CHAPTER 2

The First Focusing Step: Identify the Bottleneck

OBJECTIVES

- Learn how to identify the bottleneck
- Test your hypothesis about the bottleneck

I (Chris) drive my kids to school, and on the way, there is a stretch of road under construction. Traffic backs up as three lanes merge into two, and then into a single lane. It can take over 20 minutes to move four blocks. I like to use this time to impart wisdom to my children and wax eloquently about flow—if I listen hard enough, I can actually hear their eyes rolling. Undaunted, I explain how we in the cars are inventory piling up behind the bottleneck—an available lane—and I remind my children how lucky they are to have a father who is so wise and generous with his knowledge. At this point, my teenager in the front seat looks at me, sighs loudly, and turns up the radio; there we all sit, piled up behind the bottleneck, waiting to merge.[1]

To identify a bottleneck, look for the biggest pileup of inventory: The bottleneck will be the resource in front of the pileup. In a hospital setting, inventory can refer to patients in a specific part of the hospital, such as patients in the ED or the ICU, or it can comprise all of the hospital's patients. If there is a pileup of patients waiting to be seen by a physician in the ED, then the bottleneck may be physician staffing. If there is a pileup of patients who have been medically cleared and are waiting to be discharged from an inpatient service, then the bottleneck may be a lack of skilled nursing or assisted living facilities to accept patients upon discharge.

In the initial stages of process improvement, the biggest bottleneck will likely be obvious. In fact, the biggest pileup will probably be an issue that everyone has been complaining about for years, and the bottleneck will be sitting right in front of that pile.

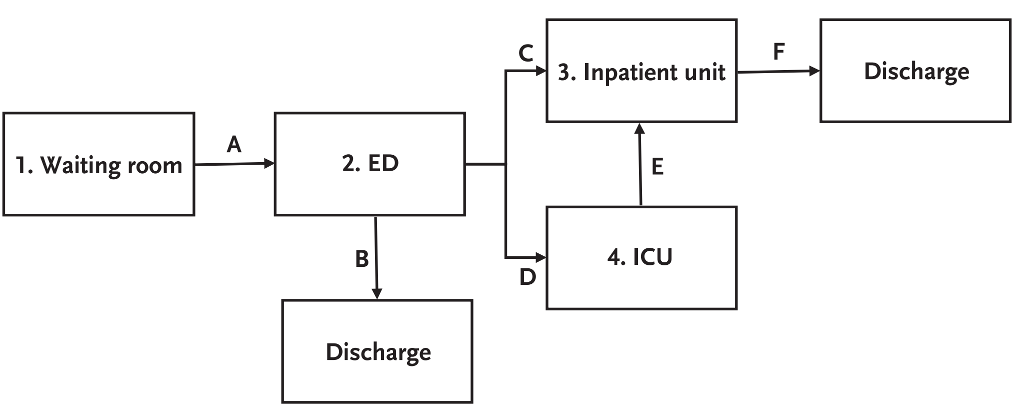
Whether obvious or not, it is helpful to go through the formal process of identifying the bottleneck. This ensures that all stakeholders working to improve flow agree on the bottleneck. It is also good practice for subsequent iterations of the 5FS, in which new bottlenecks may be less obvious.

The easiest way to identify a bottleneck is to build a process map of patient flow through the entire hospital system (exhibit 2.1). This basic, high-level map provides a good starting point for identifying the bottleneck. Note that analysis should be done during typical patterns of flow and not in extreme or unusual situations where an unexpected event is causing excessive delays. Use exhibit 2.1 to find the bottleneck by looking for a pileup of inventory: Where is the patient census higher than it should be?

Waiting Room

Is there a pileup of patients in the waiting room? There is no set number of patients that defines a pileup other than some quantitative number of patients that everyone decides is too many. It can be a function of a high rate of patients who leave without being seen or a prolonged "door to bed" or "door to provider" time that signifies excessive wait times. If there is rarely a pileup of patients in the waiting room, the bottleneck is likely external to the system—a function of the marketplace. Under those circumstances, the hospital has the capacity to treat more patients in its ED without stressing the system and should focus on growing patient volumes.

