Emergency Department Overcrowding in Massachusetts: Making Room in Our Hospitals

A discussion moderated by Catherine M. Dunham, Ed.D.
Director, The Access Project
Schneider Institute for Health Policy
Heller School, Brandeis University

Thursday, June 7, 2001
8:30 to 9:00 - Registration & Breakfast
9:00 to 11:00 - Discussion
Swissotel
Lafayette Place
Boston
This Issue Brief was prepared by Michael McManus, M.D., Associate Director of the Multidisciplinary Intensive Care Unit at Children’s Hospital in Boston and Executive Director of Pediatric Services at South Shore Hospital. He is a candidate for a Master of Public Health at Harvard University.

Executive Summary

In the past year, stories of Emergency Department (ED) crowding and ambulance diversion have entered Massachusetts headlines with increasing frequency. Hospitals throughout the state now report record numbers of hours on “diversion status,” unable to accept new ambulance arrivals because they are overwhelmed. When emergency responsiveness is threatened, public confidence in the entire health care system is undermined and immediate solutions are sought. Yet those who have followed the Emergency Medical Services (EMS) systems for many years recall similar stories and concerns in the late eighties and early nineties. The question therefore arises: Is it different this time?

This Issue Brief will detail the nature of the problem, summarize the factors contributing to ambulance diversion in Massachusetts today, and conclude that yes, it is different this time, in the following ways:

Changing balances of supply and demand

During the nineties, penetration of managed care paralleled an unprecedented decrease in the demand for emergency services. Crowding problems of the early nineties came under control readily as the number of ED visits fell and local strategies for managing temporary peaks of demand were implemented. At the same time, decreasing demand for hospital services generally, combined with a changing financial environment, led to wave after wave of hospital closure. Today, there are roughly one quarter fewer hospitals and emergency departments left in Massachusetts to serve a slightly greater population.

For reasons that should be investigated, over the past two years the demand for emergency services has begun to increase once again. While the total number of visits still remains below that of the early nineties, these visits are now funneled into fewer hospitals. As a result, while hospital efficiency is higher than ever before, the entire system is now severely stressed during periods of peak demand.

The diversion problem is no longer a limited one

While earlier difficulties were confined to the largest urban hospitals during busy winter months, all regions of Massachusetts now report increasing numbers of requests for diversion status—even during summer months. It is now commonplace for several hospitals within the same region to request diversion status simultaneously, necessitating complex mechanisms for rotating ambulance flow when requests for diversion must be denied.

Diversion is now more of a hospital than an emergency department problem

Increasingly, calls for diversion status now arise because of gridlock when hospitals are full and EDs are occupied with patients awaiting admission. As a result, the frequency of ambulance diversion is better correlated with total hospital occupancy than with the number of ED visits. Hospital crowding is particularly serious in specialty units such as Intensive Care Units (ICUs) and Cardiac Care Units (CCUs) where new emergencies are likely to require placement.

Despite the frequent lack of beds, hospitals have continued to downsize internally, staffing fewer than their full complement of licensed beds in order to maintain high occupancy rates and financial viability. Now, as increased demand has become more reliable, they find it difficult to reopen closed beds due to statewide staffing shortages.

Under present operating conditions, the system’s true capacity has been over-estimated

Variability in the demand for medical services makes 100% utilization impossible. Nonetheless, in the present economic climate, hospitals have been forced to seek such efficiency and to compete for patients in pursuit of ever-higher occupancy and turnover rates. Elective admissions are both predictable and controllable; therefore, as crowding worsens they inevitably come to occupy space that was once open and available for emergencies. From a financial perspective, this is advantageous in that it minimizes empty bed time. Yet from a systems
perspective, it increases the likelihood of ED gridlock.

At the same time, recent statewide occupancy rates of 60-70% have suggested that the Massachusetts hospital system continues to operate well below its capacity. This view is almost certainly erroneous, however, because a) there are serious methodological errors in current occupancy measures and b) it fails to account for the impact of variability upon capacity. On the contrary, recent survey data suggest that Massachusetts hospitals frequently may operate at dangerously high capacity.

Manpower shortages limit the system responsiveness

Although Massachusetts has traditionally benefited from the strongest health care workforce in the nation, a booming economy has created many attractive alternatives. Over the past several years, the pool of qualified health care professionals has declined and the mean age of available nurses has increased significantly. As a result, due to staffing shortages many hospitals cannot expand or even operate their full complement of beds, regardless of reimbursement or demand.

The new challenge of matching capacity to need is a predictable product of past policies

While health care downsizing has been the necessary product of a market solution to rising costs, the lower limits of this process have never been specified. Much effort has been expended toward increasing health care efficiency, but the present ED diversion crisis demonstrates that the goals of the marketplace will not spontaneously align with the goals of public health. To move forward, then, it is important to clearly define our public health needs, to understand the capacity limits of our system, and to create a mechanism by which we can agree when additional resources are required. In the short term, resources deemed critical must be supported directly while long-term solutions are sought.

The final section of this Issue Brief outlines the initial steps toward these ends, including recommendations to:

1. Determine the true nature of changing demand for emergency services and encourage access to medically-suitable alternatives.
2. Develop and support operations management strategies for improving patient flow and relieving ED gridlock.
3. Devise an ongoing method for measuring, monitoring, and adjusting overall hospital capacity.
4. Address current health care workforce shortages.

Introduction

Crowding in hospital emergency departments (EDs) has been a nationwide problem for more than a decade. Undeniably, EDs have come under increasing stress throughout the developed world as standards of living rise and access to quality healthcare is increasingly considered a right of citizenship. At the same time, downward pressure on hospital costs and the perception of “overcapacity” in the healthcare system has encouraged the closure of many hospitals across America through a market-based process that has yet to complete. As access expands and hospitals close, a supply-demand imbalance seems inescapable.

One manifestation of this supply-demand imbalance has been overcrowding in many of the nation’s EDs. As crowding worsens and waiting times increase, it becomes ever more difficult for ED staff to evaluate each new patient in a timely fashion. Overwhelmed, they must close their doors and request that new ambulance arrivals be diverted to alternative (and often distant) care sites. As a result, cost of care rises while quality deteriorates. The system’s burden is shifted, then, to those least able to bear it—the most critically ill.

Emergency medical services (EMS) in Massachusetts are organized regionally with all hospitals falling into one of five EMS regions. When overwhelmed, individual EDs request diversion status from their regional communication centers. That is, they request that new ambulance arrivals be diverted from theirs to neighboring institutions. In general, diversion status is requested for one of three reasons: 1) the ED staff is occupied and unable to promptly care for new arrivals, 2) the ED is physically filled and has no available bedspaces, or 3) critical support facilities within the hospital (for example, intensive care unit beds) are unavailable. Historically, organized ambulance diversion was suggested as an appropriate and effective means for relieving pressure on temporarily overcrowded EDs.

The issue of ED crowding and ambulance diversion first gained national attention in the late 1980’s. At that time it was marked by numerous
articles in both the medical literature\textsuperscript{1,10-14} and the lay press.\textsuperscript{15-18} Throughout the country, local governments organized task forces, hospitals hired consultants and medical associations issued position papers to cope with the problem. Here in the Commonwealth, emergency department crowding was the subject of at least three separate studies commissioned by the Massachusetts Medical Society\textsuperscript{19}, the Massachusetts Hospital Association\textsuperscript{20} and more recently a survey by the Department of Public Health.\textsuperscript{21}

In 1992, the U.S. Senate Committee on Finance, concerned by reports of diminished access to care among the uninsured,\textsuperscript{22} commissioned a nationwide study of hospital emergency departments.\textsuperscript{23} Based upon a survey of 689 U.S. hospitals, the General Accounting Office (GAO) observed that visits to emergency rooms had indeed, increased significantly over the preceding decade but that utilization was uneven. At that time, the problem of treatment delays was most prevalent in large, urban hospitals and was exacerbated by inappropriate use of emergency services by those with non-urgent conditions (estimated at up to 43% of all visits). Overall, 89% of patients appeared to be receiving timely care and only 7% of patients with urgent conditions experienced significant delays. As a result, no national policy interventions were recommended and local, primarily management-based efforts dominated the past decade.\textsuperscript{24-29}

In Massachusetts, local management strategies coupled with decreasing demand for services seemed to address adequately the problems of emergency department crowding for the rest of the nineties. In the past two years, however, these issues have reentered the spotlight as all EMS regions now report skyrocketing rates of ambulance diversion. Local media again are filled with related stories and, among health care professionals, anecdotal reports of compromised care abound. Once again, a task force has been convened, this time by the Massachusetts Department of Public Health and others, to search for solutions. As in other states,\textsuperscript{30,31,32} experts warn that the public health “safety net,” or provider of last resort, of emergency care has begun to fray.

