Emergency Department Crowding:
Vicious Cycles in the ED

June 2011

J. Bradley Morrison, PhD
MIT Sloan School of Management
MIT Engineering Systems Division
Brandeis International Business School
morrison@mit.edu
(781) 736-2246

Robert L. Wears, MD, MS
University of Florida
Imperial College London
École des Mines de Paris

We are grateful for support provided by NIH Grant 1 R21 HL098875, the National Heart Lung & Blood Institute, and the Office of Behavioral and Social Science Research
Emergency Department Crowding: Vicious Cycles in the ED

Morrison and Wears

Abstract

Over the past several decades, demands on the United States emergency and trauma care system have grown dramatically, but the capacity of the system has not kept pace. The result is a widespread phenomenon of crowded emergency rooms, especially in urban hospitals, which has become a major barrier to receiving timely care and has been implicated in adverse medical outcomes. This paper develops a stylized system dynamics model to examine the dynamics of patient flow in emergency departments. Simulation results show that increased ED resilience can come from relaxing bed constraints or from more human capability to cope with increasing workloads. The vulnerability of this system is rooted in the critical interaction between physical constraints imposed by the environment and the human capability of the staff to work at high performance levels under conditions of worsening workload pressure.
Introduction

ED / hospital crowding is an international problem, affecting hospitals throughout the English-speaking world. The problem first became apparent in US EDs in the 1980s, and was thought to be of crisis proportions by the end of that decade. The American College of Emergency Physicians issued a position statement (American College of Emergency Physicians 1990) and several policy recommendations (American College of Emergency Physicians 1990) at was then called “emergency department overcrowding” in 1990, but the problem only continued to grow (Derlet and Richards 2000; Goldberg 2000; Kellermann 2000; Zwemer 2000). Eleven years later, in 2001, the Society for Academic Emergency Medicine (SAEM) made crowding the theme of its yearly Consensus Conference; entitled The Unraveling Safety Net, the Conference resulted in the dedication of an entire issue of the Society’s journal, Academic Emergency Medicine, to a group of papers on the crowding problem (Adams and Biros 2001; Baer, Pasternack et al. 2001; Derlet, Richards et al. 2001; Gordon, Billings et al. 2001; Kelen, Scheulen et al. 2001; Reeder and Garrison 2001; Schneider, Zwemer et al. 2001; Schull, Szalai et al. 2001). Despite this attention, crowding has only gotten worse in the ensuing years (US General Accounting Office 2003; Kellermann 2006), culminating in a 2006 Institute of Medicine report that warned that the system was on the verge of total breakdown (Institute of Medicine 2006); despite this attention, and a plethora of interventions aimed at mitigating it, crowding seems to have been monotonically increasing over the past 25 years or so.

There have been multiple attempts to develop a workable definition of crowding (Hwang and Concato 2004). A recent systematic review of the crowding literature (Hoot and Aronsky 2008) concluded that the American College of Emergency Physician’s consensus definition seemed to encompass most of the important and relevant aspects of the problem: “Crowding occurs when the identified need for emergency services exceeds available resources for patient care in the
ED Crowding: Vicious Cycles

emergency department, hospital, or both.” (American College of Emergency Physicians 2006) This definition highlights crowding as an imbalance between supply and demand, and, as modified by Pines to include an impact on the quality of care (Pines 2007), has been widely accepted among researchers. Asplin et al (Asplin, Magid et al. 2003) advanced the understanding of the crowding problem by developing a conceptual model that provided a practical and now widely accepted framework for research, policy and management addressing crowding. The model (see Figure 1) partitions the problem space into 3 interacting components: input, throughput, and output, and has become generally accepted in healthcare in discussions of the crowding issue. Input factors reflect the sources and aspects of patient inflow; throughput factors reflect bottlenecks and delays within the ED; and output factors reflect bottlenecks in other parts of the healthcare system that might affect the ED.

Crowding has multiple, complex, interacting causes, and many ‘obvious’ causes have been discredited (Derlet and Richards 2000). Roughly 1/3 of the papers Hoot and Aronsky (Hoot and Aronsky 2008) included in their systematic review concerned research into the causes of crowding. These works tend to naturally fall into two separate areas, one concerned with general, long term trends and conditions, and the other with more specific, often local, triggering factors.

The long term trends are summarized by growing demand and falling supply. From 1995 to 2005, annual ED visits increased by 20% (from 96 to 115 million) and per capita ED visits by 7% (from 37 to 40 visits per 100) (Nawar, Niska et al. 2007). During the same period, the number of EDs decreased by 381, the number hospitals decreased by 535, and the number of hospital beds by 134,000 (Nawar, Niska et al. 2007; Health Forum 2008). In this view, crowding (and its consequences) is the inexorable result of long-term secular trends.

