Chapter 2
Interdependency of Hospital Departments and Hospital-Wide Patient Flows

Alexander Kolker

Abstract This chapter presents a quantitative analysis of patient flows for a typical hospital-wide system that consists of a set of interdependent subsystems: Emergency Department (ED), Intensive Care Unit (ICU), Operating Rooms (OR), and Inpatient Nursing units (NU) including an effect of patient readmission within 30 days of discharge. It is quantitatively demonstrated that local improvement of one subsystem (ED) does not necessarily result in performance and throughput improvement of the entire system. It is also demonstrated that local improvement targets should be aligned to each other in order to prevent unintended consequences of creating another system bottleneck, and worsening the performance of downstream units.

Keywords System flows • Simulation • Throughput

1 Introduction

Modern medicine has achieved great progress in treating individual patients. This progress is based mainly on life science (molecular genetics, biophysics, biochemistry) and development of medical devices and imaging technology. However, relatively few resources and little technical talent have been devoted to the proper functioning of the overall health care delivery as an integrated system in which access to efficient care is delivered to many thousands of patients in an economically sustainable way. According to the report published jointly by the Institute of Medicine and National Academy of Engineering, a big impact on quality, efficiency, and sustainability of the health care system can be achieved using health care delivery engineering methods (Reid et al. 2006).

A. Kolker (✉)
API Healthcare, Hartford, WI 53027, USA
e-mail: alexanderkolker@yahoo.com

R. Hall (ed.), *Patient Flow: Reducing Delay in Healthcare Delivery*,
International Series in Operations Research & Management Science 206,
DOI 10.1007/978-1-4614-9512-3_2, © Springer Science+Business Media New York 2013

A system is a set of interdependent elements (subsystems) that form a complex whole that behaves in ways that these elements acting alone would not. Validated models of a system enable one to study the impact of alternative ways of running the system, alternative designs, different configurations and management approaches. System models enable one to experiment with systems in ways that cannot be used with real systems. A mathematical model of the system reveals important hidden and critical relationships that can be leveraged to influence the system's behavior in a desired direction.

Large systems are usually deconstructed into smaller subsystems using natural breaks in the system. The subsystems can be modeled and analyzed separately, but they should be reconnected back in a way that captures the most important interdependency between them. Goldratt and Cox (2004) state "A system of local optimums is not an optimum system at all; it is a very inefficient system." Similarly, Lefcowitz (2007) summarized "...maximization of the output of the various subsystems should not be confused with maximizing the final output of the overall system." Thus, analysis of a complex system is usually incomplete and can be misleading without taking into account subsystems' interdependencies.

An insight that systems behave differently than a combination of their standalone independent components is a fundamental management principle. A summary of other fundamental management principles is presented by Kolker (2012) and Hopp and Lovejoy (2013). The latter authors note "To qualify as a principle an insight must be both highly general (applicable to many settings) and stable (relevant now and in the future)... Overlooking the things that can be captured as principles can lead to fundamental errors. Hence, understanding management principles is extremely valuable as a starting point for managing operations."

The objective of this chapter is to present a quantitative illustration of the mentioned above fundamental principle using, as an example, patient flow for a typical hospital-wide system. The system consists of a set of interdependent subsystems: Emergency Department (ED), Intensive Care Unit (ICU), Operating Rooms (OR), and Inpatient Nursing units (NU). An effect of patient readmission within 30 days of discharge is also included in the system's overall patient flow. It will be quantitatively demonstrated that ED improvement targets (local improvements) should be aligned with capacity of the downstream units to handle increased patient flow out of ED in order to prevent unintended consequences of creating another system bottleneck, and worsening the performance of downstream units. This type of problem is a particular case of dynamic supply and demand balance (Kolker 2012).

Three basic components should be accounted for in these types of problems: (1) the number of patients (or, generally, any entities) entering the system at any point of time; (2) the number of entities leaving the system at any point of time after spending some variable time in the system, and (3) the limited capacity of the system that restricts the flow of entities through the system. All three components affect the flow of entities that the system can handle. A lack of proper balance between these components results in the system over-flow, bottlenecks or, sometimes, underutilization. It is widely acknowledged that the most powerful and

versatile methodology for quantitative analysis of the proper balance and dynamic variability in complex systems is discrete event simulation (DES). This methodology is used in this chapter for the analysis of a hospital-wide patient flow.

