Section 3

Chapter 23

Optimizing patient flow through the emergency department

Kirk Jensen and Jody Crane

Key learning points

- The benefits that improving flow in your emergency department can provide.
- How tools developed in industry can help an emergency department manage and improve flow.
- How an emergency department can implement the necessary methods to optimize patient flow.

Introduction

One of the most important ways you can improve how well your emergency department (ED) works involves using a concept not limited to healthcare settings, in fact one commonly encountered by businesses of all kinds: flow. We can define flow in this sense as the movement of customers through a business operation. In the case of the ED, it is specifically the movement of patients through the department as they are evaluated, treated, and released or admitted to the hospital (Fig. 23.1). At certain points, small changes in the flow of patients can lead to long delays. But small changes you in turn make in managing that flow can also lead to dramatic cuts in the time patients spend in the ED. Making the movement of those patients more effective can thus pay big dividends in improving their satisfaction – and the quality of their care – as well as improving the satisfaction of your staff correspondingly. And this in turn can result in improved financial success for the hospital. In this chapter we will look at some ways to go about improving flow.

Matching demand and capacity

In simple terms, demand is how many of your resources are being used at a particular moment, and capacity is how many resources you have in relation to the number of patients. Matching demand to capacity is an important component of optimizing flow. When demand exceeds capacity, waiting times grow longer and patient frustration increases. Physicians, nurses, and technicians become stressed. When capacity exceeds demand, your resources – human and otherwise – are not being used effectively; in fact, they are being wasted. Patients coming into your department are unscheduled, the number fluctuating randomly, so matching the demand for services to the resources that provide them may seem a hopeless challenge. In fact, however, we have learned that the number of patients who will come to your ED is predictable.

The arrival of patients, plotted over time, follows a mathematical pattern known as a Poisson arrival pattern. Though specific numbers vary, this pattern holds generally for EDs everywhere.

Because we know what the pattern is, we can implement steps to match the capacity to the demand. Essentially there are two approaches: “smoothing” patient demand, or managing your resources to...
Section 3: Operational principles

Figure 23.1 The life cycle of a patient visit. The various stages of a visit to the ED constitute input, throughput, and output. How effectively a patient goes through the stages is flow – smooth or not.

Figure 23.2 The components of patient flow: door to doctor, doctor to decision, decision to disposition, disposition to discharge. Breaking an ED visit into its components and then analyzing the processes in each component presents opportunities to improve those processes and thereby improve flow.

Queuing

The system that patients move through involves a series or network of queues. If you have more than one patient waiting for a service, you have a queue (even one person waiting is technically a queue, but one person waiting usually does not result in problems in the ED). Queuing theory is the mathematical study of waiting lines, developed by engineers to provide models for predicting how systems that serve random arrivals rather than constant or scheduled
arrivals behave. The first work on queuing theory, which studied telephone traffic congestion causing delays that led to customers hanging up without being served, was published more than a century ago, so the phenomenon is well understood by this point.\(^1\)

An important implication of queuing theory is that as the use of service providers increases, the waiting increases. The problem is that it does not increase proportionally, but exponentially. In other words, when queuing systems with high degrees of variation reach high rates of use, they tend to go bad fast. At usage rates above about 85% of capacity, waiting times take off like a rocket lifting off from the earth (Fig. 23.3).

By those same mathematical principles, however, making small changes that affect the usage rate so that it drops below the takeoff point can have a dramatic impact on waiting times. When you act to match capacity to demand, you reduce the likelihood that queuing will reach that point.

**Variation**

We see variation all the time in the ED: patients have different illnesses or injuries requiring different levels of resources and amounts of time to assess and treat; healthcare providers have different levels of skill, training, and experience and take different amounts of time to perform similar actions. Variation can also cause an exponential increase in waiting times: if several patients arrive requiring extensive resources, delays can accumulate for other patients, and again a point is reached where waiting times take off.

The more variation can be smoothed, the more efficient the flow of patients will be. Some of the ways we will look at for matching demand to capacity also help manage variation.

**The psychology of waiting**

When waiting is inevitable, businesses have developed methods based on the psychology of waiting to manage those delays. These are built on several observations researchers have made about people waiting.\(^2\) Actions that make waiting *seem* shorter to patients include keeping them occupied, starting the process quickly, calming their anxiety, telling them how long their waiting will likely be, explaining why they are waiting, treating patients equitably in making them wait compared with other patients, and letting them wait with friends or family rather than alone. Another point to keep in mind is that people are more willing to wait when they perceive the service they are waiting for to be valuable. If we consider these actions, frequent communication clearly becomes a key. So does continual monitoring of waiting areas.