While not denying the influence of these general causal factors, work on specific factors has addressed issues such as ED use for non-urgent problems, by the uninsured, or by frequent users; and issues related to internal ED operating efficiency.
ED Crowding: Vicious Cycles

Work on crowding was initially held back by a number of assumptions, or “folk models” about its causes that ultimately proved to be false, or at least misleading (Newton, Keirns et al. 2008). For example, it has been widely thought that ED crowding is due to increased numbers of patients with relatively trivial, non-emergent problems, to increasing numbers of uninsured patients, or to “frequent flyers” – repeat visits by a small number of patients (Washington, Stevens et al. 2002). None of these hypotheses have been substantiated, and there is countervailing evidence for each (Sprivulis, Grainger et al. 2005). For example, Schull et al (Schull, Kiss et al. 2007) studied 110 EDs and 4.1 million patient visits in Ontario, and found that low-complexity patients contributed only trivially to length of stay and physician treatment times (32 and 13 seconds per patient, respectively). The same group also showed that ambulance diversion was not associated with either low complexity patients or with throughput factors, but was associated with output factors (Schull, Lazier et al. 2003). The results were similar across moderate and high volume EDs, and were robust to variations in the definition of low complexity. These results suggest that attempts to divert low-complexity patients to alternative sources of care are unlikely to substantially improve ED flow or to alleviate ED crowding. While this study does not dismiss the concern about nonurgent ED use as a policy issue – patients should not be forced into using the ED because they have no alternative – it does show that diverting low urgency patients away from the ED will not have a significant impact on crowding.
Figure 1. The input-throughput-output conceptual model of crowding (Asplin, Magid et al. 2003).

The problem was first framed as “ED crowding”, and initial work considered the ED in isolation – input and output factors were considered uncontrollable or at least outside the scope of ED managers who were dealing with the problem; in addition, 20 years ago, many ED inefficiencies did exist. However, as these inefficiencies were gradually wrung out of ED systems of care, the potential for alleviating crowding by addressing throughput issues has diminished. The weight of recent research has led to the conclusion that “… ED crowding is a local manifestation of a systemic disease” (Hoot and Aronsky 2008), and that effective solutions will have to set a scope that includes both input and output factors (Litvak, Long et al. 2001; Forster, Stiell et al. 2003; Richardson 2003). For example, systematic hospital restructuring has been shown to lead to subsequent crowding (Schull, Szalai et al. 2001). In another study, Rathlev et al (Rathlev, Chessare et al. 2007) retrospectively analyzed 93,000 visits at a single academic ED to describe the association of various input, throughput, and output factors on ED
length of stay. The only factors that were associated with increased length of stay were output factors: hospital occupancy, number of ED admissions to the hospital, and number of elective surgical admissions. The organizations that have had the greatest success in managing crowding have been those that recognized the hospital-wide nature of the patient flow problem and designed initiatives to address ED output at the organizational level (Cardin, Afilalo et al. 2003; Asplin and Magid 2007).

ED / hospital crowding leads to poorer outcomes in a variety of important conditions and patient groups, in brief, it hurts patients and degrades the quality of care (Bagust, Place et al. 1999; Richardson 2006; Sprivilis, Da Silva et al. 2006; Weissman, Rothschild et al. 2007). Crowding has been associated with delays in treatment (JCAHO 2002), increases in inpatient length of stay (Richardson 2002), particularly in the elderly (Liew and Kennedy 2003) and with increased mortality in hospitalized patients (Richardson 2006; Sprivilis, Da Silva et al. 2006). One of the earliest symptoms of crowding was the problem of ambulance diversion (Goldberg 2000; Eckstein, Isaacs et al. 2005; Burt and McCaig 2006; Sprivilis, Da Silva et al. 2006). Crowding has been associated with lower quality care for chest pain patients (Diercks, Roe et al. 2007), and delays in ED care (Schull, Morrison et al. 2003) and in delivery of definitive care such as fibrinolysis or catheterization in acute myocardial infarction (Schull, Vermeulen et al. 2004), and in worsened cardiac outcomes (Pines and Hollander 2007). It is associated with delays in antibiotic administration in serious infections (Fee, Weber et al. 2007; Gray and Baraff 2007; Pines, Localio et al. 2007) and deficient pain management (Hwang, Richardson et al. 2006) in the ED. In hospital care, crowding is associated with increases in adverse events (Cameron 2006), and in premature discharges from inpatient care (Baer, Pasternack et al. 2001; Jack, Chetty et al. 2009). Virtually every group of patients have been affected, but vulnerable populations, such as children (Committee on Pediatric Emergency Medicine 2004; Lorch, Millman et al. 2008) or the elderly are particularly susceptible (Hwang, Richardson et al. 2006).

ED – hospital crowding has shown “policy resistance” and has resisted efforts to alleviate or mitigate it. One of the striking observations about the ED-hospital crowding problem is its
ED Crowding: Vicious Cycles

persistence despite general agreement that it hurts both patients and health care organizations (Bagust, Place et al. 1999; Bayley, Schwartz et al. 2005; Falvo, Grove et al. 2007; Falvo, Grove et al. 2007). Multiple authors have raised the question of why it persists and in fact has worsened, in the face of multi-faceted attempts to control it (Kellermann 2000; Agrawal 2007; Kelen and Scheulen 2007; Moskop, Sklar et al. 2008; Moskop, Sklar et al. 2008; Viccellio 2008). This seems to be a classic case of “policy resistance”, arising, as Sterman (Sterman 2000) has suggested, from an incomplete understanding of the problem; essentially, researchers have been “looking in the wrong place” for insights into the crowding problem (Lane, Monefeldt et al. 2000).

