing and has to be captured in the optimization model. The physician is typically viewed as the bottleneck resource, and thus, the objective of the model is to minimize physician idleness. In summary, each task is uniquely identified, involving a patient, a resource, a duration, and a sequencing category. We now define the elements needed to formulate the optimization model and then express the model in mathematical terms.

Decision variables:

- s_u = the start time of task u
- w_l = the amount of time in the system for patient l = completion time—appointment time for patient l
- $x^k_{u,v} = 1$ if task u occurs before v for resource k, zero otherwise
- $y^l_{u,v} = 1$ if task u occurs before v for patient l, zero otherwise

Sets:

- $T = \{$Tasks to be sequenced during the time horizon (e.g., one clinic day)$\}$
- $P = \{$Patients scheduled to visit the clinic during the time horizon$\}$
- $R = \{$Resources involved in the clinic (the nurses, the physician, the dietician, etc.)$\}$
- $R_k = \{$Tasks to be performed on resource $k\}$
- $P_l = \{$Tasks to be performed on patient $l\}$
- $O_l = \{$Registration task(s) for patient $l\}$
- $A_l = \{$Tasks of category A to be performed on patient l, where A tasks have to occur before the physician sees the patient$\}$
- $B_l = \{$Tasks of category B to be performed on patient l, where B tasks involve the physician$\}$
- $C_l = \{$Tasks of category C to be performed on patient l, where C tasks have to occur after the physician sees the patient$\}$
- $D_l = \{$Tasks of category D to be performed on patient l, where D tasks can occur either before or after the physician sees the patient$\}$

Model parameters:

- t_u = the duration of task u
- a_l = the appointment time of patient l
- $p_l = \sum_{u \in P_l} t_u$
- $\alpha_l = \frac{1}{p_l}$
- M is a nonrestrictive large value in the context of this environment (e.g., the sum of all the processing times)

The objective function can then be defined as:

$$\min \sum_{l}^{P} \alpha_l w_l \tag{9.1}$$

The objective is fulfilled by determining the starting time of each task, thus providing a sequence of tasks. To ensure that the sequence of tasks is valid, we insert the following constraints, where each equation below represents a set of constraints:

$$s_u \geqslant a_l \quad \forall l \in P, \forall u \in P_l \tag{9.2}$$

$$M(1-x^k_{u,v}) + s_v \geqslant s_u + t_u \quad \forall k \in R, \forall u,v \in R_k \quad \text{such that} \quad u < v \tag{9.3}$$

$$Mx^k_{u,v} + s_u \geqslant s_v + t_v \quad \forall k \in R, \forall u,v \in R_k \quad \text{such that} \quad u < v \tag{9.4}$$

$$M(1-y^l_{u,v}) + s_v \geqslant s_u + t_u \quad \forall k \in P, \forall u,v \in P_l \quad \text{such that} \quad u < v \tag{9.5}$$

$$My^l_{u,v} + s_u \geqslant s_v + t_v \quad \forall k \in P, \forall u,v \in P_l \quad \text{such that} \quad u < v \tag{9.6}$$

$$y^l_{u,v} = 1 \quad \forall l \in P, \forall u \in A_l, \forall v \in B_l \quad \text{such that} \quad u \neq v \tag{9.7}$$

$$y^l_{u,v} = 1 \quad \forall l \in P, \forall u \in B_l, \forall v \in C_l \quad \text{such that} \quad u \neq v \tag{9.8}$$

$$y^l_{u,v} = 1 \quad \forall l \in P, \forall u \in O_l, \forall v \in A_l \quad \text{such that} \quad u \neq v \tag{9.9}$$

$$y^l_{u,v} = 1 \quad \forall l \in P, \forall u \in O_l, \forall v \in D_l \quad \text{such that} \quad u \neq v \tag{9.10}$$

$$w_l \geqslant s_u + t_u - a_l \quad \forall l \in P, \forall u \in P_l \tag{9.11}$$

$$s_u, w_l \geqslant 0 \quad \forall u,l \quad x^k_{u,v}, y^l_{u,v} \in \{0,1\} \quad \forall u,v,k,l \tag{9.12}$$

