Crowding, Karenina, and Complexity: Patient Flow in Evolving Health Care Ecosystems

David A. Petrie, MD*

Tolstoy begins Anna Karenina with this observation: “Happy families are all alike; every unhappy family is unhappy in its own way.” The retrospective observational study by Cheng et al.¹ in the current issue of this journal looking at the factors associated with failure to meet emergency department length of stay targets confirms again²-⁵ that perhaps the opposite is true of patient flow in emergency departments (EDs): “Unhappy EDs are all alike, but every happy ED is happy in its own way.” That is to say, the most common and most significant factor in what is known as overcrowding (a word on semantics in a second) is, by far, the boarding of admitted patients in the ED.

In Cheng’s study,¹ the odds ratio of missing ED length of stay targets for admission to the intensive care unit (ICU) was a whopping 364 (upper 95% confidence interval [CI] 307) if the bed request duration was greater than 6 hours, and 217 (upper CI 154) if there was “access block” in the ICU. During this time, the average hospital occupancy was 97.6% (so that almost half the time it was over 100%; and if this hospital is like others, the average occupancy of general/community medicine beds, often the destination of the majority of ED admits, may be higher than that). We know from a recent paper from Quebec⁶ that increases in hospital bed occupancy are associated with an increase in 30-day adverse outcomes even after adjustment for patient and ED characteristics. Commenting on that paper, an editorial entitled “Just Another Crowding Paper”⁷ laments that the finding of a 10% increase in ED bed occupancy ratio being associated with a 3% increase in death might be seen as old news. Yet, if this mortality rate was discovered to be due to an unrecognized complication of a new drug, or a new infectious disease outbreak, it would be front-page news.

Cheng et al.¹ clearly described by observation what we already known with mathematical certainty given the nonlinear nature of queues:⁸ if random variation in demand cannot be smoothed to any significant degree (i.e., in EDs), and capacity is fixed, and the ability to tolerate wait times is limited (another paper out of Ontario⁹ showed that wait times correlate with mortality in a dose–response relationship that suggests causation), then high occupancy rates in the hospital (i.e., greater than 85%⁸,¹⁰), leads to the ED being used as a holding area for admissions, displacing incoming ED patients into hallways and waiting rooms, which equals missed ED length of stay targets, unhappy EDs, and, most alarmingly, poor patient outcomes. We are all alike in that way.

That is not to say that ED leaders and front-line staff can throw up their hands, point fingers, and blame all dysfunction on “the system.” Happy EDs (in happy hospitals) have problem-solved and adapted to the myriad dynamic and interdependent factors that contribute to ED overcrowding. Like seeds that evolve to become drought-tolerant over time when there is a lack of water, EDs have evolved over time to be more efficient at rationalizing their operations when there is a lack of beds. And that is the real issue in this overcrowding/access block discussion: the problem is not “overcrowding” in the vernacular sense of the word; it is the inability to access an appropriately staffed and equipped ED stretcher, within a reasonable amount of time, relative to the need of the next sick and/or complex frail patient in the queue. There could be hundreds of low-acuity patients waiting, and this might be a problem for the fire marshal, but it does not meaningfully...
contribute to the real problem. However, if there are two ambulances with undifferentiated and potentially sick patients waiting for only one available ED stretcher, patient safety is compromised.

It is clear that previous attempts to model ED overcrowding as a simple problem (like baking a cake gone wrong) have inadvertently perpetuated the real problem. The “common-sense” assumption that overcrowding in the ED is caused by too many people in the recipe, and that the obvious way to fix that is to reduce the number of people who “shouldn’t be there,” has led to years and years, and millions and millions of dollars, of doomed-to-fail solutions. A recent meta-analysis of interventions aiming to reduce the use of EDs by low-acuity patients has shown a disappointing amount of futility (though not universally so) in this endeavour. We have failed by solving the wrong problem.