**Constraints**

We mentioned earlier that the patient’s experience in the ED involves a series of queues. When patients enter the ED, their experience consists of a series of interactions with healthcare workers – possibly as many as 20 queuing interfaces. To improve the interactions within this network of queues, you must understand the theory of *constraints*. A management theory introduced credit: Chuck Noon, PhD, UT PEMBA
Section 3: Operational principles

by Eliyahu Goldratt about 25 years ago, it defines a constraint as any resource, mindset, or policy that prevents an organization from moving closer to its goal. It classifies a system’s resources as “bottlenecks” and “non-bottlenecks” based on the demand placed on them and their capacity. A bottleneck resource has a capacity less than the demand placed on it. In the ED, for instance, triage often forms a bottleneck, when too many patients arrive for the triage nurse to effectively segment without queues forming. A physician is similarly frequently a bottleneck, when more patients appear than that doctor can diagnose and treat efficiently enough to handle the volume. A non-bottleneck resource, on the other hand, has a capacity equal to or greater than the demand placed on it. Changing processes to improve a non-bottleneck resource may make them more efficient but will not improve throughput time in the ED; a hospital, for example, might add a patient care coordinator in the ED – often an effective step. But if adding a coordinator does not improve throughput time as a whole in a particular ED, then the lack of one was not a bottleneck in that ED. Working on bottleneck resources and improving them so that their capacity can meet the demand is the only way to move an organization closer to its goal. Working on non-bottleneck resources does not improve flow – it is a waste of time.

To manage constraints, an ED needs to repeatedly remove or reduce the constraint constituting the greatest relative bottleneck. In the queuing framework, managing bottlenecks means moving resources from high states of use to lower states of use, thus realizing the exponential benefits in the reduction of waiting achieved by improving highly utilized providers.

The operational details: how to manage flow

Once you understand the principles that underlie flow management, you can put into practice procedures based on them. Because the concepts are interrelated, the procedures often are as well; using a method to match capacity with demand, for example, can affect variation as well. In this section we examine a number of specific ways to improve flow in your ED.

Measuring patient demand

To match your capacity to the demand requires measuring patient demand and then developing a system to meet it. You want the answers to these questions: How many patients are coming into your ED each month? Each day? Each hour? When are they coming during the day? As the Poisson pattern indicates, these numbers are predictable. To obtain them, track data in your ED over time and analyze the pattern, then match your staffing coverage at particular times – of the day, of the week, of the year – to the pattern. Determine not only the numbers but what services these patients will require at various times; plot use of lab and radiology services at different times, for instance, and plan your service capacity accordingly.

In tracking the data, break the patient arrivals into categories such as chief complaint, number of emergency service arrivals, severity levels, and ancillary utilization. Knowing the patterns in these categories enables you to predict specific numbers for your department. Also predictable is that variation will occur, so in addition to matching service capacity to patient demand in advance by tracking the data, develop a plan for how you will respond when the demand rises unexpectedly. An ED can, for example, arrange for patients who need beds and are going to be admitted to board in hallways of inpatient units rather than in the ED, and it can arrange to use vacated space of units that operate only during daytime hours for patient overflow. Training staff to perform some functions of another position if necessary is another method that can alleviate high demand – having technicians start intravenous administration, or having a triage nurse who does not have any patients at the moment help discharge patients from the nearby fast track – and track variation in detail to discern patterns in it as well, for they do occur even in random variation. For a more detailed discussion on predicting demand and handling it, see Mayer and Jensen’s work on flow.

Shaping demand

You may think you can do nothing to affect patient demand, and that all your efforts to improve flow should be directed toward managing service capacity. In fact, though, you can take some steps to shape demand. One is to introduce and publicize scheduled appointments for less acute complaints. Another is to promote vaccination for the flu among your patients and in your community. Decreasing the number of people likely to come down with an illness shapes the resulting demand for your department’s services.
Segmenting patient flow

Dividing patients into segments based on the severity of their condition allows you to effectively handle multiple patients simultaneously, and to smooth flow. In emergency medicine, there are three basic streams of patients based on their condition – low acuity, mid acuity, and high acuity. A system based on the Emergency Severity Index (ESI), which provides five levels in which to sort patients, further refines the streams.5 (Other five-level triage systems are equally effective for segmenting patient flow by severity and resource needs: see Chapter 19.) By segmenting streams of patients, you create unique pathways customized for each particular stream and can design individual processes for each. Patients in one stream follow the same basic steps throughout their visit. The basic streams may divide further: for instance, the high-acuity stream may subdivide into substreams of acute heart attack, acute stroke, acute traumatic injury, and acute life-threatening infection with end-organ compromise. Each of these substreams would have a well-defined process with separate steps.