This Issue Brief will begin with a supply-side inventory of the present hospital system in Massachusetts—its capacity, vigor and direction. The changing demand for hospital and particularly emergency room services will then be described along with an appraisal of the present balance of supply and demand. The dimensions of the ambulance diversion problem will then be presented, along with a model for understanding the forces currently stressing the EMS system. Ambulance diversion will be considered as a symptom of the larger problems of systemic saturation and hospital overload. After detailed discussion of each stressor and our present knowledge gaps, potential policy options will be proposed.

Section 1: Hospital Supply and a Decade of Change in Massachusetts

The optimal size of a healthcare system for a given population is unknown and, as a subjective issue, is perhaps unknowable. Historically, determination of the number of hospital beds or emergency departments necessary to service a given population has been a descriptive rather than prescriptive process. In the United States, the 1950’s, ‘60’s and ‘70’s were spent expanding health care in an attempt to bring state-of-the-art medicine within the reach of all Americans. By 1980, however, rising costs had raised concerns of excess capacity and prompted attempts at regulatory control. Nonetheless, until managed care in California made 2.5 beds per 1,000 residents seem reasonable, conventional wisdom held that a ratio closer to 4 per 1,000 was most appropriate.

In 1982, there were approximately 4.5 hospital beds for every 1,000 Massachusetts residents, and an average of approximately 4.4 beds per thousand residents nationwide. By 1993, when the Government Accounting Office (GAO) issued its report and a new President was seeking to reform healthcare, these ratios had already declined to 3.43 and 3.28, respectively.\textsuperscript{33} In part, the historically generous supply of beds in Massachusetts may have reflected the national and international reputation of our hospitals, with the “export” of hospital care to out-of-state residents accounting for approximately 5% of all discharges.\textsuperscript{34}

By 1990, however, national healthcare costs had reached 13.6% of GDP and Massachusetts Medicaid expenditures were rising an unacceptable 20% per year. In response, hospital revenues were deregulated in the Commonwealth in favor of a market-based solution. Over the decade of the nineties, then, Massachusetts hospitals transitioned from full cost reimbursement, “reasonable and customary” physician fees, and medical training premiums to Diagnosis Related Groups (DRGs), competition, and managed care.
By 1999, the number of Massachusetts residents participating in managed care plans reached 3.27 million (up from 1.58 million in 1990), over half the state’s population. In response, both the number of hospitalizations and total hospital days in the Commonwealth have fallen precipitously (Figures 1 and 2). Because urban areas tend to operate with much lower hospital-to-resident ratios than rural areas, it is noteworthy that Massachusetts, a largely urban state, now operates at ratios approaching the nationwide average. Thus, although admission rates have fallen somewhat nationwide, the drop in Massachusetts has been more pronounced.

**Figure 1:** Hospital Admissions per 1,000 Residents

![Figure 1: Hospital Admissions per 1,000 Residents](image1)

*Source: AHA Hospital Statistics 1992-2001*

**Figure 2:** Hospital Inpatient Days per 1,000 Residents

![Figure 2: Hospital Inpatient Days per 1,000 Residents](image2)

*Source: AHA Hospital Statistics 1992-2001*
At the same time, hospital efficiency in Massachusetts remains very high by national standards. More admissions are turned over per bed (Figure 3), and the average length of a hospital stay has declined 20% over the decade.

Figure 3: Average Number of Admissions per Bed


Loss of Hospitals and Beds in the 1990’s

Competition carries with it the reality that there will be both winners and losers. Until recently, market forces have successfully curbed utilization, but the price has been a dramatic change in the Massachusetts health care infrastructure. Despite a slowly growing (and aging) population, over the last decade nearly one quarter of the hospitals in Massachusetts have closed or been consolidated into larger systems, (see http://www.state.ma.us/dhcfp/pages/pdf/hospmass.pdf), leaving the total number of operating beds now 28% below 1990 levels (and nearly 10% below the national average) (Figure 4). Today, Massachusetts ranks 34th among the 50 states in beds per 1,000 residents, despite continued patronage of its hospitals by out-of-state residents. Amidst continuing financial pressures, in FY99 nearly one half of all Massachusetts hospitals reported negative patient care margins. As the impact of the Balanced Budget Act of 1997 continues, further downsizing seems inevitable.

* As discussed in the text, reported numbers of available beds vary among sources. For consistency, data obtained from the American Hospital Association in its Hospital Statistics publication will be presented unless otherwise specified.
In the past ten years, the number of EDs available to Massachusetts residents has followed the decline of full service acute care hospitals—101 in 1990 to 77 in 2000 (a reduction of 24%). Perhaps a more revealing statistic is the number of hospital beds now standing behind the state’s emergency department volume. Beyond general efficiency, factors influencing the equilibrium ratio of beds to visits include geography, degree of urbanization, and local practice patterns. In 1993, there was, on average, one operating hospital bed in Massachusetts for every 135 ED visits. This was significantly higher than the national average of 107 visits per bed, yet below the densely-populated Northeast region generally. In the nineties, however, the number of ED visits per operating bed in Massachusetts has steadily risen by 30% to 176 with some individual hospitals experiencing ratios over 200 (Figure 5).
Notably, over the last two years the ratio of beds to visits has increased sharply. After a period of relative stability during the middle nineties, 1999 saw an abrupt rise from 156 to 176 visits per bed. On the supply side, this was associated with the net loss of approximately 400 beds due to closures of Symmes Hospital (in Arlington), Malden Hospital, and Boston Regional Medical Center (in Stoneham). On the demand side, since 1998 the number of ED visits has increased for reasons that are, as yet, unclear (see Section 2). Early survey data from 2000 suggest that this trend continues.

Yet consolidations and closures are not the only sources of diminishing hospital capacity. The Department of Health Care Finance and Policy lists 17,274 licensed beds in Massachusetts acute care hospitals but only 16,910 operating beds. Because this latter number represents a weighted average, the true number of operating beds at any particular time may be even lower. For example, only 15,515 staffed beds were reported to the American Hospital Association in 1999. This discrepancy suggests that many of the Commonwealth’s remaining hospitals have downsized internally—closing beds that they once operated, or failing to grow with demand into the beds for which they once planned. Barriers to full capacity utilization include financial constraints, variability, and staff shortages, each of which will be discussed below.
Section 2: Rising Demand for Emergency Department Care

Emergency department (ED) utilization in Massachusetts has traced an interesting pattern over the past decade (Figure 6). Despite a modest increase in the population (6.0 million in 1990 to 6.35 million in 2000\textsuperscript{39,40} annual ED visits declined significantly until 1998, when the number of visits stood 15% below levels just six years before. While a determination of the reasons for this requires further study, one correlate has been the penetration of managed care (Figure 6).

![Figure 6: Emergency Visits and HMO Penetration](image)

Sources: ED visits- AHA Hospital Statistics 1992-2001;\textsuperscript{33} HMO Penetration Rate- Massachusetts DHCFP\textsuperscript{37} and American Association of Retired Persons

In 1999, however, Massachusetts saw its first absolute increase in ED visit volume in over a decade. Preliminary survey data for FY2000 suggests a further rise of between 4 and 5% has also occurred.\textsuperscript{38} As depicted in Figure 7, this rise is more than what would be expected from the population increase recently documented in 2000 census figures.\textsuperscript{49} If sustained, this trend will soon return statewide emergency department volume to the 1990 levels where ED crowding first became a serious concern. This time, however, signs of system wide stress have surfaced earlier as more visits are being channeled into fewer remaining hospitals. It would seem that the utilization reductions brought by managed care were sufficient to minimize ED crowding and mask the steady loss of hospitals throughout the nineties. These benefits now have been exhausted, however, and more serious difficulties are appearing as demand again becomes difficult to control.

\textsuperscript{‡} According to the U.S. Census Bureau’s Census 2000 Redistricting Data\textsuperscript{40} the population in MA increased 333,672 (5.5%) from 1990-2000 (state rank: 13/50 in size, 41/50 in growth). Figure 8 population figures for 1992-99 are U.S. Census Bureau estimates.
The drivers behind this new increase in ED demand have yet to be clearly identified. Changes in ED volume typically arise from changing patterns of disease (prevalence, severity) or changing patterns of usage. As discussed below, there are reasons to suspect that both forces are operative in Massachusetts. Yet absent more detailed datasets, it is difficult to pinpoint the precise reason for this rise. While prudent layperson legislation has been cited as one driver, the rise in ED volume seems to antedate its implementation in Massachusetts. A second factor, the decline of managed care products for seniors, is also a potential contributor, but it is difficult to substantiate this hypothesis using presently available information. Because the rise is coincident with the introduction of several managed care consumer protection measures these may also have played a role. Finally, an increase in inappropriate ED use, perhaps due to declining access to community-based care, has been suggested.

For whatever reason, a relative statewide supply/demand imbalance has appeared. The average hospital in the Commonwealth now sees approximately 35,000 ED visits per year, well above both national and regional averages (Figure 8). More importantly, this statistic conceals significant local and regional imbalances, with 20 hospitals in Massachusetts handling more than one half of all of the state’s emergency visits in 1999.

