CHAPTER

SCHEDULING AND CAPACITY MANAGEMENT 12

Operations Management in Action

Once upon a time, a patient at Second Street Family Practice in Auburn, Maine, had to wait from 60 to 90 days to be seen for a routine check-up. Then, when the day of the appointment finally arrived, the patient might wait nearly 20 minutes in the waiting room and another 20 for the exam to begin. But thanks to strong leadership, impressive teamwork, and effective tools, patients wanting care from Second Street, even routine check-ups, are now seen the same day they call. The average time patients spend flipping through magazines in the waiting room has dropped to around seven minutes; the exam room wait is down to eight. What's more, staff say they like the new system much better, and patient surveys show that about 90 percent of patients notice and are pleased with the changes as well.

[Clinic leadership], who had been reading and learning about advanced access scheduling, recognized it as the antidote for their frustrations. Developed by Mark Murray, MD, and Catherine Tantau, RN, consultants in Sacramento, California, and promoted by [the Institute for Healthcare Improvement (IHI)] in its office practice programs and on its website, advanced access uses queuing theory to reengineer the standard appointment scheduling system,

OVERVIEW

Matching the supply of goods or services to the demand for those goods or services is a basic operational problem. In a manufacturing environment, inventory can be used to respond to fluctuations in demand. In the healthcare environment, safety stock can be used to respond to fluctuations in demand for supplies (see chapter 13), but stocking healthcare services is not possible. Therefore, capacity must be matched to demand. If capacity is greater than demand, resources are underutilized and costs are high. Idle staff, equipment, or facilities increase organizational costs without increasing revenues. If capacity is lower than demand, patients endure long waits or find another provider.

To match capacity to demand, organizations can use demand-influencing strategies or capacity management strategies. Pricing and promotions are often deployed to influence demand and demand timing; however, this strategy typically is not viable for healthcare organizations. In the past, many clinics, hospitals, and health systems used the demand-leveling strategy of appointment scheduling; more recently, many have moved to advanced-access scheduling. Capacity management strategies allow the organization to adjust capacity to meet fluctuating demand; they include using part-time or on-call employees, cross-training staff, and assigning overtime. Effective and efficient scheduling of patients, staff, equipment, facilities, or jobs can help leaders match capacity to demand and ensure that scarce healthcare resources are used to their fullest extent.

This chapter outlines issues and problems faced in scheduling and discusses tools and techniques that can be employed in scheduling patients, staff, equipment, facilities, or jobs. Topics covered here related to scheduling tools and approaches include

- hospital census and resource loading,
- staff scheduling,
- job and operation scheduling and sequencing rules,

(continued)

leaving the majority of slots on any given day open for patients who call that day.

The benefits of advanced access go beyond improved scheduling, says IHI director Marie Schall. "It improves quality and continuity," she says. "People can get problems checked sooner rather than later, and they see the same provider virtually every time. We know that continuity contributes to better overall quality." Schall says that through its Breakthrough Series Collaboratives on Reducing Delays and Waiting Times and its IMPACT network, as well as its work with the Veterans Health Administration on improving access to care, IHI has worked with about 3,000 practices to introduce advanced access.

Source: Excerpted from IHI (2012).

OVERVIEW *(Continued)*

- patient appointment scheduling models, and
- advanced-access patient scheduling.

The scheduling of patients is a unique, but important, subproblem of patient flow. Since the mid-twentieth century, much patient care delivery has moved from the inpatient setting to the ambulatory clinic. Because this trend is likely to continue, matching clinic capacity to patient demand becomes an even more critical operating skill. Beyond operational considerations, if capacity management can be deployed to meet a patient's desired schedule, marketplace advantage can be gained. Therefore, this chapter focuses on advanced access (same-day scheduling) for ambulatory patients. Related topics covered in this chapter include

- advantages of advanced access,
- implementation steps, and
- metrics for tracking the operations of advanced-access scheduling systems.

Many of the operations tools and strategies detailed in earlier chapters are demonstrated here to show how to optimize the operations of an advanced-access clinic.

Hospital Census and Rough-Cut Capacity Planning

For many healthcare organizations, the admittance rate and number of occupied beds provide a good indication of the demands being placed on the system. For hospitals, these numbers often can be measured on the basis of the overall patient census. Most hospitals report their census daily and hourly to manage the available beds in the system. However, what many healthcare organizations fail to understand is that the census also provides a view into the resource needs to appropriately staff a system. Exhibit 12.1 shows a three-month view of a census for Vincent Valley Hospital and Health System (VVH). The pattern is

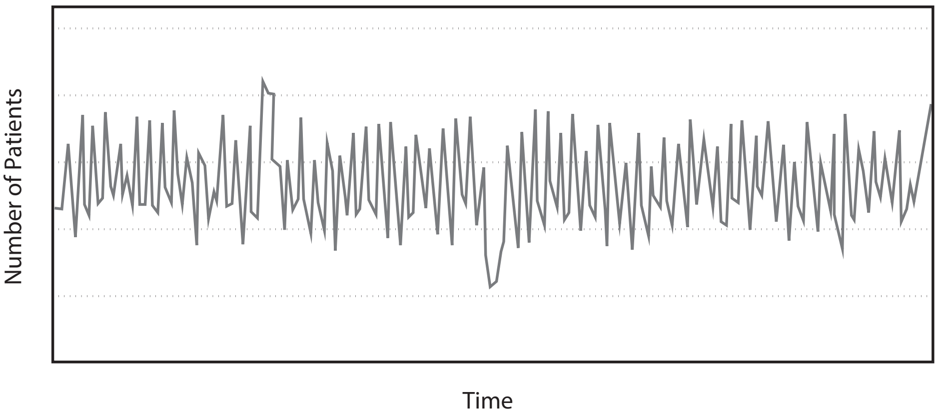


EXHIBIT 12.1
Daily Census at VVH

remarkably similar to most hospitals in that a large amount of variance exists in the patient population on a daily basis. This variance can become magnified when observing the census on an hourly basis.

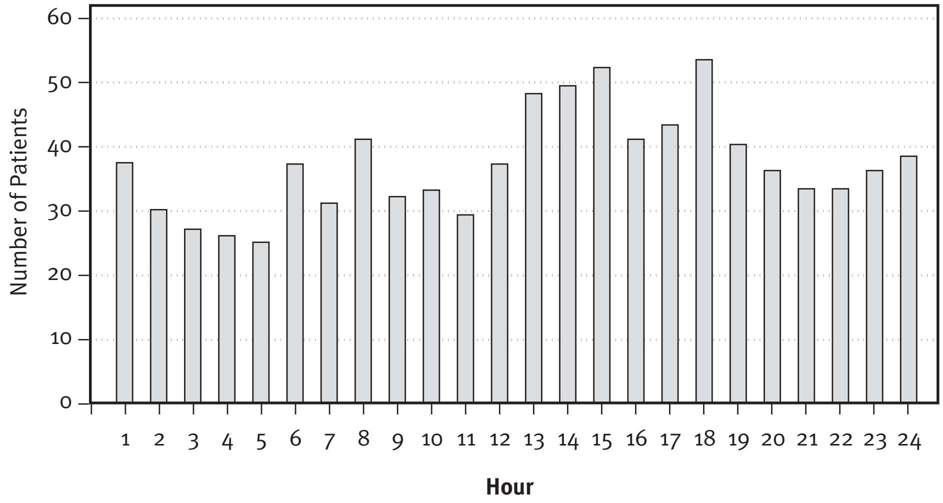
Rough-cut capacity planning
The process of converting the overall production plan into capacity needs for key resources.