Patients who fall into the two least severe categories (4 and 5 in the ESI segmentation model), and many in the middle category (3) can be fast-tracked. Fast track, we emphasize, is something you do, not somewhere you put patients (for more on this concept, see Jensen and colleagues’ work on flow6). Your fast-track process should enable you to quickly and efficiently assess and treat patients whose complaints are not severe. You should make sure you have appropriate fast-track staff to meet demand, particularly in the hours when you know the demand will be highest.

Triage: removing a potential bottleneck

Look closely at your triage process, because it can easily become a bottleneck to patient flow. As with fast track, triage is something you do and not somewhere you put patients. Consider whether some aspects of triage would fit more effectively at another stage in the process. Triage should be brief: gather information sufficient to direct the patient into a segment based on severity of condition.

EDs traditionally have had four different queues that all patients enter: triage, physician assessment, treatment, and discharge. Each queue makes the patient wait. An alternative approach is to use a one-queue process, in which patients receive assessment, treatment, and discharge in one series of steps, without additional waiting times. Team triage is an effective way to implement this process.7 The team consists of several providers, including nurses and doctor, as well as mid-level providers, such as physician assistants or nurse practitioners, who can diagnose and treat patients more fully than a nurse can but not as fully as a physician can, and lab and radiology technicians, who quickly assess and register patients, then send them to a treatment area or a results-waiting area (both nearby). The least severe and the most severe streams can be efficiently sorted into the proper channel for treatment, either the fast track or a bed. The mid-level patients can be treated in the team triage area and discharged. This is one process that helps prevent a bottleneck from forming in triage.

Optimizing your resources

If you want to improve flow in your ED (Fig. 23.4), having the right staff, the right staff mix, and the appropriate amount of staffing to meet projected demand is critical. Hiring and motivating excellent staff is the beginning point. As we noted in regard to triage, forming an effective team from physicians, nurses, mid-level providers, and technicians is equally important. And as we emphasized in regard to the fast track and predicting demand in general in the ED, so is having enough resources in place to handle that demand and provide high-quality care. Fast-tracking patients is one of the most effective actions you can take to improve flow – but the fast-track process must include sufficient resources and space.

The results waiting area is where patients who do not need beds can stay while awaiting results of lab work or tests. Monitor this area closely and keep in mind the practices emphasized in our discussion of the psychology of waiting. Providing magazines or televisions, for example, helps keep patients occupied and reduces anxiety.

We mentioned the importance of sorting patients into the right channel. Determining who needs a bed is a key component. Not all patients in the ED need beds. You should consider beds as resources, and directing patients who do not need one into the fast track, or to results waiting, or for treatment in team triage rather than into a bed is a means of avoiding bottlenecks. To borrow from business again, we can point to restaurants and the concept of turning over tables. The more you turn over beds,
the more patients you can treat with the same high level of care – you are increasing capacity without adding any resources. We like to use the phrase “keep your vertical patients vertical” – in other words, patients with less acute needs should keep moving without occupying beds. Make sure you have processes established with specific criteria for who needs a bed and who does not.

Obviously, your staff members are resources. To optimize their effectiveness, they should concentrate on carrying out only the tasks that they are best suited to perform. Doctors, in other words, should not be carrying out activities that secretaries or scribes can do. Nurses should not be entering data on computers. In addition, persons in these roles should constantly perform those functions they are suited for and intended to perform. Physicians should be kept busy seeing patients.

Case study 23.1. The virtual bed

Responding to a nearly limitless volume of patients in its urban setting, the ED staff at Metropolitan Hospital implemented the concept of a physician in triage, starting a shift from 11 a.m. to 7 p.m. consisting of a physician, a technician, and a nurse in a project known as “Virtual Bed.” The team operated directly behind triage, and instead of placing the patient back in the waiting room the triage nurses would move the patient to the treatment area. There, the physician would perform a brief exam, order the necessary diagnostic tests, and begin any basic treatment that could be initiated at that point. The patient would then be sent back out to the waiting area where the lab and radiology would begin the necessary tests, placed in a set of rooms labeled “Virtual Beds.” When the testing was complete, the patients would be moved to the back and discharged by a treating physician.

The benefits of this system were undeniable. The patient would be seen by a physician within minutes of arrival, was quickly made “in-process,” and had pain or other pressing issues addressed promptly. The waiting time had turned into “results waiting” time and the patient length of stay dropped.