Exhibit 2.1: Process Map of Hospital Patient Flow



Emergency Department

If there is a pileup of patients in the waiting room, is there also a pileup of patients in the ED? A pileup in the ED occurs when all ED beds are full and there is no availability to move patients out of the waiting room. Visual cues of an ED pileup include inpatient boarders taking up ED beds and gurneys occupied by ED patients lining the hallways. Additionally, the ED length of stay for admitted or discharged patients may be too long and ambulance diversion rates too high (recall that ambulance diversion is usually symptomatic of a broken hospital system rather than indicative of an ED-specific problem).

If there is not a pileup of ED inventory, then the bottleneck exists at arrow A in exhibit 2.1—the outflow of patients from the waiting room to the ED. Causes could include inefficient registration or triaging processes, as well as delays in rooming patients such as when a charge nurse purposely leaves ED beds empty. A more granular process map is then made to focus on the specific steps at A, and that map is used in subsequent focusing steps. If, however, there are plenty of open, staffed beds in the ED, then a project designed to discharge patients from the ED more quickly will likely have no significant effect on overall flow. This is an example of how an hour saved on a nonbottleneck is a mirage.

If there is a pileup of ED inventory, then you need to discover whether ED lengths of stay are long for discharged patients, boarders, or both. There may be a pileup of ED patients who will ultimately be discharged from the ED or an accumulation of admitted patients (boarders) waiting for a hospital bed in either an inpatient unit or the ICU (or both).

If the ED pileup consists of discharged patients, then the bottleneck resource exists at B in exhibit 2.1, and the hospital should focus on the process of evaluating, treating, and discharging patients from the ED. It would be useful to create a more detailed process map tracking a patient's movement from the time the patient reaches the ED up to and including discharge. If patients are also backing up from the ED into the waiting room, it may be helpful to include the waiting room in the process map, even though the bottleneck may not exist in the waiting room. Reorganizing workflow in the waiting room may be an important strategy to break the ED bottleneck in subsequent focusing steps.

Inpatient Units

If there is a pileup of ED boarders waiting to move to an inpatient unit, is there also a pileup of inventory on the inpatient units? The hospital needs to identify whether, for example, all of the inpatient beds in a unit are full or whether there are empty beds that are staffed, clean, and ready to accept patients. The hospital should also determine whether its average inpatient length of stay (when adjusted for acuity) is longer than published benchmarks.

If there is not a pileup on the inpatient units, then the bottleneck exists at C in exhibit 2.1. Subsequent focus should be on the process of moving patients from the ED to inpatient units, including how beds are ordered and assigned, how physician handoff is conducted, how nursing reports are given, and how patients are transported to the accepting unit.

If there is a pileup of inpatients, are they predominantly patients who are medically ready to leave the hospital (i.e., patients for whom a discharge order has or can be written)? If the answer is yes, it indicates a delay, either in writing the discharge order or in the patient physically leaving the hospital despite being discharged. If patients are waiting for the treating physician to write a discharge order, then the delay exists upstream of the discharge order, and the physician may need to start his or her day earlier or round on potential discharges first. If, on the other hand, patients have discharge orders but are unable to leave the hospital—for example, they are waiting for family members to pick them up or for a room at a rehab or skilled nursing facility to become available—then the delay exists downstream of the discharge order. In either scenario, the bottleneck exists at F in exhibit 2.1.

If the pileup is not the result of delays before or after discharge orders have been written, then the focus of improvement must include the time during which the patient is actively receiving treatment and has not yet been cleared for discharge. Are inpatient lengths of stay longer than published benchmarks suggest they should be? Benchmarks allow a hospital to compare its own length of stay to standards based on the performances of other hospitals (adjusted for patient volumes and acuity). If patients are not medically cleared for discharge within the expected time frame, then subsequent strategies must address patient care from the time patients are admitted until the time they receive discharge orders. Questions to ask during this part of the process can include: Are patients typically responding to treatment as expected and ready for discharge within that predicted time frame, or are they staying longer than expected? Are patients receive imaging studies, social work evaluations, physical therapy, and medication teaching in a timely fashion? Are nosocomial infections or iatrogenic complications keeping patients in the hospital longer than anticipated?

