Variability in Surgical Caseload and Access to Intensive Care Services

Michael L. McManus, M.D., M.P.H.,* Michael C. Long, M.D.,† Abbot Cooper,‡ James Mandell, M.D.,§ Donald M. Berwick, M.D., M.P.P.,‖ Marcello Pagano, Ph.D.,# Eugene Litvak, Ph.D.**

Background: Variability in the demand for any service is a significant barrier to efficient distribution of limited resources. In health care, demand is often highly variable and access may be limited when peaks cannot be accommodated in a downsized care delivery system. Intensive care units may frequently present bottlenecks to patient flow, and saturation of these services limits a hospital’s responsiveness to new emergencies.

Methods: Over a 1-yr period, information was collected prospectively on all requests for admission to the intensive care unit of a large, urban children’s hospital. Data included the nature of each request, as well as each patient’s final disposition. The daily variability of requests was then analyzed and related to the unit’s ability to accommodate new admissions.

Results: Day-to-day demand for intensive care services was extremely variable. This variability was particularly high among patients undergoing scheduled surgical procedures, with variability of scheduled admissions exceeding that of emergencies. Peaks of demand were associated with diversion of patients both within the hospital (to off-service care sites) and to other institutions (ambulance diversions). Although emergency requests for admission outnumbered scheduled requests, diversion from the intensive care unit was better correlated with scheduled caseload ($r = 0.542, P < 0.001$) than with unscheduled volume ($r = 0.255, P < 0.001$). During the busiest periods, nearly 70% of all diversions were associated with variability in the scheduled caseload.

Conclusions: Variability in scheduled surgical caseload represents a potentially reducible source of stress on intensive care units in hospitals and throughout the healthcare delivery system generally. When uncontrolled, variability limits access to care and impairs overall responsiveness to emergencies.

AFTER more than a decade of downsizing, many hospitals throughout the country have begun to experience stress related to diminished capacity. This is particularly evident in emergency departments where overcrowding and ambulance diversion are now widely recognized as public health problems threatening national preparedness.1–3 Frequently, emergency department crowding reflects saturation of intensive care beds and other critical services within the hospital, as emergency patients requiring admission cannot be accommodated on filled inpatient units.4–9 At the same time, medical inpatients and postoperative patients compete for the same resources, resulting in operative delays, case cancellations, and prolonged stays in postanesthesia care units.10 In response, targeted reexpansion of hospital infrastructure could be considered, but workforce shortages and financial limitations often present formidable barriers. Alternative solutions to the crowding problem are therefore necessary.11

Variability in the demand for any service presents a significant challenge to the efficient distribution of limited resources. In health care, when hospital occupancy is high, peaks of demand necessarily produce crowding, staff overloads, and unmet patient needs. Therefore, amid diminished capacity, improving the healthcare system’s response to variability represents an opportunity for simultaneous gains in effective capacity, cost-efficiency, improved outcomes, and patient satisfaction. A precondition to such improvement, however, is a deeper understanding of the nature and sources of variation in demand. Without such understanding and appropriate management of variability, systems such as healthcare organizations become inefficient, overwhelmed, and frustrating for all.

Hospitals operating at high capacity provide useful models for studying the impact of variability on the quality of and access to medical care. Pediatric tertiary hospitals, which serve a “safety-net” function in the care of seriously ill children, are of particular interest because they often provide specialized services for which there are no suitable alternatives. Hospitals for children are typically smaller than most urban general hospitals and are therefore particularly vulnerable to the stresses accompanying wide swings in demand. These stresses may be indirectly manifested in care delays, procedure cancellations, nursing turnover, and patient dissatisfaction, but are unambiguously apparent when crowding within the hospital forces off-service “boarding” (patient placement into alternative care sites within the hospital) or complete refusal of new admissions. To explore the nature and impact of variability on this aspect of the quality of medical care, we applied variability methodology12 to investigate the relationships between fluctuations in demand or volume and access to critical care services in a large, urban children’s hospital.