In Massachusetts, Chapter 141 of the Acts of 2000 defines “Emergency Medical Condition” as follows: a medical condition, whether physical or mental, manifesting itself by symptoms of sufficient severity, including severe pain, that the absence of prompt medical attention could reasonably be expected by a prudent layperson (emphasis added) who possesses an average knowledge of health and medicine, to result in placing the health of a beneficiary or another person in serious jeopardy, serious impairment to body function, or serious dysfunction of any body organ or part, or, with respect to a pregnant woman, as further defined in section 1867(e)(1)(B) of the Social Security Act, 42 U.S.C. § 1395dd(e)(1)(B).
Section 3: Ambulance Diversion and Emergency Room Overload in Massachusetts

Directors of all five Massachusetts EMS regions now report that ambulance diversion is a significant public health problem. While previously confined to isolated circumstances during busy winter months, calls for diversion status are now common throughout the year. Preliminary information from the first three months of 2001 suggest that the problem has worsened considerably since 2000, which previously was the worst year in recent memory.

When an ambulance is diverted from the nearest point of care, a chain of impact follows. For the patient, care is postponed when he or she can least afford it; ambulance transports are inherently time-sensitive and any delay represents diminished quality of care. For the “second choice” hospital, an increased ambulance diversion load contributes to crowding with unfamiliar patients whose records are unavailable and who may require services beyond its usual capabilities. For the ambulance team, increased transport and turnaround times mean diminished responsiveness and availability. For the general public, overcrowded EDs and ambulance diversion mean loss of timely access to emergency care, possible admission to hospitals far from home and, usually, lost choice of treating physician.

EMS Region IV includes metropolitan Boston and seven of the twenty busiest emergency departments in the state. During 2000, overwhelmed Region IV hospitals required more than 5,600 hours of ambulance diversion. Through the entire year, the ten busiest departments in Region IV together accounted for nearly 4,800 diversion hours or 1.5 hours per day each. One center averaged nearly three hours per day on diversion for the entire year. As depicted in Figures 9 and 10, the frequency of diversion was seasonal and worsened throughout the year with nearly one quarter of the Region’s EDs on diversion each day in December 2000. Despite more restrictive policies, diversions in EMS Region IV during the month of March quadrupled from 2000 to 2001.
EMS Region I encompasses least densely-populated Western Massachusetts. Until 1999, ambulance diversion in Region I almost never occurred. In 2000, diversion was requested 40 times, averaging 4 hours per request. The frequency of diversion in this region has also increased steadily such that by February 2001, Region I hospitals already had requested diversion status 27 times.

For an ED, EMS diversion status represents both a worst case condition and a significant underestimate of the crowding problem. When an entire region is overloaded, best practice guidelines dictate that hospital emergency departments come off diversion and that new patients be received on a rotating basis. In addition, most regions now follow diversion protocols which limit both the total time permitted on diversion status and the total number of ED closures within the same area. As a result, an increasing number of diversion requests are being denied. This necessity now gives rise to an even more serious problem—ambulances stranded in ED bays, unable to transfer patients to an overwhelmed hospital staff.
Section 4: Who Goes on Diversion and Why?

During calendar 2000, one quarter of the EDs in Massachusetts requested 100 or more hours of ambulance diversion. In general, ambulance diversion is requested by a hospital for one of three reasons: (a) the ED waiting time is incompatible with the hospital’s EMTALA (Emergency Medical Treatment and Active Labor Act”) and quality of care obligations to evaluate each new patient in a timely fashion, (b) specific services within the hospital are saturated and unable to accommodate the specialized needs of new ambulance arrivals (i.e., no available intensive care unit or cardiac care unit beds) or (c) the hospital’s total bed capacity is filled and patients awaiting admission are now queued in the emergency room producing a hospital-wide gridlock. Of these, the last is now the most common and most significant, reflecting a dangerous situation that is not easily resolved. While problem (a) may be remedied within 1-2 hours as those waiting to be seen are eventually cared for and (b) applies to only a subset of patients, problem (c) may persist for many hours and tends to recur on a daily basis. Figure 12 illustrates the interplay of factors contributing to ED overload and ambulance diversion.

**42 USCS §1395dd. Examination and treatment for emergency medical conditions and women in labor. Also known as the federal “anti-dumping” law, EMTALA was enacted in 1986, and requires all hospital EDs receiving Medicare funds to provide screening and stabilization to all patients who come to the ED before transferring them to another facility.
Figure 11 Factors Contributing to ED Overload and Ambulance Diversion

- Increased Demand
  - successful market competition
  - elective/nonelective competition
  - medical/surgical competition

- Decreased Capacity
  - fewer hospitals
  - fewer ED's
  - fewer nurses

- AMBULANCE DIVERSION
  - unfamiliar patients
  - unknown problems
  - EMTALA mandates

- ED Overload
  - increasing process time
  - need for consulting specialists
  - complex work-ups

- Hospital Overload
  - emphasis on outpatient vs. inpatient evaluations
  - complex technology
  - CT Laboratory

- Discharge Barriers
  - late physician rounds
  - few SNF beds
  - home care issues

- Diminished Capacity
  - fewer hospitals
  - fewer staffed beds
  - ICU
  - CCU
  - floor
In 1993, the General Accounting Office noted in its report that ambulance diversion was primarily an urban problem, related more to hospital crowding than to ED saturation (Table 1 and Figure 13).\textsuperscript{23} Nationally, diversion requests due to “patient loads beyond the ED staff’s ability to treat new arrivals” were most common in small rural hospitals.\textsuperscript{23} Busy, urban EDs, in contrast, were usually capable of meeting high patient loads but tended to close when all available beds were occupied. Across Massachusetts, both types of overload are now being reported.

### Table 1  Characteristics of ED diversion in 1990*

<table>
<thead>
<tr>
<th></th>
<th>Nationwide</th>
<th>Rural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requested ambulance diversion</td>
<td>39</td>
<td>14</td>
<td>61</td>
</tr>
<tr>
<td>Requested diversion 25-100 times</td>
<td>16</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Requested diversion more than 100 times</td>
<td>11</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Diversion status lasted more than 8 hours</td>
<td>23</td>
<td>13</td>
<td>25</td>
</tr>
</tbody>
</table>

*Numbers are in percent and represent survey data from 689 hospitals nationwide.
Source: U.S. General Accounting Office\textsuperscript{23}

As suggested earlier, increasing ED volume may represent either increased \textit{appropriate} utilization or increased \textit{inappropriate} utilization. The latter occurs when increasing numbers of people turn to the ED for care of non-urgent conditions. In California, for example, that state’s medical association estimates that more than 80% Medi-Cal and uninsured patient visits (one third of total California visits) are for non-urgent conditions.\textsuperscript{30} Although such patients might be better served in a physician’s office, they often select the ED for convenience or because insufficient alternative care sites exist in their communities.\textsuperscript{42} Anecdotally, medical practices increasingly refer after hours or unscheduled patients to the local ED for care. Anxious patients may seek complex medical testing or specialty consultation in the ED rather than wait for an appointment with their regular physician. Primary care physicians may refer patients to the ED for testing or consultation beyond an office’s capability. Finally, for many uninsured patients, the emergency department is both their first entry point into the health care system and their health care source of last resort.

Yet in the ED, even non-urgent conditions must be approached as potential emergencies. Each patient must be evaluated as an unknown and the physician must rely on complex testing to definitively rule out serious illness. Where a primary care physician may employ “watchful waiting” and return “check-up” visits in the management of uncertainty, these options are not available to the emergency physician. As a result, care of non-urgent conditions in the ED is significantly more time-consuming, laborious, and resource-intensive than in a physician office. EMTALA mandates that for any patient who “comes to the emergency department… the hospital must provide for an appropriate medical screening examination \textit{within the capability of the hospital’s emergency department}.” Thus, a legal risk is run by any hospital attempting to divert patients from the ED entrance to a lower-acuity clinic. Further, a rule recently promulgated by the Health Care Financing Administration (HCFA) has clarified “comes to the emergency department” as meaning “is on hospital property.”\textsuperscript{43} It is not surprising, then, that even if triage to a non-urgent care site is medically appropriate, legal concerns have made full ED-style evaluation of all arrivals the standard practice.\textsuperscript{44} It is interesting to note that while EMTALA acknowledges “diversionary status,” ambulance operators who disregard the status and enter the campus nonetheless are considered “on hospital property.” Hospital-owned ambulances in the field are also considered “on property,” regardless of their home hospital’s occupancy status.\textsuperscript{44}
If non-urgent ED utilization is on the rise, this might be detectable in rates of hospital admission from the ED. That is, the proportion of visits resulting in admission to the hospital should decline if more of those visits are for non-urgent conditions. This, indeed, may be the case as admissions via the ED statewide have increased only about one third as fast as the total number of ED visits (Table 2). As depicted in Figure 12, the number of hospital admissions arising from the ED varies from region to region yet, overall, 13.7% of ED visits resulted in admission during 1997, 13.7% in 1998 and 13.3% in 1999 when visits began to rise. Admission rates may also be affected by changes in medical technology, practice patterns, or even the availability of inpatient beds. Further, non-urgent utilization is often difficult to define since patients frequently arrive with concerning symptoms yet later “rule-out” for serious illness (for example, “chest pain” proving to be dyspepsia). Conversely, seemingly simple chief complaints may actually herald very serious illness (for example, “abdominal pain” proving to be appendicitis or pancreatic cancer). For these reasons, identification of “inappropriate” ED visits is difficult, and any attempted diversion of non-urgent visits risks diminishing access to care. Whatever the case, more detailed statewide emergency visit datasets will be necessary to evaluate meaningfully this component of ED overload.