Rough-cut capacity planning is the process of converting the overall production plan into capacity needs for key resources. For a hospital, it means planning key resources for the demand schedule. While the day-to-day demand in healthcare systems is highly variable, the aggregate demand on a month-to-month basis can be predicted more precisely. When planning resources, hospital leaders generally consider two types of labor resources: full-time staff and contractors. By examining the census, an administrator should be able to determine, on an aggregate basis, the number of contactors needed during high-volume months. This approach is an example of rough-cut capacity planning. But many healthcare systems leave this planning until the need for additional resources arises. Because they have not paid enough attention to the required staffing levels to meet demand on an aggregate basis, these systems are forced to spend unnecessary costs to meet demand.

A hospital administrator may also use the daily census to assist in preparing workforce schedules on a weekly or daily basis. Exhibit 12.2 shows a spike in the system at VVH occurring from hour 13 to hour 19, which in most situations is the middle of the day. Many hospitals still schedule staff using standard morning, evening, and night shifts. Under that staffing model, VVH doctors and nurses are ending their shifts at the time of maximum demand on the system, resulting in increased potential for errors in handing off patients to new doctors, long patient wait times, and untimely completion of medical records.

A major cost savings can be gained for hospitals and clinics by simply matching the resources to the demand patterns in the system. In this case, staffing many doctors and nurses to overlap the peak times in the middle of the day is ideal.

EXHIBIT 12.2
Hourly Census at VVH in One Patient Care Unit



From an operations perspective, this problematic issue is easy to fix. However, in practice, several obstacles may emerge, such as contractual terms agreed to by unions and conflicting physician block scheduling.

Staff Scheduling

For minor schedule-optimization problems, where demand is reasonably known and staffing requirements can be estimated with certainty, mathematical programming (chapter 6) may be used to optimize staffing levels and schedules. As these problems increase in complexity, however, developing and applying a mathematical programming model becomes time and cost prohibitive. In those cases, simulation can be used to answer what-if scheduling questions, such as “What if we added a nurse?” or “What if we cross-trained employees?” See chapter 11 and the advanced-access section of this chapter for examples of these types of applications.

A simple example of this type of issue, and how to solve it using linear programming, is illustrated in the paragraphs that follow. (For solutions to more complex staffing issues using linear programming, see Matthews [2005] and Trabelsi, Larbi, and Alouane [2012].)

Solving Riverview Clinic Urgent Care Staffing

Nurses who staff Riverview Urgent Care Clinic (UCC), the after-hours urgent care facility of VVH’s Riverview Clinic, have been complaining about their schedules. They would like to work five consecutive days and have two consecutive days off every seven days. Different nurses prefer different days off and believe that their preferences should be accommodated on the basis of seniority, whereby the most senior nurses are granted their desired days off first.

Riverview UCC collects patient demand data by day of the week and knows how many nurses should be on staff each day to meet demand. Riverview UCC managers want to minimize nurse payroll while reducing the nurses' complaints about their schedules. They decide to apply **linear programming** to help determine a solution for this two-pronged problem. Target staffing levels and salary expense are shown in exhibit 12.3.

Linear programming
A mathematical technique used to find the optimal solution to a linear problem given a set of constrained resources.

First, Riverview UCC needs to determine how many nurses should be assigned to each of the seven possible schedules (Monday and Tuesday off, Tuesday and Wednesday off, etc.).

The goal is to minimize weekly salary expense, and the objective function is set up as follows.

Minimize:

$$(\$320 \times Su) + (\$240 \times M) + (\$240 \times Tu) + (\$240 \times W) \\ + (\$240 \times Th) + (\$240 \times F) + (\$320 \times Sa),$$

where *Su* is the number of nurses required on staff for Sundays, *M* is nurses needed Mondays, *Tu* is nurses needed Tuesdays, *W* is nurses needed Wednesdays, *Th* is nurses needed Thursdays, *F* is nurses needed Fridays, and *Sa* is nurses needed Saturdays.

The constraints are the following:

- The number of nurses scheduled each day must be greater than or equal to the number of nurses needed each day.
  $Su \geq 5$
  $M \geq 4$
  $Tu \geq 3$
  $W \geq 3$
  $Th \geq 3$
  $F \geq 4$
  $Sa \geq 6$

EXHIBIT12.3
Riverview UCC Target Staffing Level and Salary Expense

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Nurses needed per day	5	4	3	3	3	4	6
Salary and benefits per nurse-day	$320	$240	$240	$240	$240	$240	$320

- The number of nurses assigned to each schedule, where the schedules are denoted by a letter of the alphabet from A to G, must be greater than zero and an integer.

  Number of nurses for schedule A (B, C, D, E, F, or G) ≥ 0

  Number of nurses for schedule A (B, C, D, E, F, or G) = integer

Exhibit 12.4 shows the Excel Solver setup of this problem.

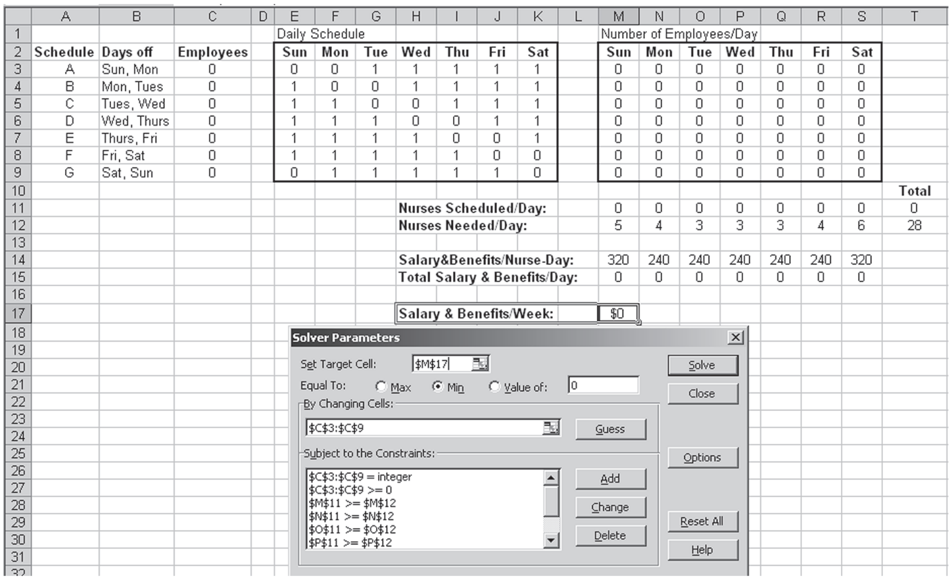
As illustrated in exhibit 12.5, Solver finds that the Riverview UCC needs to employ six full-time equivalent nurses and should assign one nurse to schedules A, B, C, and D; two nurses to schedule E; and no nurses to schedules F and G. The total salary expense with this optimal schedule is calculated as follows.