Equation (9.2) ensures that a patient's tasks cannot start before the patient's appointment time. Equations (9.3) and (9.4) capture the fact that a given resource can only process one task at a time and that any task is either before or after another task. These two equations capture these requirements as they are the integer program formulation of a disjunctive pair, meaning that either one or the other constraint, but not both, must hold for a solution to be feasible. Similarly, Eqs. (9.5) and (9.6) capture the fact that for a given patient, only one task can be performed at a time and that any task is either before or after another task. With Eqs. (9.7) through (9.10) we account for the categorization constraints on the order in which tasks have to be performed for a given patient, where any task in A_l should be before any task in B_l, any task in B_l should be before any task in C_l, and registration tasks should be performed before any other task. Finally we use Eq. (9.11) to capture the total time in system for each patient.

To test this example, we coded the model in the AMPL language and solved several small instances using CPLEX (ILOG). Based on an actual pediatric clinic as a convenient example, in each instance solved, we kept the number of resources constant, capturing registration, two nurses, a physician, a dietician, and a social worker, totaling six different caregiver resources. We assumed one unit of each resource (one person), so that the resource can only handle one patient at a time. Also, for convenience, we assumed the patients arrived every 15 min, and, given the short registration task, we assumed this created no initial patient waiting. We were able to solve these problems in a reasonable amount of time, but further work refining the model and developing solution techniques (or heuristics) will likely be necessary for clinics having more patients, providers, and provider types.

5 Opportunities for Future Research and Practice

The improvement of outpatient clinics' operational effectiveness is a broad topic area with many avenues to explore further. Below, we discuss three: (1) the effects of disruptions on efforts to optimize the clinic plan; (2) multidisciplinary coordination; (3) the use of operational data to improve future planning and scheduling activities; and (4) the challenges of managing human behavior in these complex operational environments.

5.1 Disruptions

Since optimizing the clinic plan, such as is accomplished by the model in Sect. 4, is a planning problem, typically solved one or more days prior to the start of the clinic session, we can afford the luxury of finding optimal solutions. But, further research on reformulations (see, e.g., Camm et al. 2008) may improve computation times and enable real-time use of this model. Naturally, this MIP uses deterministic data. However, given the almost certainty of disruptions (e.g., tardy or no-show patients or longer-than-expected consultation times) occurring during clinic operation, we suggest taking one of two approaches. First, the solution process could incorporate an iteration loop that takes the above model's solution and simulates the clinic with realistic and stochastic processing times. Research is needed to develop appropriate measures of robustness (ability to remain a "good" plan even in the face of ordinary disruptions) for the clinic and, using these metrics, revise the optimization model to develop a new clinic plan. This could be repeated until there is an acceptable and reasonably robust plan. Other approaches that might be appropriate here could be the use of safe scheduling, where buffers are added to task times to take into account potential task time changes (Baker and Trietsch 2009).

A second alternative (or perhaps complimentary) approach to handling disruptions would involve regularly re-solving the optimization model during the clinic session. Once we enter the real-time flow management (RFM) phase (i.e., at the start of the clinic), disruptions will occur, making the current plan potentially difficult to adhere to, if even still feasible. With each event epoch (e.g., patient arrivals, task completion, patient departure) or on some predetermined schedule (e.g., every 10 min), we could attempt to re-optimize the clinic plan to get the patient flow back on track to try to meet the clinic's operational goals. Unfortunately, these disruptions can happen often, and solving a mixed-integer program to make adjustments may take too long. Plus, system nervousness might dictate less frequent changes. So, fast heuristics, such as swapping upcoming tasks, may be needed to guide rapid updates to the current clinic's operating plan. Research into practical algorithms, both from a run-time point-of-view and also from realistic clinic change possibilities, is needed. For instance, a new solution that tells the next arriving patient (due to arrive in 5 min) to come 1 h later is likely not a reasonable response.

5.2 Coordination

A related area of opportunity concerns the growing use of multi-specialty clinics, where patients with very complex conditions may end up seeing physicians and specialists from several different areas of medicine (e.g., cardiology, immunology, and pulmonology) in a single visit. Since these specialists will be coming from different parts of the organization, if not from different organizations entirely, scheduling them in a way that works with their other commitments as well as produces a reasonably compact schedule for the patient is increasingly desirable. Analytically, as the scope of the scheduling problem grows, dimensionality can become a significant barrier to producing optimal schedules. Therefore, again, heuristic-based approaches will likely need to be developed.