| Table 2: Emergency Visits and Admissions with Emergency Charges, 1997-99 |
|--------------------------|-----------------|-----------------|-----------------|-----------------|
| FY97 2,481,814 | 20,891 | 339,852 | - |
| FY98 2,491,005 | 9,191 | 340,369 | 517 |
| FY99 2,565,777 | 74,772 | 342,936 | 2,567 |

Source: DHCFP hospital discharge data

The other side of the unnecessary visit coin, however, is the ED “observation” stay. These are patients with concerning complaints who await test results or must complete a period of close observation before they can safely be released. Before managed care, such patients routinely would have been admitted to the hospital. Today, however, the priority is to avoid admission whenever possible. The ED, then, must absorb the inappropriately long stays of many patients who have significant illness that warrants inpatient care.
**Figure 12**: Inpatient Discharges with ED Charges by Year and EMS Region

![Bar graph showing inpatient discharges with ED charges by year and EMS region.]  
- **Source**: Division of Health Care Finance and Policy 403 cost report and hospital discharge data

**Physician Specialists in the Emergency Department**

In their 1990 report to the Massachusetts Medical Society, Abt Associates identified incomplete specialty coverage as a factor contributing to ED overload. Hospitals rely heavily on specialists within their medical staffs to provide needed care to emergency room patients. When this coverage is unavailable, delays are common and transfer often necessary. Time “on call” is traditionally uncompensated and individual physicians apply to the patient’s insurer for payment on a fee-for-service basis.

For a variety of reasons, it has become increasingly difficult for physicians to receive compensation for specialty consultations rendered in the ED. In addition, as subsequent referrals are primarily dictated by practice plans and insurance arrangements, the ED is no longer a necessary or reliable vehicle for specialists to build and sustain their practices. Since heavy night work interferes with effective office practice the following day, some practitioners may elect to withdraw from a medical staff altogether rather than bear the burden of ED call. Others may simply be unwilling or unable to leave their practices to answer ED calls during regular office hours. While withdrawal from the medical staff is not possible for hospital-based specialties, willing office-based physicians (such as psychiatrists, neurologists and pediatricians) may be in short supply. Other specialties may simply be underrepresented in a specific locale leaving an insufficient number of practitioners available to share ED call. Among these, surgical subspecialties are most frequently cited (Table 3).
Table 3 Specialty Coverage in 89 Massachusetts Hospitals

<table>
<thead>
<tr>
<th>Specialty</th>
<th>% ED’s experiencing coverage difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurosurgery</td>
<td>26</td>
</tr>
<tr>
<td>Orthopedics</td>
<td>26</td>
</tr>
<tr>
<td>Plastic Surgery</td>
<td>26</td>
</tr>
<tr>
<td>Hand Surgery</td>
<td>18</td>
</tr>
<tr>
<td>General Surgery</td>
<td>11</td>
</tr>
<tr>
<td>Obstetrics/Gynecology</td>
<td>10</td>
</tr>
</tbody>
</table>

Source: Abt Associates survey, 1990

Whenever specialty consultation is required, length of stay in the ED increases and crowding is exacerbated. As general demands on an ED increase, so does its reliance upon specialty consultation. When consultations are permitted for convenience rather than urgency (for example, a plastic surgeon called for repair of a simple laceration), the volume of such calls increases. Over time, scarce specialists become even less available, time-consuming transfer arrangements must be made more often, and as patients are more regularly transferred to larger EDs, system-wide crowding increases.

ED Diversions and the “Full House”

In its 1993 report, the GAO observed that ED crowding was most frequently associated with delayed admissions and hospital occupancy rates exceeding 60%. Review of regional EMS data and discussion with emergency room directors also points to hospital crowding as the prime cause of ambulance diversion in Massachusetts today. In EMS logs, frequently cited reasons for diversion status requests are “no CCU beds,” “no ICU beds,” or “too many boarders” (patients remaining in the ED after admission due to lack of an available floor bed). This echoes both GAO and Abt Associates surveys of a decade ago where filled hospital beds prompted a large share of ambulance diversion requests (Figure 13).

Figure 13: Reasons Cited for Ambulance Diversion in the U.S.

Source: U.S. Government Accounting Office, 1993
It is perhaps not surprising that EDs requesting diversion are attached to the busiest, most successful hospitals. What is surprising, however, is that the number of ED diversion hours is better correlated with total hospital occupancy than with ED volume (Figure 14). Analysis of diversion logs together with FY99 cost report data demonstrates that the frequency of ambulance diversion in EMS Regions I, IV, and V is strongly correlated with total hospital occupancy (correlation coefficient = 0.515, two-tailed significance < 0.001) but less so to ED volume itself (correlation coefficient = 0.269, two-tailed significance = 0.08). Three variables, hospital size, occupancy and ED volume are also strongly correlated with one another. The overall picture is clear: Massachusetts prefers large, busy hospitals, these hospitals usually have busy EDs, and these EDs now frequently must divert patients when the house is full.

Figure 14  Ambulance Diversion and Total Occupancy in EMS Regions I, IV, V

As discussed above, since 1993 the average length of a hospital stay in Massachusetts has declined. It is notable here that no correlation is observable between a hospital’s overall average length of stay and either its time on diversion or its overall occupancy. This would seem to argue that crowding in specific institutions is not a function of beds being “blocked” by excessively long average hospital stays. Yet hidden within total occupancy statistics are unit-specific figures. For example, while a hospital’s total occupancy rate may lie in the 60% range, specialty units within it may be operating at or above full capacity. In such small, specialized units, one or two patients with protracted courses effectively decrease the operating capacity of the unit and interrupt flow considerably. In the case of intensive care units, this is particularly true. These units then become bottlenecks when high occupancy becomes the norm. In some regions of Massachusetts, particularly EMS Region IV, ICU utilization recently has increased (Figure 15).
In 1999 there were 1,139 ICU and 188 CCU beds in Massachusetts with average annual utilization rates of 60% and 71% respectively (DHCFP 403 cost report data). Closer inspection of the data, however, reveals that 22 intensive care units operated with average annual occupancy rates above 70% and eleven operated above 80%. Among CCUs, one third operated with average census rates above 80%. In some institutions, intensive care unit crowding rivals the ED and can be the primary cause of patient diversion.66

**Internal Downsizing**

High occupancy rates in Massachusetts hospitals reflect both system-wide and internal downsizing. Just as entire hospitals have closed under financial pressure, so too have individual hospitals downsized internally, closing beds or decreasing staff in low volume areas. By design, the lower occupancy rates of the ‘70’s and ‘80’s became unacceptable in the market-driven ‘90’s. Indeed, reduction of “excess capacity” was one explicit goal of deregulation. As a result, FY99 cost report data showed nearly 300 acute care beds removed from the system in this manner, 57 of which were ICU beds (Table 4).
As noted previously, these figures likely represent a significant underestimate of the unstaffed beds in Massachusetts. Yet even if this estimate is accurate, from an inpatient services perspective it constitutes the equivalent of an entire hospital removed from service. More importantly, it represents potential room for immediate re-expansion without new construction. In times of critical overload, a few extra beds, particularly critical care beds, could dramatically decrease the frequency of diversion requests.

Section 5: Utilization and Variability: What is “Capacity”?

If overcrowded EDs are, in large part, the result of overcrowded hospitals, why are overall hospital occupancy rates only 60%, and why must many hospitals continue to downsize internally? The answer lies in understanding the relationship between utilization, capacity, and variability.