* Department of Anesthesia and the Multidisciplinary Intensive Care Unit, Children’s Hospital Boston; Associate Professor, Harvard Medical School. † Senior Anesthetist, the Massachusetts General Hospital; Assistant Professor, Harvard Medical School. Adjunct Associate Professor of Management, Boston University. ‡ Senior Analyst, Program on Variability. § Professor of Health Care and Operations Management, Boston University School of Management. ¶ President, Children’s Hospital Boston; Professor, Harvard Medical School. CEO, Institute for Healthcare Improvement. # Professor of Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

Received from the Department of Anesthesia and the Multidisciplinary Intensive Care Unit, Children’s Hospital Boston, Boston, Massachusetts. Submitted for publication April 2, 2002. Accepted for publication December 3, 2002. Support was provided solely from institutional and/or departmental resources.

Address reprint requests to Dr. McManus: Children’s Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115. Address electronic mail to: michael.mcmanus@hms.harvard.edu. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.
Materials and Methods

All requests for admission to the 18-bed medical-surgical intensive care unit of a large, urban children’s hospital over a 1-yr period were analyzed. Sources of requests included the emergency department, operating rooms, general patient wards, and referring community hospitals (the latter by telephoned requests for transfer). All requests were categorized as scheduled or unscheduled. Medical admission requests included all patients requiring intensive care for acute medical illnesses and, by their nature, were considered to be “unscheduled.” Surgical requests included those patients requiring intensive care because of surgical illness or medical complications of surgical illness. A surgical admission was considered “unscheduled” when it arrived via the emergency department, through urgent in-hospital or interhospital transfer or following emergency surgery. All other surgical admissions were deemed “scheduled.” Staff intensivists made all classifications.

The disposition of each request was then classified according to one of two possibilities: admission to the medical-surgical intensive care unit, or refusal of admission. All refusals were categorized further by final disposition as those diverted to an alternative care site within the hospital (such as the postanesthesia care unit) or referred out to another institution because all staffed intensive care beds were occupied (fig. 1). Admission, diversion, and referral decisions were all made on a case-by-case basis by a staff intensivist charged with the responsibility of weighing bed availability and patient care needs. Patients initially diverted but later relocated to the primary unit after stays elsewhere were considered refusals. Patients referred to another institution for reasons other than bed availability (e.g., burn patients referred to a regional burn center) were excluded from analysis.

The number of requests for admission, their category, and their ultimate disposition were recorded daily. For comparison purposes, mean, range, and standard deviations in the daily number of requests were calculated for each category (scheduled and unscheduled). As a measure of the relationship between request flow and rejection from the system, associations between daily requests for admission and refusals were calculated as Pearson correlation coefficients. The effects of the total number of requests on these relationships were also examined by partial correlation. Two-tailed P values are presented throughout.

To determine the factors best explaining the number of patients diverted from the intensive care unit, a Poisson regression model was fit to the data (Stata statistical software, Stata Corporation, College Station, TX) and goodness-of-fit was verified by chi-square test. Explanatory variables included day of the week, the number of unscheduled admission requests, the number of schedulable admission requests, and lagged variables of 1 to 6 days as surrogates for lengths of stay. The relative importance of each variable is expressed as its correlation coefficient plus or minus its standard error. The chi-square test was used to test the equality of two effects.

To more intuitively describe the overall relationships, the variability of each type of request was also measured as the sum of its residuals. The sum of residuals was computed for each request type by adding the absolute value of the differences between each day’s actual number of requests and the daily average number of requests for that type (scheduled or unscheduled) over the sample period (sum of residuals = Σ |# each day’s requests − # average daily requests|). The practical significance of this simple quantity is that it provides an absolute measure of the number of times a supply/demand imbalance would occur in a system staffed to meet average demand. Ideally, resources would always be equal to demand. If inflow is variable but resources limitless, sufficient excess capacity could be maintained to meet transient demand peaks. When resources become limited, however, hospital units are commonly staffed for anticipated average demand. In this setting, imbalances occur whenever actual demand differs from this average. Higher demand, then, stresses the system by requiring staffing adjustments or overtime hours, whereas lower demand leaves empty beds and unoccupied health care professionals. The sum of residuals quantifies these disparities and represents an aggregate measure of the imbalances introduced into the system by variability.