Intensive Care Unit

If there is a pileup of ED boarders waiting to move to the ICU, is there also a pileup of patients in the ICU? In other words, look to see whether ICU beds are always full. If there is not a pileup of ICU patients, then the bottleneck exists at D in exhibit 2.1, and the process for improvement is analogous to examining C.

If there is a pileup of ICU patients, it is possible that all of the patients need ICU-level care. If so, the hospital should examine how its ICU lengths of stay compare to published benchmarks, similar to the inpatient exercise above. If lengths of stay are longer than expected, subsequent strategies should focus on patient care delivery while in the ICU and similar questions asked about patient response to therapy, timely ancillary services (e.g., labs, imaging), and nosocomial infections. If lengths of stay in the ICU are not longer than expected, it may be that the ICU does not have enough beds and the hospital needs to add capacity (this focusing step, elevating the bottleneck, is discussed in subsequent chapters).

On the other hand, the pileup may be the result of patients boarding in the ICU while waiting for an available inpatient (non-ICU) bed. If inpatient beds are available, the bottleneck exists at E in exhibit 2.1. If inpatient beds are not available, the next steps are analogous to those laid out earlier for inpatient units. In the event of competition for an inpatient bed between an ICU patient ready to transfer out and an ED boarder, the available bed should usually go to the ICU patient first (assuming that the ICU is full—otherwise, there is no urgency to transfer the ICU patient out).

In Short . . .

In short, look for a configuration where a significant number of patients are waiting on a unit followed by a relatively small number of patients waiting on the subsequent unit. Such a scenario indicates the likely location of the current bottleneck.

WHERE DO WE GO FROM HERE?

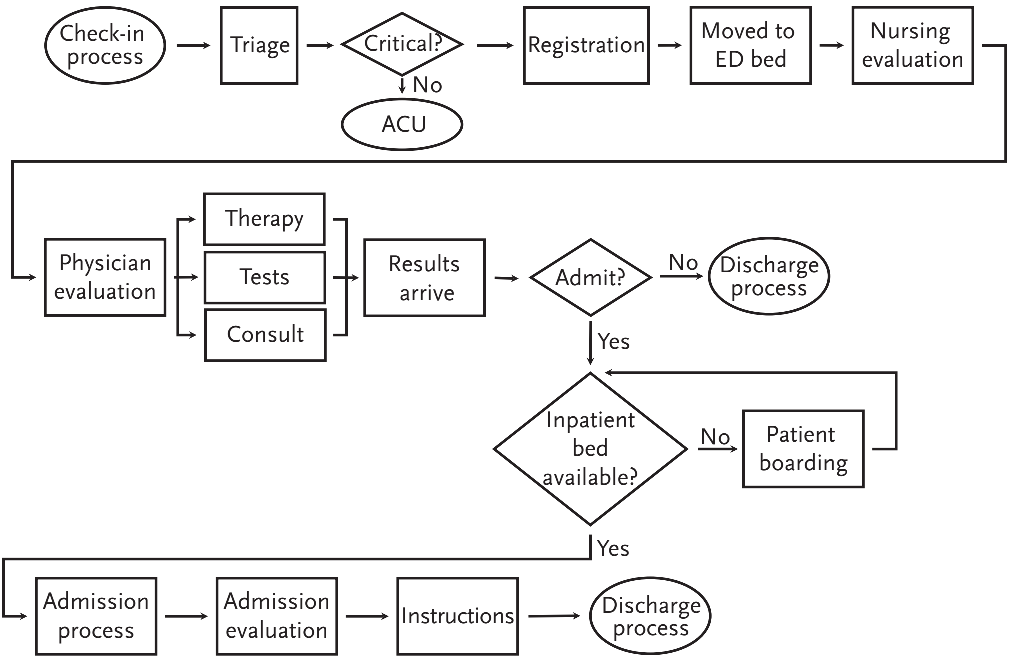
After you have identified the bottleneck, try creating a process map that details patient flow on either side of the bottleneck. If the bottleneck resource is in the ED, for example, create a detailed process map of the components of patient flow leading into and out of the ED (i.e., patients in the waiting room trying to get into the ED and patients in the ED trying to leave the ED). Exhibit 2.2 presents a more granular process map of patient flow through the ED.