Crowding exhibits many of the characteristics that are best addressed in a system dynamics approach. It shows non-linear dynamics analogous to phase shifts in physics (Hollnagel and Sundström 2006; Wears and Perry 2006; Woods, Wreathall et al. 2006), punctuated equilibria in biology (Gould 1989), or domain shifts in ecology (Holling 1973; Lesne 2008). Hwang and Lichtenthal’s characterization of slowly developing organizational crises seems apt here (Hwang and Lichtenthal 2000). In this paradigm, a slow change in a critical variable, which may be well known and easily identified, leads to a relatively sudden and discontinuous change in the behavior of the system when a threshold value is crossed; this is often accompanied by hysteresis – although a small increment in the critical variable may have led to a large change in the system, a subsequent small decrement will not restore the system to its previous state (Anderies, Walker et al. 2006; Walker and Salt 2006).

In addition, crowding shows delayed feedback loops (Hollander and Pines 2007) and complex interactivity. “Access block” – the inability to move admitted patients out of the ED because no inpatient beds are available – is associated with increased length of stay in hospitalized patients, which of course makes crowding and access block worse (Richardson 2002; Forster, Stiell et al. 2003; Liew and Kennedy 2003). Attempts to alleviate crowding often place pressure on physicians to discharge patients from the hospital sooner, but premature discharges lead to an increase in return visits to the ED by patients who are more complex, tend to stay longer, and are more often re-admitted (Baer, Pasternack et al. 2001; Jack, Chetty et al. 2009).
ED Crowding: Vicious Cycles

Many of the proposed interventions for crowding offer temporary respite but are either unsustainable or in the long run counterproductive. Where inpatient capacity is truly inadequate, increasing the supply of inpatient beds is of course indicated, but as a general solution is clearly unsustainable. Improving ED throughput by increasing departmental efficiency has been a central focus of effort, but recent studies of crowding have shown that both input and throughput factors are not associated with crowding, whereas output factors were (Rathlev, Chessare et al. 2007). Essentially, it seems that throughput factors have been optimized already, because the ED managers have been closest to the problem for many years and these factors are within their span of control; thus there is little further to be gained by incremental increases in ED efficiency (Karpiel 2004; King, Shaw et al. 2004; Patel, Derlet et al. 2006; Shah, Fairbanks et al. 2006; Worster, Fernandes et al. 2006). Other popular solutions, such as moving “boarded” patients from ED hallways to hallways on inpatient wards (Viccellio 2001), simply shift the location of the problem without addressing it in a fundamental way. Similarly, ambulance diversion has been shown to shift crowding from one hospital to another, and sometime to trigger a series of ‘tit-for-tat’ diversions that simply further increase congestion in the system (Asamoah, Weiss et al. 2008).

A final, minimal approach to the problem has been to manage it by fiat. The Joint Commission has declared ED – hospital crowding unacceptable, and that organizational leadership should “… develop and implement plans to identify and mitigate … overcrowding” (Joint Commission on Accreditation of Healthcare Organizations 2003) without notable effect. In the UK, crowding became a cause célèbre and led to a “4 hour mandate” – an NHS regulation that patients in the ED must be either admitted, transferred or discharged within 4 hours of the time they first signed in to the department (Department of Health 2000), enforced by financial sanctions on the organization for breaches. An analysis of the effect of this mandate shows a shifting of the problem – a sharp peak in hospital admissions and ED discharges just at 4 hours (Locker and Mason 2005). One of the effects of the 4 hour mandate in UK hospitals has been that the majority of these “admissions” are to a unit which is another part of the ED in all but name,
ED Crowding: Vicious Cycles

satisfying the technical requirements of the rule but having less effect on the problem (Weber, Mason et al. 2011).

Because of its dynamic complexity, delayed feedback loops, and social-behavioral components, the problem seems ideally suited to a system dynamics approach (Homer and Hirsch 2006), but it has been infrequently used. A Pubmed search for the terms ‘system dynamics’ and ‘emergency’ in any text field yielded only 6 citations, but only 2 of these were directly relevant. (By comparison, a search for ‘pancreatitis’ yields almost 45,000 citations). One of these studies was narrowly focused on laboratory response time and its effect on ambulance diversion (a proxy for crowding) (Storrow, Zhou et al. 2008); it showed a strong association between laboratory turnaround time and several measures of ED efficiency. The other (Lattimer, Brailsford et al. 2004) examined ED use at a regional rather than an organizational level, and predicted that ED volumes would increase, leading to increases in hospital occupancy and eventually “bottlenecks” – _i.e._ crowding – in the region. One additional paper not listed in Pubmed focused primarily on the tradeoff between beds for emergency admissions and those for elective surgery admissions, but not on the origins and persistence of crowding itself (Lane, Monefeldt et al. 2000).

