Operational Margin: The Critical Final Pathway in Patient Flow

Joseph Guarisco, MD FAAEM FACEP
Operations Management Committee, Co-Chair

I am one of the early developers and a huge proponent of a patient flow concept known as *split flow*, described in *Common Sense* in the three-part series "Cracking the Code."[1-3] As early as ten years ago, I and others were thinking about this issue. Patient flow pioneers include Mary Ellen Bucco, MBA and Kevin Roche, PhD at Banner Health in Phoenix; Chris DeFlitch, MD at Penn State; Jody Crane, MD MBA in his early days at Mary Washington; and Jeff Finkelstein, MD at Hartford Healthcare in Connecticut. All experts in the field, including me, have been exposed to dozens if not hundreds of emergency departments (EDs) around the country and have participated in the dismantling of poorly performing workflow systems and the installation of what we considered to be our proprietary, best practice workflow scheme. However, I must admit that I have seen less efficient, less cost-effective, and more complex workflows succeed in terms of managing patient throughput. So what gives? Why does one workflow system succeed and another fail? The answer is *operational margin*, which I believe is the critical final pathway in any successful workflow.

What do we know about this topic? There are three noteworthy reviews of the subject. Joe Twanmoh, MD MBA, president of SG2 Consulting in Baltimore, and co-chairman of the AAEM Operations Management Committee, has written and presented extensively on what he calls "Myths in Operations Management," in which he reviews the literature on workflow systems and attempts to draw some conclusions on whether or not there is a breakthrough workflow best practice.[4] There is also the ACEP Task Force on Crowding's "Emergency Department Crowding High-Impact Solutions," commented on by Rick Bukata, MD, in his review article in *Emergency Physician Monthly*, "ED Throughput a Fixable Problem."[5,6] The ACEP task force divided throughput solutions into three categories: high-impact, no impact, and low-impact but expensive. High-impact solutions focus on creating bed capacity in the ED by creating bed capacity in the hospital — moving admitted patients out of the ED faster by discharging inpatients earlier in the day and by load-leveling surgical cases. The entire list is below.

High Impact

- Move admitted patient to hallways upstairs
- Discharge inpatients before noon
- Load level elective admissions and surgical cases

No Impact

- Build a bigger ED
- Discharge units
- Hospitalist co-management
- Ambulance diversion

Lower Impact and Expensive

- Bedside registration
- Fast Tracks
- Observation units
- Physician in triage
- Cancel surgeries
- Scribes
- POCT lab
- Advanced triage
- Adding nurses and support staff

In his commentary on the findings of the ACEP task force, Rick Bukata says "we put a man on the moon; surely we can shorten the wait in the emergency department." In my frequent talks on the concept of throughput, I take a more direct approach by asking, "You're telling me you have a patient waiting, with the provider ten the feet away, and you can't seem to get the two together in a reasonable amount of time. Is that your engineering problem?" Bukata's listing of throughput solutions is below.

- Provider at triage
- Advanced triage protocols
- Pull till full
- APPs
- Scribes
- Fast Tracks
- Super Tracks
- Split flow
- Production incentivized providers
- Point of care testing
- Full capacity protocols

Like Dr. Bukata, I am worried that after 35 years of practice we will end our careers with the issue of capacity management still unsolved. Despite appearing to identify a list of solutions, does the Holy Grail of patient flow still elude us? Eugene Litvak, PhD, CEO of the Institute for Healthcare Optimization in Newton, Massachusetts, has detailed the fatal flaw most managers make in attempting to solve the capacity and supply-demand issues that face emergency departments and entire hospitals. Dr. Litvak says, "As long as patient flow is ignored you cannot solve overcrowding, and as long as you have excessive patient volume and excessive workload you are going nowhere."[7] Ignoring mathematical realty dooms you to failure.

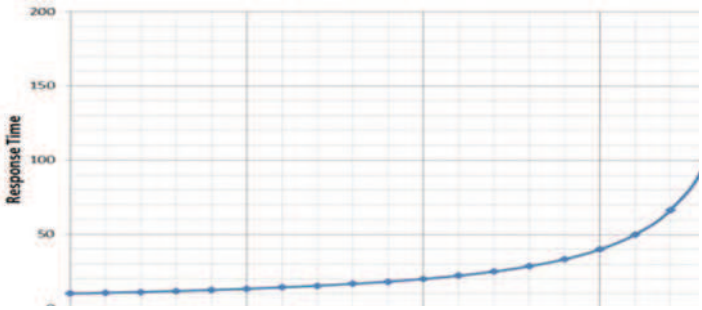
What then is the critical final pathway to efficient patient flow? What is the conceptual approach that manages capacity and delivers consistently smooth patient flow? It is operational margin. A workflow system without operational margin cannot succeed.

To explain, let's look at two distinct environments where the concept of operational margin has been mastered. The first is the fire station. Firefighters are available 24 hours a day — housed in the fire station 99% of the time rather than out fighting fires. The community's goal is to staff that fire station at such low utilization that should a fire develop, there is a 99% probability that those firefighters will be available to fight it. The second is the freestanding emergency department. Freestanding EDs generate patient satisfaction scores over 98%. They utilize providers at a rate of roughly one patient per hour, versus the average ED rate of two patients per hour. Additionally, bed utilization at a freestanding ED averages 700 patients per bed per year, versus an average of 2,000 patients per bed per year in traditional EDs. You can see that freestanding emergency departments operate at much lower resource utilization rates than traditional emergency departments.

Both the fire station and the freestanding ED demonstrate the concept of operational margin. Both create a high probability that the critical resource will be available when needed — a concept that is largely ignored throughout health care, but an industry failure that is now being capitalized on by urgent care clinics and freestanding emergency departments.