Consider the theoretical example a small hospital consisting of only 20 beds. There are 20 x 365 = 7,300 bed-days available, by one measure of capacity. On a simple level, hospitals are reimbursed sometimes according to occupied bed days and, to that extent, this measure is a relevant one. Theoretically, if 1,460 patients are admitted in the course of a year, each requiring a 5-day length of stay, and each arriving precisely as an occupied bed opens, 100% of this kind of capacity is utilized. Increased throughput (as reflected in discharges per unit time, another gauge of capacity) can then be achieved by working to decrease the length of stay. Indeed, in response to changing reimbursement patterns, this has been the focus of most hospital management teams over the past decade and is reflected in an average length of stay now 20% below 1992 levels. 53

Unfortunately, the continuous, homogeneous patient flow model just described is only theoretically possible. In practice, while it may be impossible to achieve a standard length of stay for all patients, the second requirement of perfectly-timed arrivals is approximated by using waiting lines or “queues.” That is, patients requiring care may simply wait in line to be admitted as soon as a bed becomes available. In this manner, empty bed time is minimized and utilization maximized. During periods of high demand, the queue length increases; when demand is light, the queue length decreases. This simple concept manages much of the demand for goods and services throughout our society and, in health care, is very useful for non-urgent medical needs.

The Problem of Variability

Lengths of stays vary, patients with urgent conditions cannot be made to wait in line, and the number of usable beds in our model hospital may change unpredictably with staff availability. Variability in these parameters makes 100% utilization impossible. In hospital care generally and the emergency department in particular, it is impossible to predict with any certainty when the next patient will arrive. Even where seasonal changes in flow can be anticipated, exact forecasting is impossible. Amidst these uncertainties, utilization and occupancy rates are better analyzed in terms of probabilities.

To better understand this, consider the dilemma of the local car wash. Summertime may be busier than winter and weekends more than weekdays. Yet, a fixed number of service bays must be constructed and the owner must determine what number of sites will minimize waiting time and still stay busy enough to be worth the investment.
Because the number of service sites is fixed, but demand for service is variable, the solution is not obvious. Clearly, however, some sites will need to be empty sometimes, and at other times there will be waiting lines. To determine the best number of sites to build, we need a model that takes into account the fact that people requesting car washes arrive when they want to, and not according to the evenly-flowing schedule that an owner might prefer.

Queuing theory is a branch of mathematics that is widely used in engineering and industry for just these kinds of problems—understanding and modeling systems where a fixed capacity must be matched to variable demand. In the car wash, variables for our model to consider include the arrival rate of customers, the time it takes to run each wash (service time), and the number of bays available (servers). If we assume more-or-less random arrivals over the course of a year, using queuing formulae we can calculate the fraction of time that a service site probably will be empty or the length of the lines that probably will form on busy days.

Because arrivals for medical care are similarly random, many hospital systems can be modeled in much the same way. In this case, relevant parameters include the number of beds (servers), the length of stay (service time), and the arrival rate of new patients. If a simulation model of our twenty-bed hospital is constructed, the likelihood that its beds will be filled can be readily calculated using queuing theory. Using such a model, Table 5 describes the likelihood that a given number of beds will be occupied if our imaginary hospital experiences 1000 randomly-arriving admissions and a 5-day average length of stay.

<table>
<thead>
<tr>
<th>Total number of beds occupied</th>
<th>Fraction of time this occurs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>1%</td>
</tr>
<tr>
<td>7</td>
<td>2%</td>
</tr>
<tr>
<td>8</td>
<td>4%</td>
</tr>
<tr>
<td>9</td>
<td>5%</td>
</tr>
<tr>
<td>10</td>
<td>8%</td>
</tr>
<tr>
<td>11</td>
<td>9%</td>
</tr>
<tr>
<td>12</td>
<td>11%</td>
</tr>
<tr>
<td>13</td>
<td>11%</td>
</tr>
<tr>
<td>14</td>
<td>11%</td>
</tr>
<tr>
<td>15</td>
<td>10%</td>
</tr>
<tr>
<td>16</td>
<td>9%</td>
</tr>
<tr>
<td>17</td>
<td>7%</td>
</tr>
<tr>
<td>18</td>
<td>5%</td>
</tr>
<tr>
<td>19</td>
<td>4%</td>
</tr>
<tr>
<td>20</td>
<td>3%</td>
</tr>
</tbody>
</table>

Average Occupancy 67%
Recalling from our first example that 1,460 5-day admissions could be accommodated in our 20-bed hospital if patients could be made to wait in line, we now find that even caring for 1,000 such admissions is a problem if they cannot wait. With variable arrivals and no waiting lines, all twenty beds will be occupied only 3% of days, at least 6 beds will be occupied at all times, and most often there will be 10-15 beds occupied. Average occupancy will be around 67% and, despite what would seem to be sub-capacity overall utilization, entry to the hospital will still be blocked 2.6% of the time (approximately 10 days per year). As hospital managers, we are left the thorny problems of deciding for how many beds to hire staff and what to do on those 10 days per year.

At year’s end, it may be difficult to understand why a hospital with only 67% occupancy might divert ambulances 10 days a year, but variability dictates that this is necessarily so. It is important to recognize that this does not reflect hospital staff inefficiency, because both utilization and probability of rejection from the system are functions of the same three variables (arrival rate, length of stay and number of beds), when arrivals are random, higher occupancy rates are always accompanied by higher rejection rates. In our model hospital, decreasing the length of stay will indeed decrease rejection rates, but occupancy will fall as well. As the number of admissions increases and occupancy rises, so too will the frequency of rejection. Because calculated probabilities are extremely sensitive to the number of beds, the frequency of rejection will be extremely sensitive to staffing levels, individual bed closures or the presence of individual patients with very long stays.

In reality, because hospitals seek both short lengths of stay and high occupancy rates amidst highly variable demand, periods of complete saturation and patient rejection are inevitable. It must be recognized, then, that variability limits the maximum utilization of any system and that management of variability must be included in any discussion of “capacity.” In other words, “capacity” depends upon how much waiting is allowed and how much rejection is tolerable. Ambulance diversion represents rejection of those who cannot wait. To the extent that it is related to inpatient capacity, it can only be avoided by decreasing variability or by tolerating a certain number of empty beds.

In this light, the relationship between ambulance diversion and hospital occupancy rates is not surprising. In the United Kingdom, hospital occupancy rates approach 79% and mathematical modeling there suggests that levels of 85-90% will be regularly associated with bed crises.48 This has also been demonstrated in modeling Intensive Care Unit (ICU) services here in Boston.45 It becomes concerning, therefore, that in this past winter (2000-2001), which was marked by relatively mild influenza and respiratory syncytial virus prevalence49 (see Appendix A) hospital occupancies climbed far above 90% (Figure 16).

Measuring “Real” Capacity

Why does the crowding problem now seem particularly acute while annual occupancy rates in Massachusetts remain in the 60-70% range? At least three reasons appear likely. First, utilization rates as presently measured are nearly always based upon methodologies (such as midnight sampling) that significantly underestimate true daytime utilization.50-51 Here in Massachusetts, for example, a survey of all hospitals in the state during the first week of February 2001 disclosed that census as measured by total patients/staffed beds at midnight differs tremendously from census measured in the traditional fashion (registered patients/licensed bed at midnight). As depicted in Figure 17, hospitals in EMS Region IV reported an average daily census of more than 96% when measured at noon and only 77% when measured at midnight. While 96% would suggest a serious bed crisis, 77% might be considered under-utilization. The analogy has been suggested that measuring hospital census in the traditional manner is much like trying to measure daily traffic at Logan Airport by viewing a midnight snapshot of the runways.
The second reason that reported utilization rates may underestimate true rates is that annual averages fail to capture peaks in demand associated with seasonal variations in disease. Variability in the prevalence and nature of disease severely alter the short-term demand for hospital services. While it is widely recognized that many respiratory diseases, including influenza, are common in the winter and less so in the summer, the extent of this variability can be surprising. Figures 17 below and Appendix B illustrate the seasonal variations in illness and occupancy throughout the Commonwealth. With respiratory disease in particular, seasonal variations are extreme and predictably peak each winter. In hospitals maintaining 70% *average* annual occupancy, associated transient peaks in demand may easily push occupancy to critical levels. Smaller, specialized, critical care units are even more sensitive to such swings. Thus, a community may be adequately bedded in the summer and seriously under-bedded in the winter.

Finally, annual statewide averages fail to reflect uneven occupancy rates among hospitals that reflect demographic changes, patient preferences, specialty availability, insurance contracting and physician referral patterns. When comparing Massachusetts to more well-studied health care systems in Canada and the UK, it is important to distinguish between a system where regionalization of care is strict and one in which patients freely choose amongst competing hospitals. In a market-based system, successful hospitals run full, attract both elective and emergency patients, and are staffed closer to average demand than to peaks.
As more variability funnels into fewer hospitals, peaks of demand intensify. Exacerbating the problem further is the simple fact that patient load is both randomly and non-randomly variable. Examples of random or natural variability include the aforementioned sporadic appearance of diseases and accidents. Random variability can be modeled and managed using mathematical tools. Examples of non-random or artificial variability include man-made irregularities such as patient preferences, physician availability, and the weekday to weekend differences in elective caseloads. Non-random variability cannot be easily modeled, but it can be reduced by changes in practice patterns and by limiting the choice of where and when elective care is received.