Several other approaches have been explored, including discrete event simulation (Bagust, Place et al. 1999; Hoot, LeBlanc et al. 2008), queuing theory (Litvak, Long et al. 2001; Litvak, Buerhaus et al. 2005), and other engineering methods (Levin, Han et al. 2007; Levin, Dittus et al. 2008). While these approaches have provided useful insights, they have not addressed the central issue of whether the structure of the system itself produces the phenomenon of crowding.

Therefore, the broad, overall objective of this paper is to use system dynamics modeling (Sterman 2000) to study the problem of emergency department (ED) and hospital crowding in order to inform departmental, organizational, regional, and societal policies and interventions.
aimed at alleviating it. For example, a system dynamics understanding of crowding would be useful in the following ways:

- Developing early warning capabilities of a potential overcrowding crisis
- Identifying leverage points for managing dynamic and unexpected changes in patient demand or organizational capacity to respond
- Identifying potentially dysfunctional interventions to be avoided, *i.e.*, that might provide short term relief but ultimately make the overall problem worse.

The model development and analysis that follow are motivated by ethnographic observation of the day-to-day operating practices in the emergency department, including a level 1 trauma center, of a large, inner-city teaching hospital and by one author's first-hand experience as an emergency physician. The paper draws on data sources (not presented here) comprising observations, interviews, archival data, and the literatures in medicine, health care, the management sciences and organizational theory to inform the development of a system dynamics model and analysis that explores the phenomenon of emergency room crowding, with a particular focus on how the people and systems on the front lines adapt and adjust to cope with the challenges of excess demand.

**Model Development**

The input-throughput-output framework shown in Figure 1 is the starting point for our model development (Asplin, Magid et al. 2003). We begin by carefully distinguishing the stocks and the flows. Stocks are accumulations, such as the accumulation of patients in the ED. Flows cause increases or decreases in stocks. The framework depicts three sources of inputs that generate demand for ED care, which is the inflow to the stock of patients in the ED. The figure also shows two paths by which patients exit the ED, which are outflows from the stock of patients in the ED. Thus, "patient disposition" and "leave without treatment complete" are two outflows from the stock. The outflow labeled patient disposition comprises three possibilities - admit, transfer, or discharge to the ambulatory
ED Crowding: Vicious Cycles

care system. Finally, the figure also shows that patients returning from the ambulatory care system constitute another inflow to the stock of patients in the ED.

Figure 2 uses the traditional icons for system dynamics models to depict the stock and flow structure of this system. Stocks are represented by rectangles. Flows are represented by the pipe and valve icons. Each stock and flow is labeled with a variable name.

![Figure 2](image)

**Figure 2.** The stock and flow structure of the input-throughput-output framework. Stocks are depicted by rectangles. Flows are depicted by the pipes and valves. Clouds represents sources and sinks that are considered outside the model boundary.

The aim of the remainder of this paper is to develop and analyze a conceptual model of patient flows that allows us to examine some, but perhaps not all, meaningful aspects of the dynamics of ED crowding. The modeling process is iterative, and the choice of what to include in a model is based on the purpose of the model (Randers 1980; Homer 1996). Our purpose here is to begin to understand how patient management practices in the ED interact with elements of the broader health care system within which the ED functions, so we have chosen to include one aspect of patient management - decision making for patient disposition - and one aspect of the hospital system - admission to the wards.
We present the model here in stages, beginning with a model that focuses on the physical movement of patients, expanding on the structure shown in Figure 2. We turn our attention first to the admission process. When an ED physician (or physician team) decides that the proper disposition for a patient is to be admitted to the hospital wards, the decision triggers a complex process that usually leads to the physical transfer of the patient from the ED to the hospital ward. The ED issues a request for a consultation from a relevant specialist or general practitioner with admitting privileges. If the consulting physician concurs with the ED physician’s recommendation to admit the patient, the consulting physician writes admitting orders, initiating a request for assigning a bed to this patient. Once the patient has a bed assigned, the transport personnel in the hospital may physically move the patient to the hospital ward. The structure shown in Figure 3 adds the stocks and flows describing these key steps. The large rectangle around the three stocks of patients Awaiting Consults, Awaiting Assigns, and Awaiting Transport signals that these patients are typically still physically located in the ED. (For the purpose of this early conceptual model of the dynamics of ED patient flow, we will ignore the outflows for LWOBS and Transfers shown in Figure 2.)

**Figure 3.** The stock and flow structure with detail on hospital admissions.

The rates of patient flows will depend on various factors, including factors based on waiting times and processing times, available resources, and other capacity constraints. The available time for consulting physician specialists is an example of a capacity constraint that can affect the rate of Consults. The time required for a consulting physician to become free and to travel to the ED to see a patient contributes to waiting time. The time for communicating with the ED physician and evaluating the patient constitute processing time. Similarly, there are various
ED Crowding: Vicious Cycles