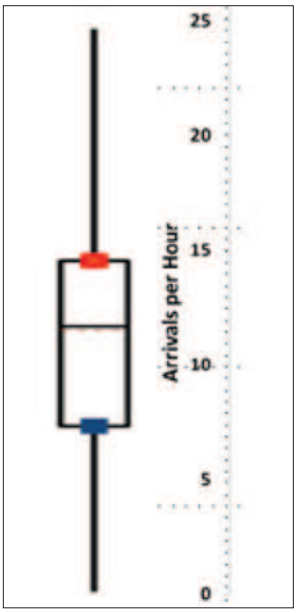
Continued on next page

If you study the classic response time-utilization curve below, you will see that response time as a function of utilization does not degrade in a consistently *linear* way — as utilization approaches 100% it degrades *exponentially*.



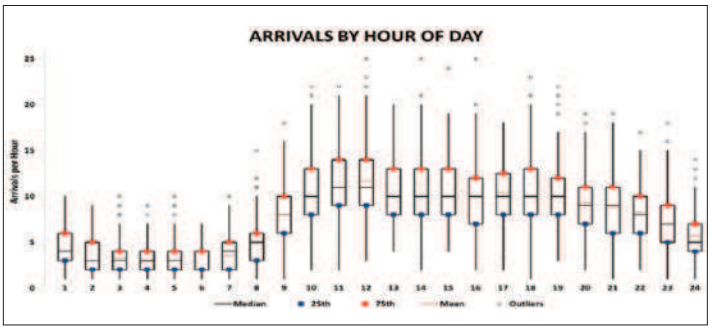
Fire departments and freestanding emergency departments operate on the early, flat part of that curve. Traditional emergency departments — and for that matter, hospitals in general — operate on the steeper part of the utilization curve. This is not a problem if resource demand is flat, consistent over the course of time, and demonstrates very little variance. An example of that might be an outpatient clinic, where demand is filtered (leveled) by the appointment system. This is obviously not the case in the ED, where resource demand (patient flow) is extremely variable. Unless you are content to let performance and service metrics deteriorate, you must keep your ED operating on the lower portion of that curve. That is mathematical reality, and ignoring reality doesn't make it go away. Living on the steep part of the response time-utilization curve means you have zero operational margin, leaving no room for the management of even slight variance. Unfortunately this is standard operating procedure in emergency departments, which are notorious for their lack of surge capacity (long waiting times). Lack of understanding and fear of escalating costs are the most common reasons for hospital administrators' failure to address this problem.

Another way to look at the probability of meeting variable resource demand is to use box plots.



The box plot to the left represents hundreds of thousands of data points and plots a distribution of those points. It shows that the average rate of patient arrivals per hour is 12, there is a 50% probability that between eight and 15 patients will arrive per hour, and a 90% probability that between one and 24 patients will arrive per hour. If you want enough nurses, doctors, and beds to give you a high probability of meeting patient demand most of the time, then you should plan on 15 patients per hour — or even better, 24 — rather than the average of 12. Staffing for the average is equivalent to operating on the steeper part of the utilization curve shown earlier. Any variance in demand towards the upper end of the box will result in an exponential deterioration in response time, with rapidly worsening service performance. In other words, planning and staffing for the 50th percentile of demand rather than the 75th or 90th means you have eliminated any operational margin (surge capacity).

I'll begin to wrap up with the box plot below, which shows 100,000 ED patient arrivals graphed over time of day. Staffing to the average (the horizontal black line in each box), rather than to the 75th (red dot) or 90th percentile (top of the vertical black line), depletes your operational margin and pushes you further (higher) up the slope of the response time-utilization curve shown earlier. That's not good.



A workflow system will succeed only if it creates adequate operational margin. A fast-track, a provider-in-triage, a split flow system, or even a traditional ED that incorporates none of these workflow processes can provide for excellent throughput if it creates enough operational margin. Building a bigger ED and hiring more docs and nurses, though inefficient, can provide great throughput. A provider-in-triage can also provide operational margin if there are enough providers, enough people to draw blood, and enough chairs or recliners for these patients. This was shown in a recent article in *Common Sense*.[8] A split flow system, creating virtual space by minimizing bed use and employing advanced practice clinicians for the low acuity stream, can achieve the same results at much lower cost.[9,10] The critical factor is whether or not these processes are supported and resourced to avoid over-utilization and meet demand most of the time. Operational margin is the key.

References

1. Guarisco J, Cracking the Code: Fixing the Crowded Emergency Department, Part 1 – Building the Burning Platform, *Common Sense*, September/October 2013.
2. Guarisco J, Cracking the Code: Fixing the Crowded Emergency Department, Part 2 – Creating the Analytic Model, *Common Sense*, November/December 2013.
3. Guarisco J, Cracking the Code: Fixing the Crowded Emergency Department, Part 3 – Implementing the Solution, *Common Sense*, January/February 2014.
4. Twanmoh J, Guarisco J, Myths in Operations Management, 2015 AAEM Podcast Series. http://www.aaem.org/publications/podcasts/emergency-medicine-operations-management
5. Solberg L, Asplin B, Weinick R, Magdid D, ACEP Task Force on Emergency Department Crowding: High Impact Solutions, *Annals Of Emergency Medicine* 2003.
6. Bukata R, ED Throughput: A fixable Problem, *Emergency Physician Monthly*, November 2014.
7. MacDonald I, Fiercehealthcare.com, November 2014.
8. Yun B, Raja A, White B, Mass General in the PIT, *Common Sense*, Sept/Oct 2015.
9. Guarisco J, Samuelson D, Rx for the ER, *Operations Research Management Science Today*, October 2011.
10. Crane JT, Noon C, *The Definitive Guide to Emergency Department Operations*, Productivity Press, March, 2011.