Minimize:

$$(\$320 \times 5) + (\$240 \times 4) + (\$240 \times 4) + (\$240 \times 4) + (\$240 \times 3) + (\$240 \times 4) + (\$320 \times 6) = \$8{,}080 \text{ per week.}$$

Next, Riverview UCC needs to determine which nurses to assign to which schedule on the bases of their preferences and seniority. Each nurse is asked to rank schedules A through E in order of preference. The nurses' preferences on a scale of 1 to 5, with 5 being the most preferred schedule, are then weighted by a seniority factor. Riverview UCC uses as the weighting factor the number of years a particular nurse has worked at the facility compared with the number of years the most senior nurse has worked there.

EXHIBIT 12.4 Initial Excel Solver Setup of Riverview UCC Optimization



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1					Daily Schedule								Number of Employees/Day							
2	**Schedule**	**Days off**	**Employees**		**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**		**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	
3	A	Sun, Mon	0		0	0	1	1	1	1	1		0	0	0	0	0	0	0	
4	B	Mon, Tues	0		1	0	0	1	1	1	1		0	0	0	0	0	0	0	
5	C	Tues, Wed	0		1	1	0	0	1	1	1		0	0	0	0	0	0	0	
6	D	Wed, Thurs	0		1	1	1	0	0	1	1		0	0	0	0	0	0	0	
7	E	Thurs, Fri	0		1	1	1	1	0	0	1		0	0	0	0	0	0	0	
8	F	Fri, Sat	0		1	1	1	1	1	0	0		0	0	0	0	0	0	0	
9	G	Sat, Sun	0		0	1	1	1	1	1	0		0	0	0	0	0	0	0	
10																				**Total**
11								**Nurses Scheduled/Day:**					0	0	0	0	0	0	0	0
12								**Nurses Needed/Day:**					5	4	3	3	3	4	6	28
13																				
14								**Salary&Benefits/Nurse-Day:**					320	240	240	240	240	240	320	
15								**Total Salary & Benefits/Day:**					0	0	0	0	0	0	0	
16																				
17								**Salary & Benefits/Week:**					$0							

Solver Parameters

Set Target Cell: $M$17

Equal To: ○ Max ● Min ○ Value of: 0

By Changing Cells: $C$3:$C$9

Subject to the Constraints:

$C$3:$C$9 = integer
$C$3:$C$9 >= 0
$M$11 >= $M$12
$N$11 >= $N$12
$O$11 >= $O$12
$P$11 >= $P$12

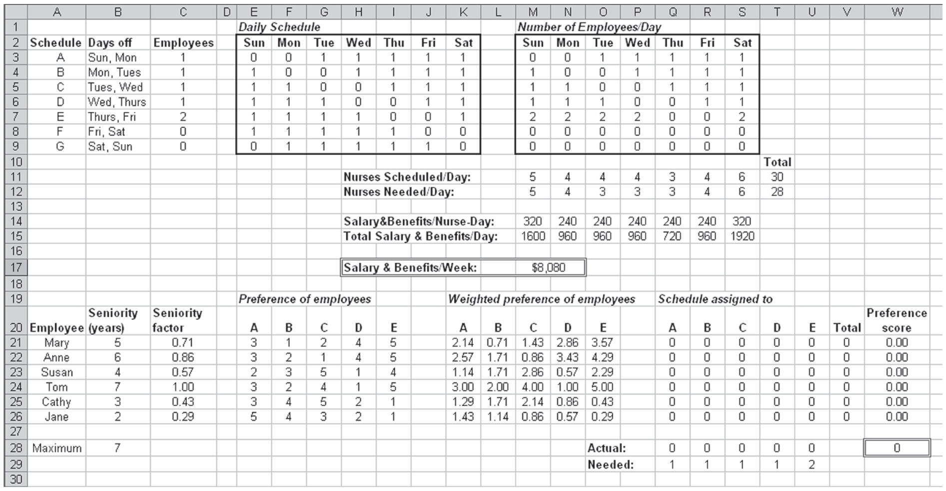


EXHIBIT 12.5 Riverview UCC Initial Solver Solution and Schedule Preference Setup

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1					*Daily Schedule*								*Number of Employees/Day*										
2	**Schedule**	**Days off**	**Employees**		**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**		**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**				
3	A	Sun, Mon	1		0	0	1	1	1	1	1		0	0	1	1	1	1	1				
4	B	Mon, Tues	1		1	0	0	1	1	1	1		1	0	0	1	1	1	1				
5	C	Tues, Wed	1		1	1	0	0	1	1	1		1	1	0	0	1	1	1				
6	D	Wed, Thurs	1		1	1	1	0	0	1	1		1	1	1	0	0	1	1				
7	E	Thurs, Fri	2		1	1	1	1	0	0	1		2	2	2	2	0	0	2				
8	F	Fri, Sat	0		1	1	1	1	1	0	0		0	0	0	0	0	0	0				
9	G	Sat, Sun	0		0	1	1	1	1	1	0		0	0	0	0	0	0	0				
10																				**Total**			
11								**Nurses Scheduled/Day:**					5	4	4	4	3	4	6	30			
12								**Nurses Needed/Day:**					5	4	3	3	3	4	6	28			
13																							
14								**Salary&Benefits/Nurse-Day:**					320	240	240	240	240	240	320				
15								**Total Salary & Benefits/Day:**					1600	960	960	960	720	960	1920				
16																							
17								**Salary & Benefits/Week:**					$8,080										
18																							
19					*Preference of employees*						*Weighted preference of employees*						*Schedule assigned to*						
20	**Employee**	**Seniority (years)**	**Seniority factor**		**A**	**B**	**C**	**D**	**E**		**A**	**B**	**C**	**D**	**E**		**A**	**B**	**C**	**D**	**E**	**Total**	**Preference score**
21	Mary	5	0.71		3	1	2	4	5		2.14	0.71	1.43	2.86	3.57		0	0	0	0	0	0	0.00
22	Anne	6	0.86		3	2	1	4	5		2.57	1.71	0.86	3.43	4.29		0	0	0	0	0	0	0.00
23	Susan	4	0.57		2	3	5	1	4		1.14	1.71	2.86	0.57	2.29		0	0	0	0	0	0	0.00
24	Tom	7	1.00		3	2	4	1	5		3.00	2.00	4.00	1.00	5.00		0	0	0	0	0	0	0.00
25	Cathy	3	0.43		3	4	5	2	1		1.29	1.71	2.14	0.86	0.43		0	0	0	0	0	0	0.00
26	Jane	2	0.29		5	4	3	2	1		1.43	1.14	0.86	0.57	0.29		0	0	0	0	0	0	0.00
27																							
28	Maximum	7													**Actual:**		0	0	0	0	0		0
29															**Needed:**		1	1	1	1	2		
30																							

The goal is to maximize the nurses' total weighted preference scores (WPSs), and the objective function is set up as follows.