One notable example of a model that describes patient flow through a large hospital was presented by Cochran and Bharti (2006). Their model of a 400-bed hospital included surgical case, emergency and direct admissions. Also, they included blocking of patients caused by the finite bed capacities of each unit, two classes of patients (emergency and regular), and probability distributions that varied with the time of day and the day of the week. The authors proceeded to develop the optimal bed allocation to balance patient load. They also showed how blocking could be decreased if elective procedures were scheduled during off-peak times.

The remainder of this chapter will illustrate principles of patient flow across a hospital system by simulating a representative case-study hospital. The simulations will demonstrate dependencies between subsystems and their impact on overall system performance.

2 Hospital System Description

This section introduces a case-study hospital system that will later be simulated under different conditions. The hospital system is a tertiary referral center and a primary teaching community hospital for Southeast Wisconsin. It was chosen to represent a typical hospital system that includes the following interdependent high-level subsystems: (1) subsystem 1—Emergency Department (ED), overall bed capacity of 25 beds; (2) subsystem 2—Intensive Care Unit (ICU), overall bed capacity of 49 beds; (3) subsystem 3—Operating Rooms (OR), overall room capacity of 6 rooms; (4) subsystem 4—Inpatient Nursing units (NU), overall bed capacity of 360 beds. A high-level flow map (layout) of the entire hospital system is shown in Fig. 2.1.

Patients transported into the ED by ambulance (~18 %) and walk-in patients (~82 %) form an ED input flow. Two months arrivals patient volume (total 8,411 patients) is included. Some patients are treated, stabilized and released home (~74 %). ED patients admitted into the hospital (~26 %) form an inpatient input flow into the ICU, OR and/or NU. We assume in our simulation that patients waiting longer than 2 h in the ED waiting room leave the ED without being seen (LNS: lost-not-seen patients). About 70 % of admitted patients are taken into operating rooms (OR) for emergency surgery, about 20 % of admitted patients move into the ICU, and about 10 % of patients are admitted from ED into the inpatient nursing units (NU).

A diversion status is declared when ED, OR, ICU, or NU are at full bed capacity. The unit diversion is defined here as the percentage of operational time when the unit is at full bed capacity and can no longer accept new patients.

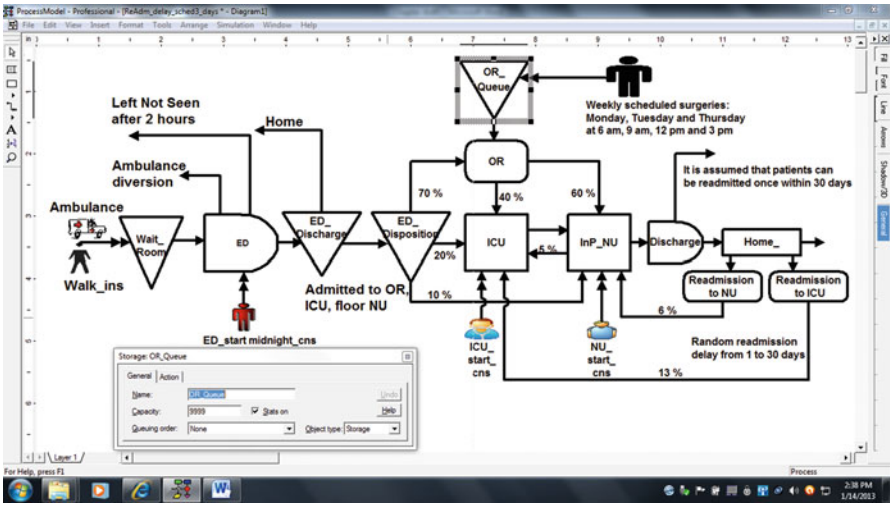

Fig. 2.1 Layout of the high-level simulation model of patient flow for a typical hospital system

About 40 % of postsurgical patients are admitted from OR into the ICU (direct ICU admission), while 60 % are admitted into the inpatient NU. However, some patients (about 5 %) are readmitted from the NU back to the ICU (indirect ICU admission).