Recognizing this perhaps before most, Asplin et al. in 2003 published the very helpful and almost universally accepted paper entitled “A Conceptual Model of Emergency Department Crowding.” This clearly articulated description led to a more complicated approach to what is really a broader systems issue. This characterized ED operations as having three distinguishable components; input, throughput, and output. Its huge success was moving the conversation (and therefore the practical solutions) from the simple era (too many low-acuity patients) into one with far more insight and effectiveness. The goal of the input–throughput–output conceptual model, which has largely been accomplished, was to “provide a practical framework on which an organized research, policy, and operations management agenda can be based to alleviate ED crowding.” And indeed, in some jurisdictions, significant improvements have resulted from this.

It is beyond the scope of this editorial to review all of the successful solutions to ED overcrowding, but improvements in ED length of stay have been achieved with strategies that optimize inpatient bed-hour utilization and therefore reduce boarding in the ED. Likewise, strategies that optimize emergency bed-hour utilization have also led to improvements in flow in many situations. Happy EDs can achieve happiness in different ways. Efficiencies can be gained with many innovative flow strategies, but only up to a certain threshold; then it becomes a capacity problem. No matter how lean an ED/hospital becomes, efficiencies will be overwhelmed by the occupancy/capacity issue at a certain point.

Unfortunately, if the Asplin model itself is interpreted too rigidly in a linear (input–throughput–output) production line, cause–effect, command and control, systems–machine manner, this may contribute to a more insidious obstacle to practical solutions; that ED overcrowding is merely a complicated (and not a complex) problem. Health systems and hospital flow do not function like machines (“well-oiled,” or not), and if we are stuck in this frame of reference, the problem will remain intractable. Patient flow is a wicked problem and is better thought of in the context of complex adaptive systems methodology (which work differently and are less predictable than machines). Patient flow in today’s healthcare ecosystems involves multiple independent agents, following different and dynamic rules (not a single set of stable system rules), with interdependent impacts on outcomes, coevolving adaptations (work-arounds, gaming), nonlinear tipping points, negative and positive feedback loops, self-organizing and emergent patterns, and no single locus of control—all characteristics of a complex adaptive system.

Experienced and more broadly informed leader-practitioner decision-makers have noted important layers to the problem (and therefore corresponding solutions) that are not explicitly represented in the input–throughput–output model, such as inadequate accountability in managing sequential and/or interdependent queues, ineffective advocacy, disengaged senior leadership, absence of public reporting and targeted incentives/disincentives, and other aspects of challenged organizational culture. Again, the Asplin model is not exclusive of these and other factors that may have significant implications for ED overcrowding. But that is one of the inherent problems of schematically representing complexity: the multiple scientific, economic, social, cultural, and psychological elements of a health (eco)system do not lend themselves to straightforward two-dimensional images, and input–throughput–output flow diagrams and descriptions.

As the ED has conceptually moved from the backdoor to the front-door of the hospital, to being described as a keystone node in a coevolving integrated network of healthcare, so too must our general orienting models evolve from simple (baking a cake gone bad), to complicated (input–throughput–output machines), to complex (ecosystem) modeling. It is not that the earlier models are wrong, but they can be said to be partial given our current understanding of ED overcrowding and the science of complexity.
Asplin’s model is still essential, but its power may be multiplied when it is interpreted through a complex adaptive systems lens.

Cheng et al. have demonstrated that the nonlinear properties of queues (when irregular demand exceeds a fixed or falling capacity) have an overwhelming impact on blocking patient flow into ED stretchers (we are all unhappy in the same way). They have also reminded us that a rigorous improvement-science approach can identify other factors and possible solutions (we can all be happy in different ways). As a keystone connector in the evolving healthcare ecosystem, emergency medicine can catalyze top-down approaches to solving the real (and wicked) problem of “overcrowding” by influencing the boundary conditions through system design and integrative thinking, advocating for incentives and sanctions, and holding other decision-makers accountable for efficiency and capacity in their interdependent queues. We can also drive system change through bottom-up positive deviance, process innovations, and rapid-cycle, iteratively driven, continuous quality improvements. But for the patients we serve, expecting relatively small efficiency gains to fix ED access block, while ignoring the tipping point dynamics of inadequate capacity, will feel like the difference between war and peace.

Keywords: ED access block, queuing theory, wicked problems, complex adaptive systems

Competing interests: None to declare.

REFERENCES