Keeping track of your patients

Tracking arrival patterns is not the only type of data collection you should do. Keep track of patients’ progress through the system while they are moving through it. A real-time dashboard enables you to accomplish this goal, either a simple whiteboard with patients listed along with where they are in the process or a more complex computer system with clearly visible monitors showing these details and
more. A patient-care team should monitor progress and results; a secretary or technician should physically accompany patients to their next location in the process when results are available.

Tracking patients as they flow through the system allows you to tell quickly when bottlenecks begin to occur, so that you can act to reduce them before backlogs cause demand to reach the takeoff point. It also helps you predict your future needs for service capacity.

Facilitating admissions

One cause of delays in the ED is the admission process to the hospital for those patients who need to enter. To address this issue, keep track of times from the decision to admit a patient to when he or she actually is admitted and record the reasons for the delays. Having this information will help you understand why delays are occurring, and it also helps you predict how many patients will need admission and what units they will go to (since this flow of patients also follows a Poisson pattern). Use the results of your tracking to communicate in advance likely numbers of patients who will be admitted, and the times of admission, to the hospital staff members responsible for placing patients.

If your ED is boarding significant numbers of patients, then flow can be impaired seriously. Taking steps to reduce this boarding is one of the most effective actions you can take to improve flow. Improving the admission process as we just discussed is one way. Another is to convince the hospital to implement a process whereby patients waiting for rooms can be moved to the hallways of the units where they will be admitted rather than remaining in the ED.

Collaborating with ancillary services

Improving flow means working not only with your team in the treatment area but also with ancillary services such as the radiology department, the lab, the pharmacy, and even housekeeping. Cutting delays even by just a little with each step can reduce bottlenecks. For example, having an x-ray technician note visibly with a card attached to a patient’s chart “x-ray back” and putting the chart in the ED doctor’s inbox can reduce waiting time for that patient. Cumulatively, such reductions in waiting time can cut overall delays in the department significantly.

Work with housekeeping in admitting patients. Housekeeping staff should be kept involved in the projection of likely admissions and in identifying empty beds as soon as they are available. Effective communication and coordination in this aspect of operations can have a significant impact on boarding in the ED and thus is one way to reduce bottlenecks and improve flow noticeably.

Examine the current processes for interaction between the ED and these other sections and try to identify similar steps, even small ones, that can improve flow and reduce delays by adjusting processes.

Case study 23.2. Super track, super results

The fast track in the ED at Acme Medical Center was typical of many in that if the main area of the ED was overrun, sick patients would overflow into the fast track, and the flow of lower-acuity patients would essentially cease. The average length of stay of low-acuity patients at the time was over two and a half hours. A flow-improvement team designed a new super track, moving supplies it needed out of the supply storage system and into a cart in the area and moving critical equipment such as printers and computers to the area to virtually eliminate the need for the providers to leave. The team assessed the arrival rates and service rates and projected length of stay to arrive at the right combination of physician assistants, nurses, and beds based on the time available to see patients. Team members ran several pilots of the new concept and proved quickly that they could see 60% more patients with 33% less staff and 75% fewer beds. The team was able to reduce the length of stay of ESI Level 4 and 5 patients from an average of two and a half hours to about 50 minutes.

Putting the theories and methods to work

In implementing methods to improve flow, you will find that using an established system of operational management will help you do so more effectively. A number of such systems exist, such as Lean management, Six Sigma, and total quality management. Health care recently has seen Lean management adapted for its specific needs.

Lean management

Lean management is an approach to operations management resulting from an intensive study of why Toyota has been so successful for the last half-century. The approach focuses on processes and on developing
people. Its implementation into health care started in aspects that were direct crossovers from manufacturing, such as inventory management. More recently, it has become more prevalent in clinical operations such as EDs and operating rooms. Lean management is characterized by several key components: creating patient value, eliminating waste, promoting flow, making continuous improvement, and developing people.

We noted in our discussion of segmenting patient flow that patients in one stream should follow the same steps. This idea incorporates the Lean principles of creating patient value and eliminating waste. All of the activities in a patient process can be classified as value-added or non-value-added. Value-added activities are those steps that move the patient closer to wellness, services that patients desire, services done right the first time. Physician examination is a good example of a value-added step. This is, after all, the main reason a patient comes to the ED in the first place. Other examples would be use of labs and radiographs—only if they change the management of the patient’s course through the system for the better or otherwise enhance the patient’s well-being—psychological well-being, for example, of knowing he or she does not have a fracture. Non-value-added activities are those steps that do not create patient value and that patients do not want. “Traditional” triage is a classic example of a non-value-added activity. No patient comes to the ED to be interviewed by a nurse in order to determine how long the patient can wait. At the extreme, triaging a patient when there are idle physicians and available beds is an obvious example of a wasteful step.