While exhibit 2.1 provided a useful starting point for identifying the bottleneck, exhibit 2.2 provides the necessary detail to identify the bottleneck itself.

CONFIRM YOUR BOTTLENECK: THE SNIFF TEST

After you have identified a potential bottleneck, test that hypothesis by envisioning how the system would change if the bottleneck had unlimited capacity. If the hypothesized bottleneck was ED physician staffing, for example, and the imagined system now has ten times the number of ED physicians working

Exhibit 2.2: Detailed Process Map of Patient Flow Through Emergency Department



at any given time, determine whether flow would improve dramatically for the whole system. (Practically speaking, we recognize that adding unlimited capacity is probably not feasible and could actually make the problem worse by increasing inventory.) If the hypothesized move helps significantly, and if it creates a new bottleneck in the system, you have probably correctly identified the bottleneck.

Let's apply what we have learned to a case study.

CASE STUDY

A Level 1 trauma center averaged 60 hours of ambulance diversion each month. The CEO estimated that every hour of ambulance diversion cost the hospital $10,000, generated bad publicity, and denied services to the local community. The hospital applied TOC to identify the bottleneck in its system: For a month, whenever the ED went on ambulance diversion, the charge nurse recorded the reason (exhibit 2.3).

Based on this data, where is the hospital's bottleneck?

(continued)

Exhibit 2.3: Reasons for Ambulance Diversion

Exhibit 2.3 shows that the most common reasons for ambulance diversion are holding admits (boarding inpatients in the ED) and high ED volume (which comprises predominantly boarded patients); this is the pileup of inventory. The bottleneck is a resource somewhere in front of this pileup, and to find it, we must next focus on the process steps downstream of the pileup. Doing so requires a more granular process map of the downstream steps, specifically C, F, and 3 in exhibit 2.1. By focusing on these steps, we will identify why the boarded patients are waiting in the ED.

If there is a delay moving patients out of the ED, then the bottleneck exists in C (not enough transporters, for example, or delays in nurses giving and receiving report). In this case, there should be plenty of cleaned, staffed inpatient beds ready for new patients. A more granular map of the steps involved in C will help identify the specific bottleneck. If it is transporters, then we would see that the boarded patients all have inpatient rooms ready, that report has been called, and that they are simply waiting for someone to physically wheel them out of the ED. The sniff test would confirm that if we added ten times more transporters and all of the boarded patients immediately left the ED, then transportation is the bottleneck. If, however, more transporters do nothing to improve boarding, then we must look elsewhere.

If there is instead a delay in inpatient rooms becoming available, then the bottleneck is likely in 3 (there are no available inpatient beds because there are not enough inpatient nurses to staff them, for example, or perhaps the beds are all filled because the inpatient physicians haven't rounded on their patients and written discharge orders) or in F (for example, the inpatient beds are all filled with patients who are ready for discharge but they need an available skilled nursing facility and there aren't any available). The bottleneck could also involve the ICU, but practically speaking, this is less common as most admitted ED patients do not go to the ICU.

If the bottleneck exists at 3 or F, then the sniff test would look at how ED boarding changes with an unlimited number of staffed inpatient beds. If ED boarding suddenly disappears, then the bottleneck is somewhere on the inpatient side and a more granular process map of the inpatient steps will further narrow the search.

SUMMARY

- To identify the bottleneck, first look for the biggest pileup of patients (inventory) in the hospital or healthcare system.
- The bottleneck is the resource sitting right in front of the pileup.

- After identifying the bottleneck, create a granular process map of the steps adjacent to the bottleneck—those that flow into or feed the bottleneck and those that flow immediately away from or out of the bottleneck.
- Teenagers don't appreciate how wise we are.

NOTE

1. Incidentally, you know how there always seem to be construction workers at these sites walking around with hardhats and orange vests, but it's not clear what exactly they are doing? They could be considered nonbottleneck resources! I suspect that if you added more of these construction workers, we'd still be sitting in traffic just as long; remember that an hour saved on a nonbottleneck resource is a mirage.