At the end of the day, it is important for policymakers to recognize that regional or statewide average utilization data (as presently collected) fail to capture the peaks of occupancy that are most responsible for ambulance diversion. It is therefore difficult to understand and intelligently address capacity questions as they are currently framed. “Excess capacity” has until recently provided the buffer necessary for hospitals to accommodate peaks in demand. If excess capacity is now removed from the system and present capacity is increasingly insufficient to meet demand, more detailed data collection will be necessary to model patient flow and optimize the use of remaining resources.

**Competing Interests**

While only about 13% of ED visits result in hospital admission, on average about 53% of all hospital admissions in Massachusetts enter via the ED. As might be expected, revenues tend to be higher in hospitals where controllable non-urgent admissions exceed unpredictable emergency admissions (correlation coefficient = \(-0.528\), \(p < 0.001\), Figure 18). Thus, a competition of sorts now

---

### Figure 17 Statewide Average Hospital Occupancy Rates by Quarter and EMS Region (Inpatient Data)

<table>
<thead>
<tr>
<th>Quarter</th>
<th>West (1)</th>
<th>Central (2)</th>
<th>Northeast (3)</th>
<th>MetroBoston (4)</th>
<th>Southeast (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>96Q1</td>
<td>57.2%</td>
<td>66.5%</td>
<td>54.7%</td>
<td>59.8%</td>
<td>56.8%</td>
</tr>
<tr>
<td>96Q2</td>
<td>61.7%</td>
<td>65.8%</td>
<td>49.3%</td>
<td>65.0%</td>
<td>65.5%</td>
</tr>
<tr>
<td>96Q3</td>
<td>62.0%</td>
<td>60.1%</td>
<td>52.5%</td>
<td>62.4%</td>
<td>62.5%</td>
</tr>
<tr>
<td>96Q4</td>
<td>58.4%</td>
<td>57.7%</td>
<td>50.0%</td>
<td>61.2%</td>
<td>61.9%</td>
</tr>
<tr>
<td>97Q1</td>
<td>56.4%</td>
<td>61.0%</td>
<td>47.4%</td>
<td>60.2%</td>
<td>61.5%</td>
</tr>
<tr>
<td>97Q2</td>
<td>63.0%</td>
<td>61.2%</td>
<td>52.9%</td>
<td>64.5%</td>
<td>62.5%</td>
</tr>
<tr>
<td>97Q3</td>
<td>62.0%</td>
<td>60.1%</td>
<td>51.3%</td>
<td>64.4%</td>
<td>61.9%</td>
</tr>
<tr>
<td>97Q4</td>
<td>64.0%</td>
<td>57.7%</td>
<td>50.1%</td>
<td>64.8%</td>
<td>70.4%</td>
</tr>
<tr>
<td>98Q1</td>
<td>61.3%</td>
<td>61.2%</td>
<td>51.3%</td>
<td>64.8%</td>
<td></td>
</tr>
<tr>
<td>98Q2</td>
<td>67.3%</td>
<td>60.2%</td>
<td>51.3%</td>
<td>64.8%</td>
<td></td>
</tr>
<tr>
<td>99Q2</td>
<td>57.7%</td>
<td>64.8%</td>
<td>50.1%</td>
<td>64.8%</td>
<td></td>
</tr>
</tbody>
</table>

*Source: DHCFP 403 cost report and hospital discharge data*
exists between non-urgent and emergency admissions; non-urgent patients, though legitimately requiring care within a reasonable length of time, can nonetheless be safely placed in a queue and admitted sequentially as new beds open. Emergency patients, in contrast, are unpredictable, cannot be queued, and must be placed immediately into open, staffed beds. As discussed previously, in a competitive market, open staffed beds are a liability to be eliminated. Thus, over time, emergencies are gradually crowded out by electives that line up to compete for any open beds.

Figure 18: Admission and Fiscal Characteristics of 76 Massachusetts Hospitals in FY 99

Elective surgery presents a special case of queued admission, since a bed is committed the moment anesthesia is induced. Because surgical schedules are set in advance, hospital census status cannot be taken into account until the day of surgery. In busy hospitals, each morning’s elective surgeries are permitted to begin based upon the day’s anticipated discharges. Once started, a bed is committed even if these discharge estimates are incorrect. As volume increases and hospitals run at full capacity, elective surgical patients effectively co-opt beds from later-arriving emergency patients. While on the surface this may seem an efficient use of hospital facilities, it is counter-intuitively true that artificial variability in elective caseload is typically as great as the natural variability in emergency department volume. That is, even though daily elective admissions could theoretically be controlled to produce a more even patient flow, in most institutions the variability in elective caseload is equal to (and sometimes greater than) random variability. In some hospitals, elective medical admissions may predominate while in others elective surgery represents the majority of patient flow. In either case, as hospitals become extremely busy, even the variability inherent in elective admissions may unnecessarily add to intermittent crowding and overcapacity.

The primary insight to be gained from this line of reasoning is that expanding resources before attempting to control artificial variability can be counterproductive. If resources are expanded first, unnecessary variability is encouraged and problems ultimately resurface at different levels. After several expansion cycles, the result is a highly variable and oversized system. Conversely, once artificial variability has been removed, continued crowding represents an unassailable argument for allocation of additional resources. These principles are exploited in an innovative approach to the problem of ambulance diversion that involves removal of excessive

Source: DHCPF 403 cost report data
variability from elective surgery schedules. In this remedy, a significant portion of admissions are converted from a random, variable flow to a more continuous, homogeneous one. As peaks of admission variability are diminished, flow is improved, and rejections from the system are minimized. This strategy, though potentially powerful, has yet to be proven and would require both careful analysis and inter-departmental cooperation for its successful implementation.

Striking a Balance

In any discussion of capacity allocation and the ratio of elective to non-elective patient loads, it is important to recognize that current circumstances reflect the priorities we have set. As demonstrated above, variability inherent in the demand for emergency services necessitates a trade-off between the number of empty beds in a system and the number of rejections from that system. Historically, “excess capacity” in Massachusetts has served to accommodate emergencies. As this capacity has been removed, overcrowding and rejections (diversions) have necessarily appeared. While in the past hospitals have absorbed empty beds as a cost of doing business, financial priorities now require them to maximize occupancy and staff to average demand. Thus, if emergency readiness is required in a market-based system, we must decide who will pay for the empty beds.

At the end of the day, difficult tradeoffs between optimal “efficiency” and optimal “readiness” are necessary. While the problem of emergency room crowding is pressing, intentionally skewing the system in favor of emergencies would be ill-conceived and carries significant risk. For example, unless adequate overall capacity exists, any mandated “emergency” reserve of staffed beds will necessarily raise costs, decrease access, and increase waiting times for those with non-urgent conditions. More to the point, unless the empty beds are paid for, such mandates will ultimately prove counterproductive as more Massachusetts hospitals are forced into bankruptcy. The challenge, then, is to match capacity to need optimally and to assign the costs of emergency readiness fairly.

Section 6: Matching Capacity to Increasing Need

As our healthcare system shrinks, it is inevitable that more efficient management of the remaining resources will be necessary. But because the demands of the marketplace do not necessarily match the demands of public health and safety, there is reason for concern that market-based health care downsizing may exceed acceptable public health limits. Indeed, the current frequency of ambulance diversions has led many to argue (in Massachusetts and elsewhere) that we already have reached those limits. Whatever the case today, it is foreseeable that factors such as growth and aging of the population will produce future health care supply-demand imbalances if these competing interests cannot be reconciled.

Yet cogent arguments for re-expansion (or even preservation) of the health care system can be advanced only when there is consensus that available resources are being used wisely. To that end, it is important to identify and eliminate factors that may impair efficient utilization of the remaining health care infrastructure. At the same time, care must be taken to preserve existing elements that may be underutilized today but will become essential tomorrow.

Identifying the Bottlenecks: Diversion Itself

Perhaps the simplest initial approach to matching supply and demand involves identification of bottlenecks in the system that impair patient flow. With regard to ambulance diversions, a significant amount of energy has already been devoted toward improving the diversion process itself.

Historically, diversion practices across the state have varied among hospitals and, from region to region, even the meaning of the words “saturation,” “overload,” and “diversion” have differed significantly. Further, early warning of impending saturation has not been available to afford regional EMS teams the opportunity of preempting diversion calls by rerouting flow before overload occurs. However, several systems throughout the United States, including Massachusetts EMS Region II, have begun pilot programs whereby real-time patient load information is available to EMS teams, ED directors and hospital administrators. The common goal of these systems is to minimize ED duress through more efficient resource management.