Results

A total of 1,978 requests for admission to the intensive care unit were received during the 1-yr period. The classification and disposition of these requests are presented in figure 1. Forty-seven percent of all requests resulted from scheduled surgical procedures, and 53% resulted from (unscheduled) medical or surgical emergencies. Table 1 summarizes key descriptive statistics for daily admission requests and their dispositions. There were no significant interruptions to full unit operations.
during the period of study. Calculated average daily census over the entire period was 15.2, producing an overall average occupancy ~85%. The average length of stay for all patients was 3.8 days, with scheduled patients experiencing shorter average stays (2.0 days) than unscheduled patients (4.9 days). The variability of requests for admission in all groups was evidenced by their large ranges, standard deviations around their means, and sums of residuals. It was noteworthy that despite their smaller volume, scheduled requests contributed more to variability within the system than did unscheduled requests.

The most obvious manifestation of systemic overload in the intensive care unit is placement of a patient “off-service” or into another institution because all primary intensive care beds are filled. To explore the relationship between this condition and the variability of demand, daily refusals were compared to daily requests for admission both in the aggregate and by type of admission request. Although it is intuitively obvious that refusals are more likely as requests for admission increase, this basic relationship is complex and routinely modified by many factors, including staffing levels, occupancy, length of stay, and specific patient needs. Even with these potentially confounding factors, a very high correlation existed between scheduled surgical requests and refusals of admission ($r = 0.542, P < 0.001$). This relationship was preserved in partial correlations controlling for the number of unscheduled requests ($r = 0.579, P < 0.001$), indicating that the relationship between scheduled requests and refusals is not merely a by-product of a coupling relationship between scheduled and unscheduled cases. Further, no correlation was found between scheduled and unscheduled requests ($r = -0.066, P = 0.102$). Notably, unscheduled requests, though representing a greater fraction of the total (1,056 of 1,978 total requests), were less strongly correlated overall with refusals ($r = 0.255, P < 0.001$) than were scheduled requests.

In the regression model, elements identified as explanatory of the number of intensive care unit refusals included the day of the week (including a holiday effect), the number of admissions, and a lagged effect representing cumulative admissions over 1 to 6 days (length of stay effect). While controlling for day of the week and length of stay, the impact of scheduled admissions on rejection was significantly greater than that of unscheduled admissions ($P = 0.047$). Overall, scheduled admissions were 25% more associated with rejection than were unscheduled admissions, despite their smaller number ($r = 0.25 ± 0.02$ and $0.19 ± 0.02$ for scheduled and unscheduled admissions, respectively).

To further define the association between scheduled requests and refusals, running serial correlations spanning sequential 30-day periods were calculated. These measures disclosed numerous intervals of high demand and extremely coupled relationships wherein nearly 70% of intensive care unit refusals were associated with scheduled requests (maximum $r^2 = 0.69$, average $r^2 = 0.41$). In contrast, unscheduled admissions showed far less coupling (maximum $r^2 = 0.49$, average $r^2 = 0.13$). Taken together, the data demonstrated that bed availability was more strongly determined by variation in scheduled demand than by variation in requests for unscheduled admission. Figure 2 graphically illustrates this relationship during a representative period.

Finally, the overall contribution of day of the week to variability was explored by comparing weekday scheduled and non-scheduled request frequencies together with their residual sums. Table 2 reveals that although the average number of requests per day appeared relatively stable, variability, as evidenced by the sum of residuals, was high. Overall, supply/demand imbalances (residuals) arose equally from scheduled and unscheduled patient flows. Simply put, it was clear that scheduled arrivals were no more predictable than emergencies.