Maximize:

Mary's WPS + Anne's WPS + Susan's WPS + Tom's WPS + Cathy's WPS
+ Jane's WPS

The constraints are the following:

- The assignment is binary, meaning that each nurse must be either assigned or not assigned to a particular schedule.
  Mary assigned to schedule A (B, C, D, or E) = 0 or 1
  Anne assigned to schedule A (B, C, D, or E) = 0 or 1
  Susan assigned to schedule A (B, C, D, or E) = 0 or 1
  Tom assigned to schedule A (B, C, D, or E) = 0 or 1
  Cathy assigned to schedule A (B, C, D, or E) = 0 or 1
  Jane assigned to schedule A (B, C, D, or E) = 0 or 1
- The number of nurses assigned to each schedule must adhere to the requirements established earlier.
  Number of nurses assigned to schedule A (B, C, or D) = 1
  Number of nurses assigned to schedule E = 2
- Each nurse can only be assigned to one schedule.
  Mary (Anne, Susan, Tom, Cathy, or Jane) A + B + C + D + E = 1

Exhibit 12.5 shows the Excel setup of this problem.

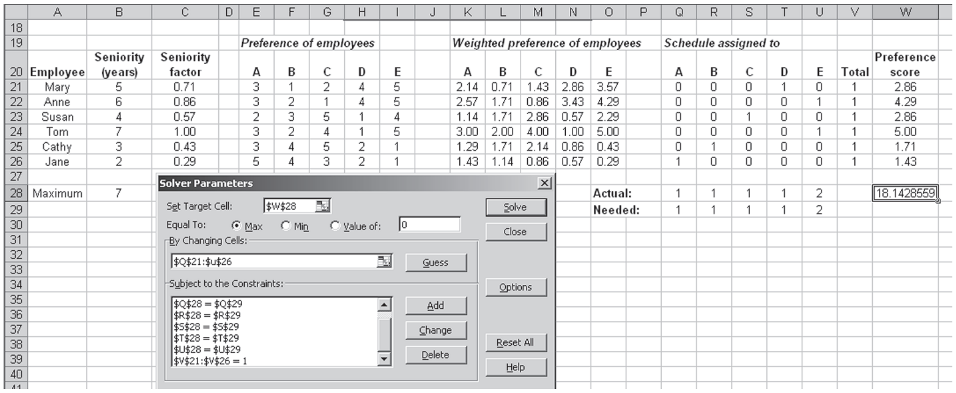
As shown in exhibit 12.6, Solver finds that Mary should be assigned to schedule D (her second choice), Anne to schedule E (her first choice), Susan to schedule C (her first choice), Tom to schedule E (his first choice), Cathy to schedule B (her second choice), and Jane to schedule A (her first choice). All of the nurses now have two consecutive days off every seven days and are assigned to either their first or their second choice of schedule. Note that even this simple problem has 20 decision variables and 41 constraints.

Job and Operation Scheduling and Sequencing Rules

Master production scheduling (MPS) is a technique used in most production-oriented environments that has direct application to the healthcare operations space. The concept behind MPS is to forecast needs for the future and build a schedule to fit those needs.

When building a master production schedule, time fences are set up to help avoid disruptions in the schedule. Typically, time fences depicted as “frozen,” “slushy,” or “liquid” are established to give the scheduling department information as to when a schedule can be adjusted. For example, a surgery center may aim for a frozen schedule for surgeries scheduled during the following week; a slushy schedule, where up to 20 percent may be adjusted, for surgeries scheduled two to three weeks in advance; and a liquid, or open, schedule for surgeries scheduled one month or more into the future. By freezing a schedule for a set period, the surgery center is able to avoid unnecessary interruptions. Interruptions in scheduling eventually lead to fewer surgeries for a variety of reasons, including the variance in time related to surgeries, extra setup time of surgery rooms, and general impact of changing surgeries at the last minute. To handle urgent surgeries when using MPS, a hospital should keep some capacity available for these situations. The net effect of this approach is increased output from the surgery because the variability associated with urgent surgeries does not affect the MPS.

EXHIBIT 12.6 Riverview UCC Final Solver Solution for Individual Schedules



			Preference of employees					Weighted preference of employees					Schedule assigned to						
Employee	Seniority (years)	Seniority factor	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	Total	Preference score
Mary	5	0.71	3	1	2	4	5	2.14	0.71	1.43	2.86	3.57	0	0	0	1	0	1	2.86
Anne	6	0.86	3	2	1	4	5	2.57	1.71	0.86	3.43	4.29	0	0	0	0	1	1	4.29
Susan	4	0.57	2	3	5	1	4	1.14	1.71	2.86	0.57	2.29	0	0	1	0	0	1	2.86
Tom	7	1.00	3	2	4	1	5	3.00	2.00	4.00	1.00	5.00	0	0	0	0	1	1	5.00
Cathy	3	0.43	3	4	5	2	1	1.29	1.71	2.14	0.86	0.43	0	1	0	0	0	1	1.71
Jane	2	0.29	5	4	3	2	1	1.43	1.14	0.86	0.57	0.29	1	0	0	0	0	1	1.43
Maximum	7											Actual:	1	1	1	1	2		18.1428559
												Needed:	1	1	1	1	2		

Solver Parameters

Set Target Cell: $W$28

Equal To: ● Max ○ Min ○ Value of: 0

By Changing Cells: $Q$21:u26

Subject to the Constraints:

$Q$28 = $Q$29
$R$28 = $R$29
$S$28 = $S$29
$T$28 = $T$29
$U$28 = $U$29
$V$21:$V$26 = 1

Job and operation scheduling views the problem of how to sequence a pool of jobs (or patients) through a particular operational activity. For example, a clinic laboratory constantly receives patient blood samples that need to be tested, and it must determine in what order it should conduct those tests. Similarly, a hospital typically has many patients waiting for their surgery to be performed, and it needs to decide the order in which those surgeries should occur.

The simplest sequencing problems consist of a pool of jobs waiting for only one resource to become available. Sequencing of those jobs is usually based on a desire to meet due dates (time at which the job is expected to be complete) by minimizing the number of jobs that are late, minimizing the average amount of time by which jobs are late, or minimizing the maximum late time of any job. Also desirable is to minimize the time jobs spend in the system or average completion time.