In general, these systems begin with common diversion definitions and policies based upon a consensus amongst hospitals as to “best practice” guidelines. Here in Massachusetts, such guidelines...
have already been introduced by the Department of Public Health and the Massachusetts Hospital Association in conjunction with the Ambulance Diversion Task Force. When diversion is necessary at a participating ED, broadcast communications are issued to hospital and pre-hospital providers so that appropriate rerouting of patients may occur. As these systems are refined, they promise to permit real-time monitoring of hospital status across regions and across the state. At the same time, they will permit accurate data collection to support better understanding of the entire diversion phenomenon.

The primary challenge to all regional patient management programs is creation of a collaborative atmosphere among providers who are otherwise incented to compete in the marketplace. Further challenges arise from the fact that, as depicted in Figure 9 and discussed above, problems in the ED are more often the symptom of hospital overload than the fundamental cause. Thus, as tighter and tighter management is necessary, any efficient statewide system will ultimately require information regarding the ongoing status of all divisions within each hospital. This information must then be used to coordinate utilization amongst competing hospitals. Presently, such information is either unavailable or proprietary.

Identifying the Bottlenecks: Inside the Hospital and Beyond

When the problem of ambulance inflow is one of ED outflow, bottlenecks within the hospital become the focus of attention. These may be numerous and largely idiosyncratic to individual institutions. While the roles of ICU and specialized unit saturation have already been touched upon, further discussion of hospital management strategies is beyond the scope of discussions here. It should be emphasized, however, that innovative solutions to hospital gridlock abound, and that some of the best are generated by providers on the front lines whose intimate knowledge of their home system allows them to craft truly workable solutions. Many of these solutions may be portable to other institutions, while others may not. Lasting solutions to Massachusetts’ patient flow bottlenecks almost certainly will involve both.

Yet common to all hospitals are obstacles to the disposition of patients who are medically ready for discharge but continue to occupy beds. Discharge barriers range from delays in paperwork to the absence of skilled nursing facility beds. In some cases, issues as mundane as timing of a ride home may delay discharge for several hours. As discussed earlier, midnight census figures fail to reflect the daytime presence of these patients and the true level of bed utilization. Thus, complaints of crowding are unsupported by census figures, ED backup is inexplicable, and the policy response is uncertain.

Because swift bed turnover and shorter length of stay are now critical to hospital survival, the issue of discharge timing has been a focus of attention for some time. In this instance, market goals and public health goals are aligned. In the last few years, numerous strategies have been implemented to address these problems including early discharge planning, rescheduling of rounds, addition of hospitalist teams, and provision of transportation to those needing rides home. One of the benefits of the Massachusetts Ambulance Diversion Task Force has been the opportunity for health care providers to share their ideas and experiences around this issue.

It is noteworthy that an increasingly important discharge barrier remains outside a hospital’s control. When a patient is ready for rehabilitation or requires discharge to a skilled nursing facility, utilization in these facilities becomes important. In these institutions, because operating margins are thin, occupancy rates consistently run above 85-90%. To accomplish this, a queue of waiting patients must be maintained. Unfortunately, while in the queue these patients often occupy beds in acute care hospitals, contributing to gridlock and ambulance diversion. An interesting irony is the occasional congestion observed when ambulances are unavailable for transfers to skilled nursing facilities because they are delayed by diversion of their acute care runs. This form of gridlock should become more common as cutbacks in reimbursement for non-urgent ambulance transfers begin to discourage provision of these services. Clearly, any final solution to problems of patient flow will require an integrated approach with inclusion of chronic care sites.

Section 7: Initial Responses

As the problem of ED ambulance diversion has intensified over the past two years, some hospitals have responded by re-expanding their facilities. In part, this is predictable given the “lumpiness” of bed closure; in a region requiring 1000 beds, if one of three 400-bed hospitals were to close, the remaining two would need to add a total of 200 new beds. Additionally, innovative expansion of
EDs to include observation or extended monitoring units permits hospitals to add scarce monitored beds in a relatively inexpensive fashion. As a result, large EDs in Massachusetts have gotten larger, with the median size of the largest (top quartile) increasing 66% since 1994 (Appendix C).

In addition to ED expansion, surviving hospitals may attempt to respond to crowding by reversing the process of internal downsizing—reopening beds that had been previously closed for inadequate occupancy. As discussed above, using the difference between licensed and reported operating beds as a guide, theoretically as many as 290 beds could be added to the Massachusetts hospital infrastructure tomorrow. Yet both reopened beds and newly constructed ones now face the same obstacle—no one to staff them.

*Health Care Workforce Shortages in Massachusetts*

The present status of nursing and the health care workforce in Massachusetts has been the subject of an earlier Massachusetts Health Policy Forum. It is no coincidence that reports of a worsening nursing shortage now hit the Massachusetts headlines as frequently as reports of ambulance diversion. Despite efforts by hospitals to increase nursing staff over the past several years (achieving a 4.8 percent increase since 1996, according to Division of Health Care Finance and Policy cost reports), staffing vacancies remain. Many hospitals in the Northeast have now resorted to international recruiting reminiscent of the nursing shortages in the early 1980’s. Certainly, hospital beds are of little use if there is no one to attend them.

In many ways, the evolution of the present nursing shortage parallels that of ED crowding. As already discussed, managed care penetration, declining admissions, falling average daily censuses, and the need to staff below peak occupancy led to downsizing and a disappointing job market for nurses during the middle nineties. As a result, nursing school enrollment declined and the average age of nurses in the United States rose to 44.3 years. Today, as practicing nurses retire, hospitals are forced to compete for, and then fight to retain, a decreasing number of new graduates. Yet new graduates are unable to function in high-acuity, critical care units (such as EDs and ICUs) and require mentoring from more experienced nurses. Unfortunately, non-patient care activities (such as mentoring) are economically difficult to support.

Statistically, Massachusetts may seem to be better defended against a national nursing shortage than many other states. In 1996, the Commonwealth ranked 1st amongst the states in nursing employment per 1,000 residents and in 1998 it ranked 2nd in its overall supply of nurse practitioners. The mean hourly wage was 16% above the national average, and Massachusetts teaching hospitals continued to attract the best and the brightest health care professionals.

Yet other statistics raise concern. Only 60% of Massachusetts nurses are employed full time (state rank = 46/50) and the growth in nurses per capita has not kept pace with the rest of the nation (4% vs. 20% increase from 1988 to 1996). Two thirds of Massachusetts nurses are now over age 40, and only 53% of all nurses choose to work in hospitals. Many have chosen to leave active practice over the past two years, as low state unemployment rates and ample opportunities elsewhere in health care continue to support high turnover rates. Simply put, nursing careers in Massachusetts have become unattractive when compared to the available alternatives.

Massachusetts hospitals that have survived this far and now attempt to expand staff to meet increasing demand are left to do so in an environment where they can ill-afford to offer higher wages, better staffing ratios and career-developing non-clinical time. Indeed, even maintaining current staffing levels and avoiding further bed closure is difficult. If a critical-care-capable nurse takes 6-7 years to produce, this situation should continue for the foreseeable future. Meanwhile, Massachusetts is left to compete with other states for a shrinking pool of qualified professionals.

The dilemma of nursing is not unique, however. All health care workers, including physicians, have faced similar pressures over the past decade and these have taken their toll on morale and depth of the labor pool. As ED evaluation and care becomes increasingly complex, an army of personnel including radiology technicians, pharmacists, and respiratory therapists become critical to the process.

---

11 With 4,340 radiology technicians in 1998, Massachusetts ranked 12/50 (70.6/100,000 population)
12 In 1998, Massachusetts ranked 24th of 50 states in pharmacists per capita (67.2/100,000)
13 There were 1,590 respiratory therapists in Massachusetts in 1991 (25.9/100,000 population, state rank: 41/50)
These professionals also are now in short supply and their positions are often the rate-limiting steps in patient flow. As noted previously, while downsizing had been the goal for the nineties, its limit and extent were never specified. Clearly, it is now extending into areas that we cannot afford—the pool of talent upon which we have long relied.

**Conclusions**

Emergency departments are the only health care sites mandated by law to care for all arrivals. As such, they cannot control their inflow and are subject to wide variability in the demand for medical services. Most of these demands are reasonable, although some may be better served in lower acuity settings. For reasons that cannot be determined adequately from presently available data, in the last two years the demand for ED services in Massachusetts has begun to rise after more than eight years of decline. While this rise has undoubtedly put added stress upon the EMS system, ambulance diversion is no longer simply a response to inflow.

Today, the hospital infrastructure standing behind the Commonwealth’s EDs is considerably smaller than a decade ago. Surviving hospitals, evolutionarily selected out for efficiency, necessarily gravitate toward controllable (queue-able) elective patient flows. From a public health perspective, this natural consequence of market competition becomes important only when the limits of capacity are touched. At these points, when high utilization rates are the norm, variable emergency flow will always unsuccessfully compete with queued elective flow for any available beds.