The Lean concept of waste addresses how to manage non-value-added activities. The goal is to eliminate as much waste as possible (you cannot eliminate it entirely). Lean management identifies a number of forms of waste; in healthcare settings, these translate to such actions as:

- Overprocessing or doing more than the patient needs, such as multiple providers asking redundant questions or insisting on triage protocols rather than getting the patient in front of the provider and ordering only what is necessary.
- Overproduction, such as generating reports that go unread.
- Not doing an activity right the first time (known as defects), resulting in having to go back and ask the patient a question again because a provider did not document the encounter as it occurred, or having to order a test or medication because the wrong one was ordered initially.
- Not using your human resources to actively engage in problem solving and process improvement.

The goal of improvement efforts under Lean management is to create value and eliminate waste within processes, continually increasing the amount of value-added activity relative to the amount of non-value-added activity, and thus increasing the value-added ratio.

Organizations using Lean management place frontline employees in control of improving the system, as these people are the ones who will have the best answers. Such organizations empower these staff members by teaching them tools to help them improve their workplace. These organizations typically implement over 90% of employee improvement suggestions, and continually examine and improve processes.

For more information on the development and implementation of Lean management, see the work of Womack et al. and Spear. For its specific implementation in health care, see the white paper on the subject from the Institute for Healthcare Improvement.

**Plan–do–study–act cycles**

Improvement efforts, to succeed, require tests of their effectiveness and subsequent adjustment. The “plan–do–study–act” (PDSA) cycle is a key component of your process to improve flow. You should run small testing cycles often rather than large ones infrequently. Under the PDSA process, you propose one change and test it for a short time.
The “plan” phase involves defining the objective and diagramming how it should be attained; specifically, what do you want to do in this cycle, who should do it, and where and when should it be carried out? Consider what you are trying to learn from the test.

The “do” phase involves implementing the plan, during which you should document any problems and observations. You organize the data you have collected during the test and start to analyze it. “Study” means analyzing fully what you learned in the test and summarizing the lessons. “Act” means deciding what changes you need to make in this plan if you are going to implement it permanently.

You then repeat the cycle, testing your change on a wider scale and assessing any adjustments, and develop new ideas to test. This process helps ensure that ideas that seem promising actually work to improve flow; testing them on a small scale and continuing to test and adjust as a result of what you learn will lead to changes that actually are effective in improving flow. Such a process helps keep your staff engaged in the improvement efforts and active in examining processes and seeking better ways to operate your ED.

Hiring and staffing

As we have emphasized, the team of healthcare professionals you put in place in your department is critical in improving flow. Getting the right team for the ED begins with focusing on hiring the right people. Once they are in place, making sure they receive the training they need, evaluating their performance, and coaching them for better performance are key actions. If people are what make organizational change effective, having the right people is crucial. Having motivated staff members empowered to examine processes, suggest changes, and help implement and test those changes will go a long way toward achieving your goals for improving flow.

The impact of smoothing flow

Improving flow brings multiple benefits to your ED. It creates a more pleasant environment for both patients and staff, increasing customer satisfaction and employee fulfillment. In doing so, it provides a safer department and enables your staff to consistently provide high-quality care. Achieving improved flow allows you to increase service capacity without necessarily adding new resources, by making existing resources more effective, adding value to the steps in your processes, and reducing waste in your system. All of these improvements also, on a practical level, will increase the profitability of your hospital.

References


Case study 23.3. Building in flow

Nightingale Hospital opened a new facility 10 miles from its main facility in 2009. During the design process, administrators asked the ED team to review the ED design in the new facility. Experienced at implementing projects to improve flow, the team was able to architecturally hardwire the principles it had developed at the main site, making flow in the new department more efficient by complementing the operational design with the layout. It contains twice as many triage rooms as originally planned, and an internal waiting room was added between the intake area and the main ED to facilitate keeping vertical patients vertical. In addition, the rooms were stocked with all of the supplies to treat most patients. The new facility was budgeted to see about 25 000 annual patient visits, but immediately handled volumes amounting to about 40 000 annual visits. The average time from arrival to seeing a doctor is 30 minutes. Length of stay averages a little over two hours, and the walk-out rate has been less than 1% every month since it opened. Patient satisfaction is usually above the 90th percentile. Carts replaced specialized rooms, so that any procedure could be done in any room.
Section 3: Operational principles