activities and delays associated with Assigns and Transports. We model these flows of (Consults, Assigns, Transports) by assuming an average elapsed time that comprises the waiting and processing times and further assume that this average time is constant. We also explicitly model how constrained availability of beds when hospital occupancy is high affects the rate of flow of Assigns. Because there is a fixed number of beds in the hospital, when the hospital approaches full occupancy, it becomes increasingly difficult to assign a bed to a patient. The rate of inflow to the stock of Admits with Beds must slow down, and indeed if the hospital is completely full must equal zero. The model captures this critical feedback process explicitly, as shown in Figure 4. The rate of Assigns is the lesser of the Desired Rate of Assigns and the Feasible Rate of Assigns. The Desired Rate of Assigns is a constant fraction per unit time of the stock of patients Awaiting Assign, representing the demand for beds from patients ready to be assigned. The Feasible Rate of Assigns represents the supply of beds that can be assigned to these patients. Beds may be available because there are empty beds (i.e., occupancy is less than 100%) and because patients get discharged, freeing their beds for reassignment. Thus, the Feasible Assignment Rate is the sum of the rate of assigning previously empty beds such that occupancy increases and the rate at which beds become available from Hospital Discharges. In most real hospitals, patients from the ED are only one source of demand for hospital beds. Others include surgical admissions and medical admissions directly from other specialties. The model here does not include other sources of demand, the bed capacity to serve them, or the decision making processes for assigning beds to these competing sources of demand. Instead, we interpret the fixed quantity of beds in the model as representing the beds allocated to patients from the ED.

Figure 4. The feedback structure of the constraint imposed by hospital bed availability.
ED Crowding: Vicious Cycles

In Figure 4, the lines with arrows are causal links. A causal link from one variable to another variable (which can be a flow) means that a change in the first variable causes a change in the second variable. For example, an increase in the rate of Hospital Discharges causes an increase in the Feasible Assign Rate. Conversely, an increase in the number of Admits with Beds causes a decrease in the Feasible Assign Rate, because the number of empty beds is lower. Together with the stocks and flows, the causal links form feedback loops. For example, imagine the stock of Admits with Beds increases (due to an inflow of Transfers). The increase in Admits with Beds causes a decrease in the Feasible Assign Rate. As this rate falls low enough, it causes the rate of Assigns to decrease. As the inflow of Assigns drops below the outflow of Transfers, the stock of Awaiting Transfers decreases, which reduces the rate of Transfers, slowing or stopping the increase in stock of Patients with Beds. The feedback process works to offset, or balance, the original change (the increase in Patients with Beds), so we designate this a balancing loop. Two such loops are labeled in Figure 4 as B1 and B2. Balancing loops bring stability to systems, often by limiting growth or moving the system towards some implied target. In this case, the loops act as controls on the inflow of patients to the wards given the physical reality that a bed must be available in order to assign a bed.

To use this model to investigate the dynamics of patient flow, we specify equations for each variable shown in the diagram. Appendix 1 presents the full equation listing. The equations translate the causal logic shown in the diagram into algebraic representations. Parameter values are required for constants such as average time delays (e.g., Avg LOS) and number of Beds. For our conceptual analysis here, we use parameter values suggested by practicing emergency physicians. Arrivals to the ED tend to be lowest in the early morning hours, rise to a peak in the late afternoon (around 4:00 or 5:00 pm) and then taper off throughout the night. The simulations in this paper all begin with an arrival flow that mimics this diurnal cycle as shown in Figure 5 generated by an average arrival rate adjusted by a diurnal multiplier. Discharges from the hospital are also subject to some of the same diurnal factors, so we adjust the endogenously generated rate of discharges by the same diurnal multiplier. We set the initial
conditions for all stocks to the long-term steady state values for midnight (because time 0 is midnight of the first the day) so the model begins near a steady-state. Figure 5 also shows the ED Census generated from simulating the model under the baseline conditions.

Figure 5. Left panel: Pattern of patient arrivals used as model inputs for the baseline and test scenarios. Right panel: Simulation results showing the total ED census in the baseline scenario.

To conduct simulation experiments with the model, we begin with the system in dynamic equilibrium as described and then introduce a change. For clarity of exposition, all of the simulations in this paper begin with the same initial conditions and then introduce at time=39 hours a one-time temporary increase in Arrivals that lasts for 10 hours after which Arrivals return to the original, baseline rate. The Arrivals graph in Figure 5 shows this surge of arrivals for one value (n = 6) of the temporary increase. The results of our first experiments, from introducing an increase of 5 patients per hour and 6 patients per hour, are shown in Figure 6. The first panel shows the Actual Wait Time for patients from the time an ED physician initiates the request for consult to the time the patient is transferred to a bed on the wards. The second panel shows the total ED Census, which is the sum of the stocks of Patients in ED plus those in Awaiting Consults, Awaiting Assign, and Awaiting Transfer. The results show the basic "physics" of the patients flows. At time 39, the increase in arrivals causes the ED census to begin to grow. Once the ED has stabilized and processed these patients, some are discharged and others are processed for admission. As the requests for admission begin to increase, the hospital beds become full. The Feasible Assign Rate drops well below the Admission Decisions and the stocks of patients Awaiting Assign and Awaiting Transfers grow. Arrivals slow
somewhat because of the diurnal pattern, bringing some relief in the congestion, but soon arrivals begin to grow again, causing the ED Census to grow as well. There are many patients still physically located in the ED, despite the fact that the ED physician and consulting specialist have already concurred to admit the patient and admitting orders have been written. Consequently, the Actual Wait Times grow. It takes quite some time for the effects of the surge in arrivals to dissipate, but they eventually do so, and over time the ED Census and Actual Wait Time returns to the original conditions. Recovery is slow, but the system has the resilience to eventually recover from the shock of additional arrivals. Figure 6 shows the results of another similar test when the magnitude of the temporary increase is 6 patients/hour. The results are qualitatively the same. These two simulations mimic the case of "access block" that has been described by other authors (Richardson 2002; Forster, Stiell et al. 2003; Liew and Kennedy 2003).