Recently, ambulance diversion has entered the headlines because Massachusetts emergency departments now find themselves attached to hospitals that are often touching those limits of capacity. When hospitals fill, ED outflow is blocked and ambulances must be diverted. To many this comes as a shock because, for rather straightforward reasons, the true degree of hospital crowding has been underestimated by traditional measures of occupancy and capacity. As a result, ambulance diversion requests are appearing in parts of the state and at times of the year that are unprecedented.

The common theme of the nursing shortage, ED crowding and ambulance diversion problems is the same as that echoing throughout all aspects of health care—a free-market solution to rising health care expenses carries costs along with benefits. The good news is that the system is behaving as predicted; in response to financial pressures, capacity has been reduced by more than one quarter. To the extent that this “excess” capacity was a driver of increasing health care costs (a debate beyond the scope of this Issue Brief), much of this problem has been eliminated. The bad news is that two critical issues must now be faced: 1) the cost of this capacity cutback is a significant reduction in our hospital system’s reserve, flexibility and responsiveness; and 2) there is reason to believe that the downsizing process will continue, perhaps extending below acceptable public health limits.

With this in mind, it is not difficult to make an argument for allocation of additional resources. Indeed, many of the stresses now contributing to the diversion problem arose from financial pressure. However, while re-expansion of hospital services may be necessary to solve the ambulance diversion problem, this now faces serious financial and manpower constraints. Further, because crowding is a complex phenomenon manifesting itself differently in different hospitals, it is not obvious how, precisely, re-expansion should occur.

The ambulance diversion issue is a signal that Massachusetts now faces serious questions concerning the overall capacity of its hospital system. To answer these questions, we first must determine what our public health needs are and what we want our system of hospitals to do. Are emergency departments serving the role for which they were intended or are they bearing the burden of weaknesses elsewhere? Will simple infusion of capital solve, worsen, or fail to affect the problem? If more capacity is necessary, what kind is required and where specifically should new resources be directed? To be confident that resources are being used effectively, we must examine more thoroughly how emergency departments function within the health care system and identify any operational burdens on the EMS system. To the extent that ambulance diversion is an ED outflow problem, we must determine the best systemic remedies. At the same time, we need to understand the reasons behind recent changes in demand for emergency services and decide whether these changes represent an aberration or a threatening trend. Armed with knowledge of who seeks care in the emergency room and why, we must determine the appropriateness of such use and the feasibility of encouraging more suitable alternatives. Finally, any lasting solution will require us to develop mechanisms through which we can agree that public safety is well served by the size of our system and that it remains sufficiently dynamic to meet future public health challenges.
Policy Recommendations

1. **Determine the true nature of changing demand for emergency services and encourage access to medically-suitable alternatives.**
   a. Develop a statewide encounter-level database for all emergency departments in Massachusetts. Data managed through the Division of Health Care Finance and Policy should include:
      i. Patient demographics
      ii. Mode of arrival
      iii. Clinical and injury characteristics
      iv. Services and procedures
      v. Charges
      vi. Payer and provider data
   
      The State Legislature can establish a start-up grant program to cover the costs of data collection. These data, if collected retrospectively, would yield useful trend analysis of the contribution of these factors to emergency department overcrowding.
   b. Develop EMTALA-compliant hospital programs and policies geared toward diverting patients with non-urgent conditions to non-emergency care settings.
   c. Explore reimbursement strategies that encourage expansion of community primary care services, including after-hours clinics.

2. **Develop and support operational strategies for improving patient flow and relieving ED gridlock.**
   a. Patient flow bottlenecks and the impact of variability upon hospital overload should be investigated. To the extent that these contribute to hospital gridlock, modern operations management strategies could be employed to increase the effective capacity of the hospital system.
   b. The Department of Public Health and regional EMS providers should be funded to complete development of a real-time open-access monitoring system of ED saturation status to be used by Massachusetts EMS providers in routing emergency patients efficiently during periods of peak demand.
   c. State legislators can establish a matching grant program for supporting hospital programs aimed specifically at reducing ED diversion and hospital crowding. Preference should be given to community-based and regional cooperative efforts. First-year funding should be based upon merit, and funding thereafter based upon proven reductions in diversion rates.
   d. The extent to which supply/demand balances in skilled nursing facilities contribute to hospital gridlock should be examined. Mechanisms for facilitating transfers between acute care hospitals and skilled nursing facilities should be developed.

3. **Devise a system for measuring, monitoring, and adjusting overall hospital capacity.**
   a. To address concerns that the entire hospital system in Massachusetts now becomes overloaded during periods of peak demand, an ongoing method for measuring, monitoring, and adjusting hospital capacity should be developed, beginning with an inventory of available hospital resources.
b. A statewide capacity monitoring system is needed. Available hospital resources vary tremendously throughout the year and even within a single day. To match supply with demand adequately, frequently updated information regarding the following is required:

i. Occupied beds
ii. Staffed/closed beds
iii. Specialty unit availability
iv. Urgent vs. non-urgent admissions
v. Seasonal fluctuations in need for services

c. The capacity to take new emergencies depends upon the availability of staffed and waiting open beds. Mechanisms must be established to assign the costs of maintaining this readiness.

4. Identify and address current health care workforce shortages

a. The inability to admit patients from the ED to inpatient beds, specialty care units, or nursing homes in a timely fashion leads to both ED overload and ambulance diversion. An inventory and functional analysis of the Massachusetts health care workforce must be conducted to identify the geographic and specialty care areas of need.

b. State legislators should support state-funded efforts to encourage high school students to consider careers in nursing and other areas of health care, to complete their education, and to pursue employment in Massachusetts acute care hospitals.

c. Means should be established for supporting paid mentoring time for hospital clinical staff, especially nurses.

Acknowledgements

Many thanks to the following individuals for their kind assistance in preparation of this manuscript:

Mr. Brad Prenney, Mr. Paul Dreyer, and Ms. Nancy Ridley, Massachusetts Department of Public Health; Mr. Jerry O’Keefe, Ms. Kathy Fuda, Ms. Diane McKenzie, Ms. Raphaela Miller, and Ms. Amy Lischko, Massachusetts Department of Health Care Finance and Policy; Ms. Linda Moriarty and Mr. Stephen Cote, Massachusetts EMS Region I; Mr. Jonathan Epstein, EMS Region III; Mr. Stephen Nelson, EMS Region IV; Mr. Larry Heidenberg, Massachusetts EMS Region V; Ms. Leslie Kirle and Mr. David Smith, Massachusetts Hospital Association; Mr. David Hannan, Mr. Richard Aubut, Mr. Richard Colette, and Dr. Robert Fitzgerald, South Shore Hospital; Professors Alan Sager and Eugene Litvak, Boston University; Professors John McDonough and Stuart Altman, Brandeis University, Dr. Charlotte Yeh, National Heritage Insurance Company, Professor Nancy Turnbull, the Harvard School of Public Health, and Senator Richard T. Moore, Joint Committee on Health Care.
References


URL: http://www.bhpr.hrsa.gov/healthworkforce/profiles/default.htm


URL: http://www.state.ma.us/dhcfp/pages/pdf/trends.pdf


43. 42 CFR § 489.24 (added by 65 FR 18434, 18548, April 7, 2000).


URL: http://www.bmj.com/cgi/content/full/319/7203/155


50. Boaden R. Occupancy probably already higher than thought. BMJ 1999;16 (July)


52. Ambulance Diversion Survey. Dreyer MP, Massachusetts Department of Public Health; Data presented to the Massachusetts Ambulance Diversion Task Force and in Senate Hearings on Hospital Crowding. Boston, MA; March 2001.


URL: http://ihp.brandeis.edu/mhpf/GT-MHPF%20Workforce%20Issue%20Brief.pdf

58. MNA Agenda of Initiatives in Response to the Nurse Shortage: Massachusetts Nurses Association; 2000.
URL: http://massnurses.org/News/001001/staffrpt.htm
APPENDIX A

Respiratory Disease Patterns as Reported to the Center for Disease Control’s National Respiratory and Enteric Virus Surveillance System.49

Figure 1: Respiratory Syncytial Virus Trends. RSV is the most common cause of bronchiolitis and pneumonia among infants and young children. Note the relatively low prevalence of disease during the winter of 2000-2001.
Figure 2: Respiratory mortality trends in the United States. Respiratory disease is one of the primary contributors to the variability of demand for medical services (see also Appendix B).
Appendix C

Size of Massachusetts Emergency Departments (in sq. ft.)

<table>
<thead>
<tr>
<th></th>
<th>n = 85</th>
<th>n = 81</th>
<th>n = 76</th>
<th>n = 77</th>
<th>n = 70</th>
<th>n = 70</th>
<th>n = 67</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY94</td>
<td>7.163</td>
<td>8.177</td>
<td>7.441</td>
<td>8.482</td>
<td>8.548</td>
<td>8.548</td>
<td>8.814</td>
</tr>
<tr>
<td>FY96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FY97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FY98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FY99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F Y = 93, 94, 95, 96, 97, 98, 99