Various **sequencing rules**, also known as the queuing priority, may be used to schedule jobs through the system. Commonly used rules include the following:

Sequencing rules
Heuristic rules that indicate the order in which jobs are processed from a queue. Also known as *queuing priority.*

- *First come, first served (FCFS)*—Jobs are sequenced in the same order in which they arrive.
- *Shortest processing time (SPT)*—The job that takes the least amount of time to complete is first, followed by the job that takes the next least amount time, and so on.
- *Earliest due date (EDD)*—The job with the earliest due date is first, followed by the job with the next earliest due date, and so on.
- *Slack time remaining*—The job with the least amount of slack (time until due date or processing time) is first, followed by the job with the next least amount of slack time, and so on.
- *Critical ratio*—The job with the smallest critical ratio (time until due date or processing time) is first, followed by the job with the next smallest critical ratio, and so on.

When only one resource or operation is available through which the jobs may be processed, the SPT rule minimizes average completion time, and the EDD rule minimizes average lateness and maximum lateness. However, no single rule accomplishes both objectives. When jobs (or patients) must be processed via a series of resources or operations, with different possible sequencing at each, the situation becomes complex and applying a particular rule does not result in the same outcome for the entire system as for the single resource. Simulation may be used to evaluate these complex systems and helps determine optimum sequencing.

For a busy resource, the SPT rule is often applied. It allows completion of a greater number of jobs in a shorter amount of time than do the other rules, but it may result in some jobs with long completion times never being finished. To alleviate this problem, the SPT rule may be used in combination with other rules. For example, in some emergency departments (EDs), less severe cases (those with a shorter processing time) are separated from more severe cases and fast-tracked to free up examination rooms quickly.

For time-sensitive operational activities, in which lateness is not tolerated, the EDD rule is appropriate. Because it is the easiest to apply, the FCFS rule is typically used when the resource has excess capacity and no jobs will be late. In a Lean environment, sequencing rules become irrelevant because the ideal size of the pool of jobs is reduced to one and a kanban system (a form of FCFS) can be used to pull jobs through the system (chapter 10).

Vincent Valley Hospital and Health System Laboratory Sequencing Rules

A technician recently has left the laboratory at VVH, and the lab manager, Jessica Simmons, does not believe she can find a qualified replacement for at least one month. This situation has greatly increased the workload in the lab, and physicians have been complaining that their requested blood work is not being completed in a timely manner.

In the past, Jessica has divided the blood testing among the technicians and requested they complete the tests on an FCFS basis. She is now considering a different sequencing rule to satisfy the physicians. In anticipation of this change, she has asked each physician to enter a desired completion time on each request for blood testing. To investigate the effects of changing the sequencing rules, she analyzes, under various scheduling rules, the first five requests completed by one of the technicians. For five jobs, 120 sequences are possible for their completion. Exhibit 12.7 shows the time to complete each blood work sample and the time of completion requested by the physician.

Exhibit 12.8 indicates the order in which jobs will be processed and results under different sequencing rules, and exhibit 12.9 compares the various sequencing rules. The FCFS rule performs poorly on all measures. The SPT

EXHIBIT 12.7
VVH Laboratory Blood Test Information

Sample	Processing Time (minutes)	Due Time (minutes from now)	Slack	CR
A	50	100	100 – 50 = 50	100 ÷ 50 = 2.00
B	100	160	160 – 100 = 60	160 ÷ 100 = 1.60
C	20	50	50 – 20 = 30	50 ÷ 20 = 2.50
D	80	120	120 – 80 = 40	120 ÷ 180 = 1.50
E	60	80	80 – 60 = 20	80 ÷ 60 = 1.33

Note: CR = critical ratio.

EXHIBIT 12.8
VVH Laboratory Blood Test Sequencing Rules

Sequence	Start Time	Processing Time	Completion Time	Due Time	Tardiness
FCFS					
A	0	50	50	100	
B	50	100	150	160	
C	150	20	170	50	170 – 50 = 120
D	170	80	250	120	250 – 120 = 130
E	250	60	310	80	310 – 80 = 230
		Average	186		(120 + 130 + 230) ÷ 5 = 96
SPT					
C	0	20	20	50	
A	20	50	70	100	
E	70	60	130	80	130 – 80 = 50
D	130	80	210	120	210 – 120 = 90
B	210	100	310	160	310 – 160 = 150
		Average	148		(50 + 90 + 150) ÷ 5 = 58
EDD					
C	0	20	20	50	
E	20	60	80	80	
A	80	50	130	100	130 – 100 = 30
D	130	80	210	120	210 – 120 = 90
B	210	100	310	160	310 – 160 = 150
		Average	150		(30 + 90 + 150) ÷ 5 = 54
STR					
E	0	60	60	80	
C	60	20	80	50	80 – 50 = 30
D	80	80	160	120	160 – 120 = 40
A	160	50	210	100	210 – 100 = 110
B	210	100	310	160	310 – 160 = 150
		Average	164		(30 + 40 + 110 + 150) ÷ 5 = 66
CR					
E	0	60	60	80	
D	60	80	140	120	140 – 120 = 20
B	140	100	240	160	240 – 160 = 80
A	240	50	290	100	290 – 100 = 190
C	290	20	310	50	310 – 50 = 260
		Average	208		(20 + 80 + 190 + 260) ÷ 5 = 110

Note: All times shown in exhibit are in minutes. CR = critical ratio; EDD = earliest due date; FCFS = first come, first served; SPT = shortest processing time; STR = slack time remaining.

EXHIBIT 12.9
Comparison of VVH Blood Test Sequencing Rules

Sequencing Rule	Average Completion Time	Average Tardiness	Number of Tardy Jobs	Maximum Tardiness
FCFS	186	96	3*	230
SPT	148*	58	3*	150*
EDD	150	54*	3*	150*
STR	164	66	4	150*
CR	208	110	4	260

*Best values.

Note: All times shown in exhibit are in minutes. CR = critical ratio; EDD = earliest due date; FCFS = first come, first served; SPT = shortest processing time; STR = slack time remaining.

rule minimizes average completion time, and the EDD rule minimizes average tardiness. Under these two rules, three jobs are tardy and the maximum tardiness is 150 minutes. After considering these results, Jessica implements the EDD rule for laboratory blood tests to minimize the number of tardy jobs and the average tardiness of jobs. She hopes adopting this rule reduces physician complaints until a new technician can be hired.

Patient Appointment Scheduling Models

Appointment scheduling models attempt to minimize patient waiting time while maximizing utilization of the resource (clinician, machine, etc.) the patients are waiting to access. Soriano (1966) classifies appointment scheduling systems into four basic types: block appointment, individual appointment, mixed block-individual appointment, and other.

A block appointment scheme schedules the arrival of all patients at the start of a clinic session. Patients are usually seen FCFS, but other sequencing rules can be used in block appointment scheduling. This type of scheduling system maximizes utilization of the clinician, but patients may experience long wait times.

An individual appointment scheme assigns different, equally spaced appointment times to each patient. In a common modification of this type of system, different appointment lengths are available and assigned on the basis of the type of patient. This system reduces patient waiting time but decreases utilization of the clinician; in other words, increasing the interval between arrivals results in a reduction of both waiting time and utilization.