Figure 6. Response to a step increase in patient arrivals from time 10 to time 20 for a step height of 4 patients/hour and a step height of 5 patients per hour.

Expanding the Model

The model in the previous section includes one important aspect of the physical constraints imposed on the ED by the fixed bed capacity of the hospital system within which it operates. In this section, we extend the model to encompass some behavioral effects of ED crowding. We include additional feedback loops in the extended model and then use it to conduct further
ED Crowding: Vicious Cycles

simulation analysis for the purpose of deepening our understanding of the dynamics of ED crowding.

The previous simulations show that constraints on bed availability cause patients to wait in the ED for extended periods. Under such conditions, the increased number of boarders in the ED results in a greater workload for the ED staff. We extend the model here to consider possible effects of the increased workload on patient management practices in the ED. There are many such possible effects, but here we explicitly represent just one. We consider the effects of workload pressure on the decision making associated with patient dispositions. Specifically, we assume that when workload gets significantly higher than the normal workload, some fraction of disposition decisions are different. Greater workload leads to a higher frequency of admissions decisions for patients that would not have been admitted under less stressful conditions - what we will call Admissions Due to Bias. These might occur because of mistakes made due to workload pressure, but they might also occur as cautious physicians facing demanding workload become more likely to lean towards choosing to admit a patient for whom the disposition decision is a rather close call - the Admission Bias increases. Greater workload can also lead to a higher frequency of discharge decisions for patients that would otherwise have been admitted - what we will call Discharges Due to Bias. To model the flows of patients with these dispositions due to bias, we adjust the stock and flow structure as shown in Figure 7. The stock of patients in the ED is now comprises a stock of Patients in ED Destined for Admission and a stock of Patients in ED Destined for Discharge. The physicians do not know a priori in which stock the patients belong, but for modeling purposes we track them separately. The figure also shows a stock of Potential Revisits that is increased by the flow of Discharges Due to Bias and decreased by the Revisit rate, as patients return to the ED through the flow of Pre-Admit Arrivals.
Figure 7. A model of patient flow in the ED showing constraints on bed availability and the effects of workload pressure on patient dispositions.

An important consequence of Admissions Due to Bias is that they increase the flow of patients generating demand for the admissions process of consult, assign, and transfer. When the bed constraints are binding, Admissions Due to Bias will cause an increase in the number of patients in the ED - and these patients still generate workload demands on ED personnel because the patients are still physically in the ED. The workload demand from a patient for whom the admission decision has already been made (i.e, a patient in the stock of Awaiting Consult, Assign, or Transfer) is considerably less than that from a patient who is still under active evaluation. Nevertheless, the former group of patients still draw on the ED resources. As shown in Figure 7, an increase in these stocks constitutes an increase in the ED Census, generating an increase in ED workload, which in turn cause the Admission Bias to climb, resulting in more Admissions Due to Bias and further increases in the stocks that form the ED Census. The feedback loop, labeled “R3,” is a reinforcing feedback loop, because it acts to reinforce the direction of a change. Reinforcing loops move systems away from stability and are often implicated in dysfunctional dynamics.
To conduct our next simulation experiments, we need to specify the relationship between increased workload and the frequency of Admissions Due to Bias and Discharges Due to Bias. The effect of workload on disposition bias is modeled as an upward sloping nonlinear function of the actual workload compared to a threshold below which the bias is unaffected. For parsimony, we use the same effect functions for both admission and discharge biases (although the model allows us to parameterize these functions separately). Figure 8 shows how the Admission Bias depends on the variable Relative Workload, which is the current ED Workload compared to a threshold based on a multiple Normal Workload. Normal Workload is set to the peak workload experienced in the baseline scenario. The multiple of the Normal Workload is set to 1.05 in the following simulations. The tolerance of 5% additional workload above normal peaks before there is any effect on performance is a type of human capability that endows the ED with resilience to withstand a threat of increased demand. The Discharge Bias is model in exactly the same manner.