And since these heuristics will be of little use by themselves, embedding them within pan-organizational, or even interorganizational, information systems, will be necessary to add the most value to practice. This problem of coordination across multiple divisions of a hospital or other large healthcare provider has been around for decades. Mayo Clinic first dealt with this issue shortly after World War II by establishing a "central appointment desk," which coordinated appointment scheduling for multispecialty patient visits (Berry and Seltman 2008). Today, fast, responsive, and automated information systems offer significant opportunities for improving the effectiveness of these scheduling coordinationactivities.

5.3 Using Operational Data Better

One important function of COMS is its use of historical clinic data to forecast consultation durations involving different combinations of care providers and patients. This is helpful in determining a clinic plan that will be less likely to disruption, as the plan will already reflect the best possible estimate of the task duration (and likely far better than a standard schedule template, which generally attempts only crude differentiation among patients). Mining those historical data in order to make better forecasts about upcoming clinic visits represents a significant opportunity for both research and practice. While standard regression models may provide more accurate predictions than taking global means, using methods to identify and disaggregate influential subpopulations may further improve the accuracy of these estimates.

Another opportunity for future research is to develop ways to draw from the advances made in goods-producing industries, such as by seeking multiple sources of data and using multiple forecasting techniques, including quantitative as well as qualitative methods (Makridakis and Winkler 1983; Fisher and Raman 1996). While historical data will likely be informative, there may be some special cause or new influence that may create deviation from historical patterns. Seeking estimates from all the nurses and physicians that have cared for this patient, or patients like him, could be a useful supplement to the historical data when developing the best prediction of how long an upcoming consultation will take.

5.4 Managing Behavior

Finally, there is the issue of modifying the behavior of patients or providers, if not both. To this point, we have largely focused on accommodating sources of variability (which can lead to uncertainty if the variability is not easily predicted) by improving both our ability to plan for it and our ability to react to it in a more timely and rational manner. A complementary strategy that exists is reducing the variability so that less uncertainty has to be accommodated. For example, while much of the discussion above has involved approaches to predicting and compensating for consultations of uncertain durations, we could combine that with managerial tools that reward more consistent adherence to the scheduled durations (or, alternately, that discourage gross deviations). While there will always be legitimate reasons to extend a consultation (or to end it earlier than planned), not all activities during a consultation might add value and could be safely eliminated to help keep the clinic on schedule. If awareness of the time elapsed during a consultation is a barrier to providers keeping true to their itineraries, a system of automated reminders, such as pages or texts sent to cellular phones, might be implemented (an excellent addition to the functionality of COMS, which will already be trackingactivities).

A better understanding of the sources of this variability will require further research but will be invaluable in helping outpatient clinics plan and manage their flow. This is especially true in complex clinics, as the variability from multiple consultations can accumulate to produce wildly varying operational performance from clinic session to clinic session. It is important to understand that variability cannot be eliminated completely. It is possible, however, to distinguish that variability created by provider input, or even the schedule itself, from "natural" (exogenous) variation, such as that arising from different patients' needs. It is desirable to minimize this first kind of variability, as suggested by Litvak (2005).

6 Conclusions

This chapter explored some of the challenges of achieving consistently excellent operational performance in complex outpatient clinics, offered a comprehensive model of how scheduling and flow might be improved, and discussed related opportunities for innovation in practice and in research. As complex outpatient visits becomes increasingly common, the problem of poor clinic flow will continue to undermine the financial viability of healthcare providers while simultaneously motivating patients to avoid care (or seek it through alternate means). If complex care clinics are to improve their operations, they must begin to consider how tools like COMS can be developed and implemented. Researchers must also contribute to this objective by developing new analytical methods and technologies. With the increasing investment in electronic health record (EHR) systems, and the evolution of new technologies, such as RFID and other sensors, there is tremendous opportunity for making dramatic gains in the operational effectiveness of complex outpatient clinics.