A mixed block-individual appointment scheme schedules a group of patients to arrive at the start of the clinic session, followed by equally spaced appointment times for the remainder of the session. This type of system can be used to balance the competing goals of increased utilization and decreased waiting time.

Finally, other appointment schemes are modifications of the first three types.

Simulation has been used to study the performance of various appointment scheduling models and rules. Although no scheduling rule or scheme has been found to be universally superior, the Bailey-Welch rule (Bailey and Welch 1952) performs well under most conditions. This rule schedules two patients at the beginning of a clinic session, followed by equally spaced appointment times for the remainder of the session.

Chow and colleagues (2011) demonstrate how to reduce the number of surgery cancellations by using an advanced computer simulation model to improve the allocation of open surgical slots in the appointment system. Using Monte Carlo simulation techniques, they increased surgical volume by more than 5 percent and reduced the number of overcapacity bed days by more than 9 percent.

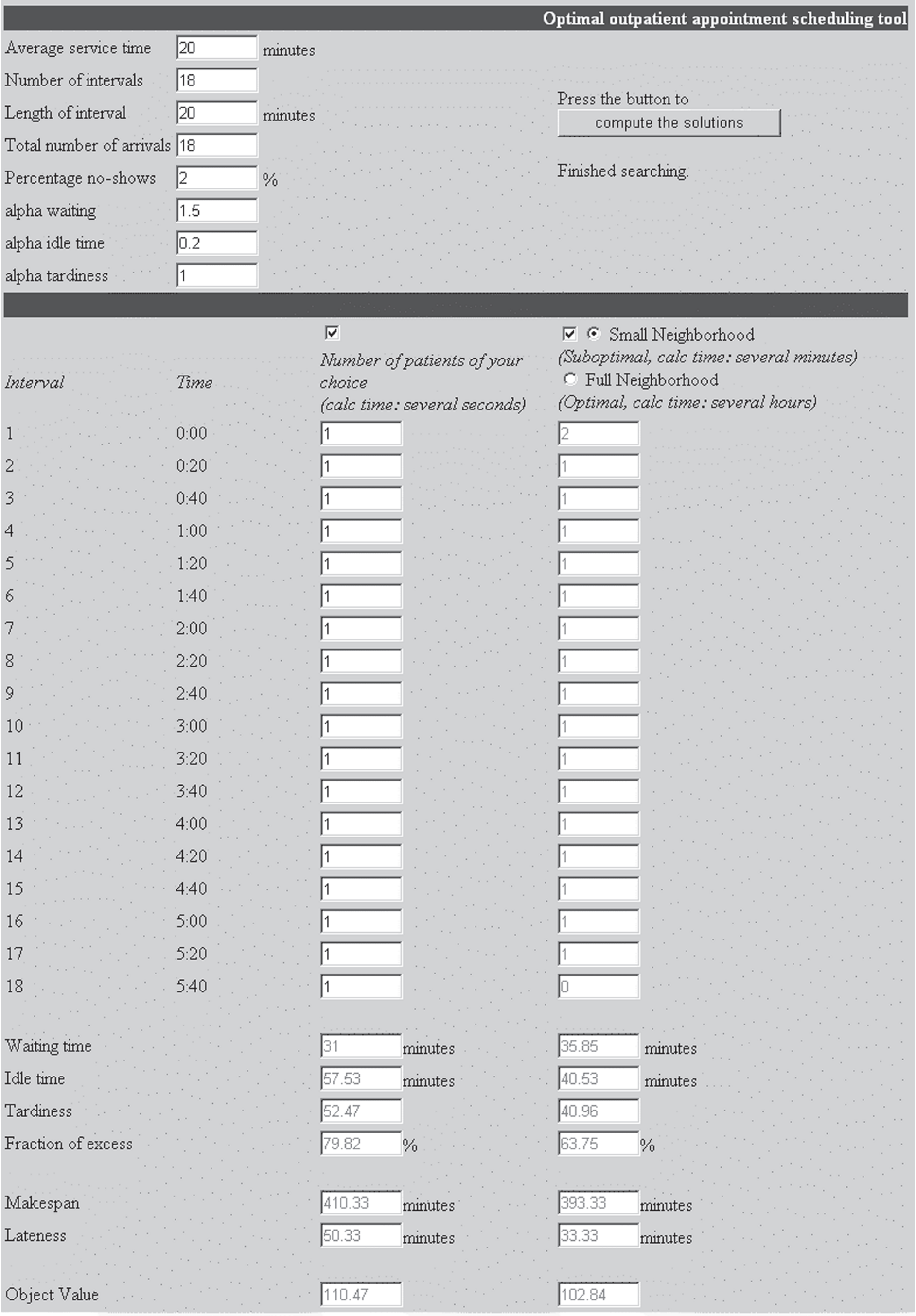
Kaandorp and Koole (2007a, 2007b) developed a mathematic model, called the Optimal Outpatient Scheduling tool, to determine an optimal schedule using a weighted average of expected waiting times of patients, idle time of the clinician, and tardiness (the probability that the clinician has to work later than scheduled multiplied by the average amount of added time). This tool uses simulation to compare the optimal schedule found using the model to a user-defined schedule.

Riverview Clinic Appointment Schedule

Physicians at VVH's Riverview Clinic typically see patients for six consecutive hours each day. Each appointment takes an average of 20 minutes; therefore, each clinician is scheduled to see 18 patients per day. The patient no-show rate is 2 percent. Currently, Riverview uses an individual appointment scheme with appointments scheduled every 20 minutes. However, clinicians have been complaining that they often have to work late but are idle at various points during the day. Riverview decides to use the Optimal Outpatient Scheduling tool (Kaandorp and Koole 2007b) to determine if another scheduling model can alleviate these complaints without increasing patient waiting time to an unacceptable level.

Exhibit 12.10 shows the results of this analysis when waiting time weight is 1.5, idle time weight is 0.2, and tardiness weight is 1.0. The optimal schedule follows the Bailey-Welch rule. Under this rule, patient waiting is increased by five minutes, but both idleness and tardiness are decreased. Riverview Clinic leaders do not believe that the additional waiting time is unacceptable and decide to implement this new appointment scheduling scheme.

EXHIBIT 12.10
Riverview Clinic Appointment Scheduling



Optimal outpatient appointment scheduling tool

Average service time: 20 minutes
Number of intervals: 18
Length of interval: 20 minutes
Total number of arrivals: 18
Percentage no-shows: 2 %
alpha waiting: 1.5
alpha idle time: 0.2
alpha tardiness: 1

Press the button to compute the solutions

Finished searching.