**Figure 8.** Admission Bias as a function of Relative Workload. Relative Workload is defined as the ratio of current ED Workload to the product of (1+Error Threshold) and the Normal Workload, which is defined as the peak workload in the baseline cycles. In the current model, the function for the Discharge Bias is identical.
ED Crowding: Vicious Cycles

The blue line Figure 9 shows the response to a temporary increase in patient arrivals of five patients/hour. The upper left and upper right panels show the Actual Waiting Time and ED Census, as in the previous simulations. The lower left panel shows the Admission Bias, and the lower right panel shows the stock of Potential Revisits (which arise due to the Discharge Bias). The response of this stylized ED department, with physicians of finite capacity, appears from the salient metrics to be quite similar to the response shown in Figure 6 when there are no biases. The increase in arrivals soon leads to constrained bed availability, blocking access and causing ED Census to grow. Wait times grow as well. The system remains crowded for an extended period, taking six or seven daily cycles to fully recover as before, to the normal peak and trough census values. However, there are some weak signals that the system has been stressed if we examine the less salient Admission Bias and Potential Revisits. During the periods of peak census, workload is higher than the threshold for tolerating excess workload, so there are some Admissions Due to Bias and Discharges Due to Bias, as seen in the graphs of Admission Bias and Potential Revisits. Nevertheless, the system recovers, despite the challenge in the form of a burst of additional arrivals.

Next, we consider the response to a slightly larger surge in arrivals. The red line in Figure 9 shows the response to an increase of six patients/per hour. Although the most immediate response appears similar to that for the smaller surge, the ultimate behavior is quite different. The hospital fills quickly as before blocking access and causing a backup of patients boarding in the ED. As before, the additional workload demand from the growing ED Census leads to an increase in the disposition biases. But now, system performance deteriorates rapidly and continues to worsen even after the surge in arrivals is over. Although the Admission Bias begins to fall immediately once the surge in arrivals has subsided, the consequence of the Admissions Due to Bias during the period of peak excess demand remain in the system - literally as boarders in the ED - keeping workloads high. As the workload is still high enough to engender some Admissions Due to Bias among the ongoing arrivals of patients, there is continued inflow of patients in the Awaiting stocks greater than the feasible outflow to the hospital wards. The system here has crossed a critical threshold, or tipping point, and we see that ED Census and
wait times continue to grow. Growing census leads to more biased disposition decisions, which in turn increases the census, and the system behavior is swept into instability by this vicious cycle. The system is not able to recover from a shock of this magnitude, a shock which is only slightly larger than the shock shown in the blue line. In a real world system, at some point additional feedbacks would surely intervene, but this simulation highlights the potential vulnerability of the system. For a sufficiently large surge in arrivals, the system crosses a tipping point beyond which the reinforcing loop R3 in Figure 7 has come to dominate the system, and the system is permanently overwhelmed.

**Figure 9.** Simulations with the model shown in Figure 7. Response to a step increase in patient arrivals from time 39 to time 49 for a step height of 5 patients/hour and a step height of 6 patients per hour.
ED Crowding: Vicious Cycles

To examine more closely the consequences of the biases in disposition decisions, we conduct the two simulations shown in Figure 10. The blue line shows the results when the only biases is the Discharge Bias decisions; that is, there are no Admissions Due to Bias. The red line shows the results when the only bias is the Admission Bias. The Admissions Due to Bias result in more patients in the ED, thus setting in motion the reinforcing loop R3. Discharges Due to Bias, in contrast, actually help the system by temporarily relieving some workload pressure. Although some fraction of these Discharge Due to Bias patients return to the ED, they leave during the period of extreme stress on the system. The model does not include adverse consequences on patient outcomes that no doubt arise from some Discharges Due to Bias, nor does it include an increase in the workload from a revisit patient that might be associated with the patient’s worsening condition.

![ED Census](image)

**Figure 10.** Simulations with the model shown in Figure 7. Response to a step increase in patient arrivals from time 39 to time 49 for a step height of 6 patients/hour with no Admissions Due to Bias (blue) and no Discharges Due to Bias (red).

Next we turn our attention to some simulation experiments that examine the sensitivity of the system to characteristics of the physical environment and the behavioral responses. What if the ED physicians are more tolerant of the excess workload? To answer this question, we vary the parameter that sets the threshold workload above which biases begin. In the previous simulations, this threshold was 5% above normal peak workload. We test a small change in this threshold by setting it to 8% and show the results in Figure 11. For comparison, the blue line
ED Crowding: Vicious Cycles

shows the same simulation as the red line in Figure 9 - a response to a surge in arrivals of 6 patients/hour. The green line shows the response when the threshold for workload tolerance is 8%. The system is now able to respond effectively to the challenge from the surge in arrivals. With fewer Admissions Due to Bias, the ED avoids crossing the tipping point and they are able to recover once the surge in arrivals is over. These results highlight an important feature of the dynamics of this system. Human capability, such as the tolerance of the ED staff to excesses in workload, is sometimes able to overcome significant challenges to the smooth performance of the system. More insidiously, precisely because the human capability is able to do so, the signal that performance is threatened is muddled.

![ED Census](image)

**Figure 11.** Simulations with the model shown in Figure 6. Response to a step increase in patient arrivals from time 39 to time 49 for a step height of 6 patients per hour. Blue: baseline. Green: Higher tolerances for workload stress. Red: Greater availability of beds.