### Discussion

This is the first investigation into the direct impact of patient flow variability on access to medical care. The primary finding is that scheduled patient flow, although theoretically controllable is, counterintuitively, more variable than the random demand of emergencies. One result of this variability is a widely ranging demand for critical care services that, in units operating at high capacity, is frequently responsible for patients being placed off-service or denied access to the hospital altogether. The practical implication of this is that hospital capacity could be increased and systemic stresses reduced simply by smoothing scheduled patient flow. Although our data were obtained from a single institution, they are consistent both with observations we have made elsewhere and with analyses published in the operations management literature. Therefore, the principles illustrated here may be applicable to other busy hospitals.
The negative impact of hospital crowding on the quality of patient care is now being recognized,\textsuperscript{16,17} and there is growing consensus both in medicine and government that steps must be taken both to insure patient safety and restore hospital “surge capacity.” Increasingly, the daily practice of anesthesiology is complicated by crowding as it delays recovery room discharges,\textsuperscript{10} forces surgical case cancellations,\textsuperscript{15,14} and drives a growing number of postoperative patients into “off-service” care sites. In some hospitals, intensive care units have become bottlenecks where scheduled patient flows compete with emergencies to produce backups, crowding, and ambulance diversion. Therefore, under present resource constraints, it is crucial that new approaches be sought to optimally match the supply and demand for hospital care.

Variability in surgeons’ operative times, hospital lengths of stays, and physician practice patterns are watched closely by hospital managers, but variability in demand itself is largely overlooked. In many industries, however, efficient management of variable demand is a much-studied issue that is recognized as critical to corporate success. In health care, because demand fluctuations are largely perceived as random or seasonal, they

![Fig. 2. Daily admission requests and refusals over a representative 60-day period. Squares correspond to the number of scheduled surgical patients requiring admission to the intensive care unit; diamonds correspond to the total number of patients denied admission (managed off-service or transferred to other institutions). Note that most rejection peaks correspond to peaks in scheduled demand.](image)

Table 2. Characteristics of Admission Requests and Sources of Variability by Day of the Week

<table>
<thead>
<tr>
<th>Day</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average no. of requests</td>
<td>All</td>
<td>6.60</td>
<td>6.48</td>
<td>6.21</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td>Scheduled</td>
<td>3.62</td>
<td>3.44</td>
<td>3.46</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>Unscheduled</td>
<td>2.98</td>
<td>3.04</td>
<td>2.75</td>
<td>2.81</td>
</tr>
<tr>
<td>Average number of rejections</td>
<td>1.85</td>
<td>1.96</td>
<td>1.92</td>
<td>1.54</td>
<td>1.70</td>
</tr>
<tr>
<td>Sum of residuals</td>
<td>Scheduled admissions</td>
<td>79</td>
<td>67</td>
<td>73</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Unscheduled admissions</td>
<td>68</td>
<td>87</td>
<td>66</td>
<td>68</td>
</tr>
</tbody>
</table>
are usually considered uncontrollable. Yet variation in the demand for hospital services is actually of two types: “natural,” such as variability in type of disease, its severity, and the arrival pattern of patients; and “artificial,” that introduced by idiosyncrasies in the systems employed to deliver care. Natural variability is associated with the episodic healthcare needs of individuals, is primarily random, and is beyond the control of healthcare deliverers. Artificial variability, in contrast, is characteristically nonrandom, more unpredictable, and related to controllable factors in the design and management of healthcare systems. As examples, the random demand pattern for trauma services exhibits natural variability, whereas day-of-the-week variation in scheduled hospital admissions is artificial. Flows subject to natural variability can be modeled using stochastic tools, whereas those subject to artificial variability resist this and appear to behave erratically.