References

Baker KR, Trietsch D (2009) Safe scheduling: setting due dates in single-machine problems. Eur J Oper Res 196(1):69–77

Baker KR, Trietsch D (2009) Principles of sequencing and scheduling. Wiley, New York

Bazzoli GJ, Brewster LR, Liu G, Kuo S (2003) Does U.S. hospital capacity need to be expanded? Health Aff 22(6):40–54

Berry LL, Seltman KD (2008) Management lessons from Mayo Clinic. McGraw-Hill, New York

Camm J, Magazine M, Polak G, Zaric G (2008) Scheduling parallel assembly lines to minimize a shared pool of labor. IIE Trans 40(8):749–758

Cardoen B, Demeulemeester E, Belien J (2010) Operating room planning and scheduling: a literature review. Eur J Oper Res 201(3):921–932

Cayirli T, Veral E (2003) Outpatient scheduling in healthcare: a review of the literature. Prod Oper Manag 12(4):519–549

Chand S, Moskowitz H, Norris JB, Shade S, Willis DR (2009) Improving patient flow at an outpatient clinic: study of sources of variability and improvement factors. Health Care Manag Sci 12(3):325–340

Dobson G, Hasija S, Pinker EJ (2011) Reserving capacity for urgent patients in primary care. Prod Oper Manag 20(3):456–473

Fiore K (2010) Living room is new ED waiting room. MedPage Today. http://www.medpagetoday.com/EmergencyMedicine/EmergencyMedicine/23315. Accessed 12 May 2011

Fisher ML, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. Oper Res 44(1):87–99

Froehle CM, Platt M, Magazine MJ (2011) Optimal outpatient clinic scheduling and rescheduling. University of Cincinnati working paper

Gul S, Denton BT, Fowler JW, Huschka T (2011) Bi-criteria scheduling of surgical services for an outpatient procedure center. Prod Oper Manag 20(3):406–417

Gupta D, Denton BT (2008) Appointment scheduling in health care: challenges and opportunities. IIE Trans 40:800–819

Hall RW (2008) Patient flow: the new queueing theory for healthcare. OR/MS Today 33:36–40

Harper PR, Gamlin HM (2003) Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. OR Spectr 25:207–222

Ho C-J (1989) Evaluating the impact of operating environments on MRP system nervousness. Int J Prod Res 27(7):1115–1135

LaGanga L (2011) Lean service operations: reflections and new directions for capacity expansion in outpatient clinics. J Oper Manag 29(5):422–433

Li Z, Ierapetritou M (2008) Process scheduling under uncertainty: review and challenges. Comput Chem Eng 32(4–5):715–727

Litvak E (2005) Optimizing patient flow by managing its variability. In: Berman S (ed) Front office to front line: essential issues for health care leaders. Joint Commission Resources, Oakbrook Terrace, pp 91–111

Makridakis S, Winkler RL (1983) Averages of forecasts: some empirical results. Manag Sci 29(9):987–996

Minella G, Ruiz R, Ciavotta M (2008) A review and evaluation of multiobjective algorithms for the flowshop scheduling problem. INFORMS J Comput 20(3):451–471

Murray M, Berwick DM (2003) Advanced access: reducing waiting and delays in primary care. JAMA 289(8):1035–1040

Peterson CL, Burton R (2007) U.S. health care spending: comparison with other OECD countries. CRS Reports for Congress, Congressional Research Service, Washington, DC. http://assets.opencrs.com/rpts/RL34175_20070917.pdf. Accessed 25 Aug 2011

Pham D-N, Klinkert A (2008) Surgical case scheduling as a generalized job shop scheduling problem. Eur J Oper Res 185(3):1011–1025

Pinedo M (2009) Planning and scheduling in manufacturing and services. Springer, New York

Salzarulo PA, Bretthauer KM, Côté MJ, Schultz KL (2011) The impact of variability and patient information on health care system performance. Prod Oper Manag 20(6):848–859

StateHealthFacts.org (2011) www.statehealthfacts.org/comparetrend.jsp?cat=8&sort=a&sub=94&typ=1&yr=63&ind=392&srgn=1. Accessed 18 Jan 2011

White D, Froehle CM, Klassen KJ (2011a) The effect of integrated scheduling and capacity policies on clinical efficiency. Prod Oper Manag 20(3):442–455

White D, Magazine M, Morelli S (2012) Who's next: using patient analysis and operations management to address non-punctual and no-show patients. University of Cincinnati working paper