Another possible improvement in this system is to free up more beds to be available for admissions from the ED. We conduct such a test in the model to see how the system behaves if there is easier access to beds by increasing the total number of beds. The grey line shows the system's response when the number of beds is two more than in the baseline scenario. The small increase in availability is enough to avoid the devastating overload, and the system is able to recover from the shock. This simulation demonstrates that, not surprisingly, changes in the physical environment (i.e., more beds) can make a system more resilient. The simulations in
Figure 11 highlight the important interaction between human capabilities and the physical environment. Both offer possible means for increasing resilience. A more physically robust workplace calls on less extreme human capability to achieve the requisite resilience to withstand a shock. Alternatively, a less robust physical setting requires more human capability to achieve the needed resilience.

The simulations in Figure 11 call attention to the interaction between patient management practices and the workplace setting in the hospital ED, highlighting that both dimensions have an important influence on patient flow dynamics and ED crowding. To further explore this critical interaction, we conduct a series of simulations in which we vary the size of the surge in arrivals (the input), the tolerance for excess workload (the human capability), and the number of beds (the physical setting). For several different combinations of bias threshold and bed availability, we conducted a number of simulations to determine the largest surge in the arrivals the system can withstand; that is, we identified the tipping points for each combination of parameters. The results are shown in Figure 12. Moving upward in this diagram represents increasing resilience - the ability to withstand and recover from a larger shock. For any given bias threshold (staying on any one line), greater availability of beds achieves greater resilience. Alternatively, for any given bed scenario (holding at one point on the horizontal access), increasing the bias threshold fosters greater resilience.
Figure 12. Results of experiments to identify the tipping points for various combinations of the Bias Threshold and Additional Beds. Results plot the influx (total number of additional patients over a 10 hour period) that pushes the system past the tipping point.

When hospital occupancy is high, the allocation of beds to ED patients is often difficult and occurs only after significant delays. We conducted a set of experiments to explore the effect of the timing of when an extra bed is made available. We use the same test scenario as before, a surge with an additional 6 patients per hour for 10 hours. Figure 13 shows the simulation results when no additional beds are allocated (blue line), which is the same as the blue line in Figure 11. The red line in Figure 13 shows the results when one additional bed is made available for ED patients 4 hours after the surge begins (t = 43 hours), and the green line shows the results when the additional bed is made available 8 hours after the surge begins (t=47 hours). The difference is outcomes is striking. When the allocation occurs 8 hours into the surge, the system does not recover from the surge. ED census levels are not as high as in the no extra bed scenario, but the census continues to grow long after the surge is over. The system has crossed the tipping point, and the additional bed allocated 8 hours after the surge begins is not adequate to resolve the situation.
Figure 13. Results of experiments testing the effect of allocating extra beds. Response to a step increase in patient arrivals from time 39 to time 49 for a step height of 6 patients per hour. Blue: No extra beds. Green: One extra bed allocated 8 hours into surge. These two scenarios push the system past the tipping point. Red: One extra bed allocated 4 hours into surge. The system recovers.

Discussion

Hospital emergency departments are complex settings that bring together a mix of health care personnel in a dynamically changing environment with a changing mix of demands amidst significantly constrained resources, such as time and space. Most of the time, these emergency departments operate at remarkably excellent performance levels, even though most of the time it seems they are operating under extremely challenging conditions. This paper uses a system dynamics model to examine some aspects of patient flow dynamics in the ED. We show that beyond a certain point, the system loses its ability to recover from increases in demand in the form of excessive patient arrivals. The simulation results highlight that the vulnerability of this system is rooted in the critical interaction between physical constraints imposed by the environment (e.g., bed availability) and behavioral factors, such as the human capability of the ED staff to work at high performance levels under conditions of worsening workload pressure.

The simulation results mimic a quintessential feature of life in the ED. Staff in EDs face an increasingly challenging mismatch between demand for their services and their nominal capacity to provide such service. Yet, although there are some occasions of failure and some
signs of deteriorating performance, EDs across the country largely continue to avoid catastrophic collapse of their systems. Human capabilities (e.g., the physician’s ability to continue to make proper dispositions in the face of adversity) compensate almost continuously for physical constraints and uncertainty. The simulation results show that increased ED resilience can come from relaxing bed constraints or from more human capability to cope. Importantly, there is a trade off of these two dimensions of bed constraints and workload tolerance. Improvement on one dimension can compensate for shortcomings in the other. In EDs where bed availability is constrained, staff that can tolerate extreme workload pressure without succumbing to disposition bias can enable the system to operate acceptably in response to greater shocks. However, more easy access to beds would enable the system to achieve the same levels of performance without the need to rely on the individuals who are more tolerant to workload excesses.

The concern arises because human capabilities are not infinite. When they get overloaded, system performance deteriorates rapidly. When operating near the tipping point these capabilities are the "buffer of last resort" that gives the system its resilience to recover. The human capability (of the ED staff in this example) to tolerate the extra workload masks the degree to which the bed constraint is threatening system performance, or at least reducing resilience.
Bibliography and References Cited


ED Crowding: Vicious Cycles


ED Crowding: Vicious Cycles


ED Crowding: Vicious Cycles


ED Crowding: Vicious Cycles


ED Crowding: Vicious Cycles


ED Crowding: Vicious Cycles