Interval	*Time*	☑ *Number of patients of your choice (calc time: several seconds)*	☑ ● Small Neighborhood *(Suboptimal, calc time: several minutes)* ○ Full Neighborhood *(Optimal, calc time: several hours)*
1	0:00	1	2
2	0:20	1	1
3	0:40	1	1
4	1:00	1	1
5	1:20	1	1
6	1:40	1	1
7	2:00	1	1
8	2:20	1	1
9	2:40	1	1
10	3:00	1	1
11	3:20	1	1
12	3:40	1	1
13	4:00	1	1
14	4:20	1	1
15	4:40	1	1
16	5:00	1	1
17	5:20	1	1
18	5:40	1	0
Waiting time		31 minutes	35.85 minutes
Idle time		57.53 minutes	40.53 minutes
Tardiness		52.47	40.96
Fraction of excess		79.82 %	63.75 %
Makespan		410.33 minutes	393.33 minutes
Lateness		50.33 minutes	33.33 minutes
Object Value		110.47	102.84

Source: Kaandorp and Koole (2007b). Copyright © 2007 Guido Kaandorp and Ger Koole.

Advanced-Access Patient Scheduling

Advanced Access for an Operating and Market Advantage

In the early 1990s, Mark Murray, MD, and Catherine Tantau, RN, were among the early adopters of **advanced-access scheduling** at Kaiser Permanente in Northern California. Their goal was to eliminate long patient waits for appointments and bottlenecks in clinic operations (Singer 2001). The principles they developed and refined have now been implemented by many leading healthcare organizations globally.

Advanced-access scheduling
A method of scheduling outpatient appointments that provides open time slots every day for seeing patients on the same day they request an appointment. Also known as *same-day scheduling.*

Because most clinics today use traditional scheduling systems, long wait times are prevalent and appointments may only be available weeks, or even months, into the future. The further in advance that visits are scheduled, the greater the fail (no-show) rate becomes. To compensate, providers double-book or even triple-book appointment slots. Long delays and queues occur when all the patients scheduled actually appear for the same appointment slot. This problem is compounded by patients who have urgent needs requiring that they be seen immediately. These patients are either worked into the schedule or sent to an ED, decreasing both continuity of care for the patient and revenue to the clinic. At the ED, patients are frequently told to see their primary care physician (PCP) in one to three days, further complicating the scheduling problem at the physician office.

Advanced access is implemented by beginning each day with a large portion of each provider's schedule open for urgent, routine, and follow-up appointments. Patients are seen when they want to be seen. This scheme dramatically reduces the fail rate, as patients do not have to remember clinic visits they booked long ago. Because no double or triple booking occurs, patients are seen on time and schedules run smoothly. Clinics using advanced access can provide patients with the convenience of walk-in or urgent care, with the added advantage of maintaining continuity of care with their own doctors and clinics.

Parente, Pinto, and Barber (2005) studied the implementation of advanced-access scheduling in a large midwestern clinic with a patient panel of 10,000. Following implementation, the average number of days between calling for an appointment and being seen by a doctor decreased from 18.7 to 11.8. However, the most significant finding was that 91.4 percent of patients saw their own PCP following implementation of the system, as opposed to 69.8 percent pre-implementation.

Implementing Advanced Access

Changing from a long-standing—albeit flawed—scheduling system to advanced access is challenging. However, an organization can increase its probability of success by following a few well-prescribed steps. In a study of large urban public hospitals, Singer (2001) developed the following methodology to implement advanced access.

Obtain Buy-In

Leadership is key to making this major change. The advanced-access system must be supported by senior leaders as well as providers. Touring other clinics that have implemented advanced access may help these groups understand how this system can work successfully.

For large systems, starting small in one or two clinical settings is best. Once initial operating problems are resolved and clinic staff are expressing positive feelings about the change, advanced access can be carefully implemented in additional clinics in the system.

Predict Demand

The first quantitative step in implementation is to measure and predict demand from patients. For each day during a study period, demand is calculated as the number of patients requesting appointments (today or in the future), walk-in patients, patients referred from urgent care clinics or EDs, and calls deflected to other providers. After initial demand calculations are performed, additional factors may be included, such as day of the week, seasonality, demand for same-day versus scheduled appointments, and even clinical characteristics of patients.

Predict Capacity

The capacity of the clinic needs to be determined once demand is calculated. In general, capacity is the sum of appointment slots available each day. Capacity can vary dramatically from day to day, as providers usually have obligations for their time in addition to seeing patients in the clinic. Determining whether a clinic's capacity can meet expected demand is relatively easy using Little's law (described in detail in chapter 11).

That said, true capacity may not be readily apparent. Singer (2001) reports that, prior to close examination, leaders at many public hospital clinics felt that demand exceeded capacity in their operations. However, several of these clinics were able to find hidden capacity in their systems by using providers effectively (e.g., by minimizing their paperwork) and converting storage space to examination areas.

Another opportunity to improve the capacity of a clinic is to standardize and minimize the length of visit times. A clinic with high variability in appointment times may find that it has many small blocks of unused time.

Assess Operations

The implementation of advanced access provides the opportunity to review and improve the core patient flow and operations in a clinic. The tools and techniques of process mapping and process improvement, particularly value stream mapping and the theory of constraints, should be applied before advanced access is implemented.

Work Down the Backlog

Working down the backlog is one of the most challenging tasks in implementing advanced access, as providers are required to see more patients per day than usual until they have caught up to same-day access. For example, each provider may need to work one extra hour per day and see three additional patients until the backlog is eliminated.

The number of days needed to work down a backlog can be determined using this equation:

Days to work down backlog = Current backlog ÷ Increase in capacity,

where current backlog equals the number of appointments on the books divided by the average number of patients seen per day, and increase in capacity is the new service rate (patients per day) divided by the old service rate minus 1.

Go Live

Once a clinic has completed the above steps, it is almost ready to go live with its advanced-access scheduling system. However, it must first determine how many appointment slots to reserve for same-day access. Singer and Regenstein (2003) report that public hospital clinics leave 40 percent to 60 percent of their slots available for same-day access while other types of clinics leave up to 75 percent of slots available.

Educating patients in anticipation of the shift to advanced access is important, as many will be surprised by the ability to see a provider the day they request an appointment. Many elderly patients may actually decline this option, as they may need more time to prepare for the appointment or arrange transportation to it.

No clinic operates in a completely stable environment, so prospectively developing contingency plans is useful. Contingencies can range from the unexpected, such as a provider being ill or called away on an emergency, to the predictable, such as increases in demand, as for routine physicals in the weeks preceding the start of school. Good contingency planning ensures the smooth and efficient operation of an advanced-access system.

Metrics for Evaluating Advanced Access

Gupta and colleagues (2006) developed the following set of key indicators that can be used to evaluate the performance of advanced-access scheduling systems:

- *PCP match*—percentage of same-day patients who see their own PCP
- *PCP coverage*—percentage of same-day patients seen by any physician
- *Wait time for next appointment* (or third next available appointment)—for example, if you are calling on Monday and an appointment is

available on Tuesday, Thursday, and Friday, the wait time for the third next available appointment is five days (Friday)

- *Good backlog*—appointments scheduled in advance because of patient preference
- *Bad backlog*—appointments waiting because of lack of slots

Most well-functioning advanced-access systems have high PCP match and PCP coverage. Depending on patient mix and preferences, the good backlog may be relatively large and still not be problematic, but a large or growing bad backlog can signal that capacity or operating systems in the clinic need to be improved.