Scheduled surgery is representative of hospital demand sources that are routinely subject to artificial variability. Although the day-to-day scheduling of such admissions is controllable and could result in a smooth source of patient flow, this is seldom the case in clinical practice. Instead, artificial variability is more often uncontrolled and extremely irregular demand patterns can result. Unlike emergencies, scheduled admission patterns are the product of a complex combination of competing priorities. As such, they do not follow a Poisson distribution and, as demonstrated here, can sometimes exhibit larger ranges, standard deviations, and overall variability. This, in turn, produces more instances of supply/demand imbalance (here quantified as “residuals”) than a similar number of emergency admissions.

Whenever resources are limited, management of variability becomes critical to the efficiency and effectiveness of a complex system. This is increasingly the case in hospitals, where resource or manpower constraints limit capacity expansion and now force operation at very high occupancy. In these systems, natural variability cannot be eliminated, but it can be managed through application of a body of operations management methodologies specifically developed to optimally allocate fixed resources to meet random demand. Artificial variability, on the other hand, is best “managed” by elimination wherever possible.

We believe that the impact of variability control on hospital operations deserves further study. Although systems operating below capacity may realize minimal gains from control of variability, those operating near capacity may benefit greatly. Similarly, systems with few scheduled admissions will clearly benefit less than systems with large scheduled flows. In any case, systems that capably address both sources of variability will function optimally under their resource constraints, and any continuing shortfalls must be met either by additional resources or by fundamental changes in the way care is delivered. This conceptual framework provides an objective basis for resource allocation within a hospital and throughout health care generally.

Contemporary operations management research demonstrates that inflow variability greatly accentuates problems of rejection and unused capacity. In busy hospitals, competing patient flows frequently collide to produce intermittent demand surges that outstrip the supply of available beds. By smoothing demand peaks whenever possible, patient flow can be improved and rejections minimized. To understand why this is so, imagine an alternative: that in which scheduled admissions are planned to arrive at a fixed rate (here, perhaps equaling the historical average of four patients per day). In such an arrangement, four scheduled beds are prepared each day and all scheduled patients are admitted into them without risk of rejection or off-service transfer. The resulting aggregate variability (numerically expressed in our investigations as the sum of residuals) is significantly reduced and the remainder attached to a much smaller volume of unpredictable emergencies. This smaller volume of naturally variable admissions might then be accommodated more efficiently in the remainder of the unit and, after accounting for variations in length of stay, the appropriate number of total beds determined using queueing theory. Although a variety of daily scheduling tools have been evaluated for improving operating room flow and virtual separation of resources has been suggested elsewhere as a more efficient means of utilizing intensive care units, reduction of demand variability has not yet been investigated.

Although the literature is largely silent on this subject, the phenomenon of erratic patient flow with intermittent periods of extreme overload has long been familiar to anesthesiologists, intensivists, and critical care nurses working in busy units. Nonetheless, clinicians and healthcare managers seldom recognize that without specific controls aimed at smoothing elective schedules, arrival of a scheduled patient is no more predictable than the arrival of an emergency. We believe that this reducible source of artificial variation explains much of the difficulty in planning and staffing critical hospital services today. Clinicians and operating room managers who do not understand this may erroneously attribute wild fluctuations in workload to “natural” variations in demand rather than to alterable choices. As a result, only two remedies are usually considered in response to acute crowding: rationing or continued addition of staff and beds in wasteful cycles of expansion. As a more sensible alternative, we propose that hospitals first seek to control artificial variability as much as possible.

Reduction of artificial variability represents a promising area where astute managers may improve patient care without intruding on the specifics of clinical decision-making. As healthcare providers continue to be challenged by cost constraints and an aging population,
it is our belief that artificial variability will increasingly be viewed as symptomatic of a maladaptive and wasteful system. As hospitals crowd further, the price of ignoring such variability will more clearly manifest itself in unmet needs and diminished quality. Therefore, to maintain patient care standards under resource constraints, those who lead healthcare organizations should seek to identify and reduce artificial variability whenever possible.

References

1. Gosselin P. Amid national prosperity, ER's see a growing emergency. The Los Angeles Times. August 6, 2001