Fears About Advanced Access and Their Resolution

Pointing out the realities of same-day scheduling can help reduce physicians' fears about change and help them make an effective adjustment to the new system. Gregg Broffman, MD, medical director of the 110-physician Lifetime Health Medical Group in Rochester and Buffalo, New York, whose group adopted same-day scheduling in the late 1990s, reported the following three common fears that physicians experience but that actually are unjustified (Olsen 2012):

- *Insatiable demand*. Physicians worry that opening their schedule will leave them swamped with work, but this is a false expectation. By carefully measuring and predicting supply and demand, advanced access ensures adequate coverage and can help determine the need to hire new clinicians to handle the workload.
- *Fewer encounters*. Use of same-day scheduling has been shown to decrease the number of annual encounters with individual patients. At the same time, it boosts the likelihood that patients will see their personal physician, rather than be worked in with the first available clinician. As a result, patients are more satisfied with their visits than they would be without advanced access. Furthermore, clinical outcomes rise while costs decrease, because a person's regular practitioner is less likely to order unnecessary tests or prescribe medication than is a clinician who is unfamiliar with the patient's history.
- *Lower revenue*. Decreased volume might suggest a dip in practice revenue, but the opposite has proven true. Clinicians who initially saw a 10 percent to 15 percent drop in encounters experienced about an 8 percent increase in relative value units, which are used to measure the robustness (or "dollar value") of an office visit. For example, when a diabetic patient makes an unplanned visit, physicians can look ahead to her next scheduled appointment and "max pack" the initial visit by

performing the future checkup that day. The visit can be coded at the higher level allowed by a more complicated encounter, and the max packing leaves an appointment open in two weeks to see a new patient.

Conclusion

Advanced-access scheduling is an efficient and patient-friendly method of scheduling the delivery of ambulatory care. However, implementing and maintaining this and other capacity management techniques are difficult unless leadership and staff are committed to their success.

Discussion Questions

1. What job sequencing rule do you see most often in healthcare? Why? Can you think of any additional job sequencing rules not described in this text?
2. How could advanced-access techniques be used for the following types of facilities?
   a. An ambulatory surgery center
   b. A freestanding imaging center
3. What are the consequences of using advanced access in a multispecialty clinic? How might these tools be applied to provide same-day scheduling?
4. Can advanced-access techniques be used with appointment scheduling schemes? Why or why not?

Exercises

1. Two of the nurses (Mary and Tom) at Riverview UCC have decided to work part time rather than their previous full-time schedule. Each prefers to work only two (consecutive) days per week. Once they become part-time employees, salary and benefits per nurse-day for these nurses will be reduced to $160 on weekdays and $220 on weekend days. Considering this savings, Riverview UCC can hire an additional full-time nurse if needed. Should Riverview UCC agree to the two nurses' request? If the clinic agrees, will additional nurses need to be hired? Assuming that part-time nurses and any new hires accept any schedule offered by Riverview UCC and that preferences for the

remainder of the nurses are the same as stated in the chapter, what new schedule would you recommend for each nurse?

2. The VVH radiology department currently uses FCFS to determine how to sequence patient X-rays. On a typical day, the department collects patient X-ray data, and these data are available on the book's companion website. Use the data to compare various sequencing rules. Assuming these data are representative, what rule should the radiology department adopt for sequencing, and why?

On the web at
ache.org/books/OpsManagement3

3. Use the Optimal Outpatient Scheduling tool (Kaandorp and Koole 2007b), provided on the companion website, to compare two appointment scheduling schemes—individual appointments and optimal scheduling—under the following assumptions. For the individual scheduling scheme, assume an 8-hour day that can be divided into 10-minute time blocks (48 time intervals), a 15-minute service time for patients, 24 patients seen according to the individual appointment scheme (a patient is scheduled to be seen every 20 minutes), and 5 percent no-shows. For the small neighborhood optimal schedule, assume a waiting time weight of 1, an idle time weight of 1, and a tardiness weight of 1. What are the differences in the two schedules? Which would you choose? Why? Now, increase the waiting time weight to 3 and recompute the small neighborhood optimal schedule. How is this optimal schedule different from the previous one? Finally, change the service time to 20 minutes and compare the individual appointment schedule scheme to the small neighborhood optimal schedule with waiting time weights of 1 and 3. Which schedule would you choose, and why?

On the web at
ache.org/books/OpsManagement3

4. A clinic wants to work down its backlog to implement advanced access. The clinic currently has 1,200 booked appointments and sees 100 patients a day. The physician staff have agreed to extend their schedules and can now see 110 patients per day. What is their current backlog, and how many days will it take to reduce it to zero?

References

Bailey, N. T. J., and J. D. Welch. 1952. "Appointment Systems in Hospital Outpatient Departments." *Lancet* 259: 1105–8.

Chow, V. S., M. L. Puterman, N. Salehirad, W. Huang, and D. Atkins. 2011. "Reducing Surgical Ward Congestion Through Improved Surgical Scheduling and Uncapacitated Simulation." *Production and Operations Management* 20 (3): 418–30.

Gupta, D., S. Potthoff, D. Blowers, and J. Corlett. 2006. "Performance Metrics for Advanced Access." *Journal of Healthcare Management* 51 (4): 246–59.

Institute for Healthcare Improvement (IHI). 2012. "Advanced Access: Reducing Waits, Delays, and Frustrations in Maine." Accessed October 4, 2016. www.ihi.org/resources/Pages/ImprovementStories/AdvancedAccessReducingWaitsDelaysandFrustrationinMaine.aspx.

Kaandorp, G. C., and G. Koole. 2007a. "Optimal Outpatient Appointment Scheduling." *Health Care Management Science* 10 (3): 217–29.

———. 2007b. "Optimal Outpatient Appointment Scheduling Tool." Accessed June 24. http://obp.math.vu.nl/healthcare/software/ges.

Matthews, C. H. 2005. "Using Linear Programming to Minimize the Cost of Nurse Personnel." *Journal of Healthcare Finance* 32 (1): 37–49.

Parente, D. H., M. B. Pinto, and J. C. Barber. 2005. "A Pre-Post Comparison of Service Operational Efficiency and Patient Satisfaction Under Open Access Scheduling." *Health Care Management Review* 30 (3): 220–28.

Singer, I. A. 2001. *Advanced Access: A New Paradigm in the Delivery of Ambulatory Care Services.* Washington, DC: National Association of Public Hospitals and Health Systems.

Singer, I. A., and M. Regenstein. 2003. *Advanced Access: Ambulatory Care Redesign and the Nation's Safety Net.* Washington, DC: National Association of Public Hospitals and Health Systems.

Soriano, A. 1966. "Comparison of Two Scheduling Systems." *Operations Research* 14 (3): 388–97.

Trabelsi, S., R. Larbi, and A. Alouane. 2012. "Linear Integer Programming for Home Health Care." *Lecture Notes in Business Information Processing* 100 (2): 143–51.