The flow map includes 30-days of patient readmission feedback loops. These loops represent a uniformly distributed random delay in the range from 1 day to 30 days. It was reported that almost one-fifth (19.6 %) of patients nationwide who had been discharged from a hospital were rehospitalized within 30 days (MPAC 2007; Jencks et al. 2009). This case study system assumes 13 % readmission into ICU and 6 % readmission into inpatient nursing units (NU), totaling the overall 19 % of patients readmitted within 30 days after discharge.

The performance of the hospital needs significant improvement. The ED is on ambulance diversion a large percentage of time and a significant percentage of patients left not seen (LNS). The ICU frequently does not have beds for ED patient admissions or delays admission of postsurgical patients. The Surgical Department is often at capacity, and elective surgeries are frequently rescheduled. The hospital management needs to decide on the following: what unit/department to start with for process improvement projects; what type of projects to select; and process improvement performance metrics.

Because patient crowding is most visible in the ED, the hospital believed that inadequate ED throughput capacity was an issue. One way of increasing ED throughput capacity is by reducing ED patient length of stay (ED LOS) (Hopp and Lovejoy 2013). This might be accomplished in several ways. For example, Cho et al. (2011) constructed a computerized consultation management system in the ED of a tertiary care teaching hospital and evaluated the influence of the consultation management system on ED length of stay (LOS) and the throughput process.

ED personnel selected the department and on-call physician in the specialty department using the consultation management software and activated the automatic consultation process when specialty consultation was necessary. If the treatment plan had not been registered for 3 h, all of the residents in the specific department are notified of the delay in the treatment plan with a SMS message. If an admission or discharge order had not been made in 6 h, all of the residents and faculty staff in the specific department receive SMS messages stating the delay in disposition. The authors report significant reductions of ED LOS after implementing the system: the median ED LOS decreased from 417.5 min in the pre-system period to 311.0 min in the post-system period. The automated consultation and monitoring process formalized communication between physicians in ED with high consultation and admission rates.

Wang (2012) developed a simulation model of an emergency department (ED) at a large community hospital, Central Baptist Hospital in Lexington, KY aimed at determining the most critical process for improvement in quality of care in terms of patient length of stay. The author identified that floating nurse, combining registration with triage, mandatory requirement of physician's visit within 30 min, and simultaneous reduction of operation times of some most sensitive procedures can all result in substantial LOS reduction.

Oredsson et al. (2011) have undertaken a systematic literature review to explore which interventions improve patient flow in ED (33 studies with over 800,000 patients in total were included, mostly in European hospitals). The authors concluded that fast track for patients with less severe symptoms results in shorter waiting time, shorter length of stay, and fewer patients leaving without being seen. Team triage, with a physician in the team, will probably result in shorter waiting time and shorter length of stay and most likely in fewer patients leaving without being seen. There is only limited evidence that streaming of patients into different tracks, performing laboratory analysis in the emergency department or having nurses to request certain x-rays results in shorter waiting time and length of stay.

The next section analyzes the effect of various targets LOS on throughput and ambulance diversion in the ED as a separate subsystem.

3 ED as a Separate Subsystem: Effect of Patient Length of Stay on ED Ambulance Diversion

Emergency Department (ED) ambulance diversion due to "no available beds" has become a common problem in most major hospitals in the USA. A diversion status due to "no available ED beds" is usually declared when the ED census is close to or at the ED bed capacity. An ED remains in this status until beds become available when patients are moved out of ED (discharged home, expired, or admitted into the hospital as inpatients). The percentage of time when ED is on diversion is one of the

important ED performance metrics, along with the number of patients in queue in the ED waiting room, ED patient waiting time, and the percentage of patients left before they are seen (LNS). ED diversion results in low quality of care, dissatisfaction of patients and staff, and lost revenue for hospitals.

Patients' length of stay (LOS) in ED is one of most significant factors that affect the overall ED throughput and ED diversion (Blasak et al. 2003; Gunal and Pidd 2006; Miller et al. 2003; Simon and Armel 2003). There are generally two major groups of ED patients with different LOS distributions: (1) patients who are subsequently admitted as inpatients into the hospital (OR, ICU, floor nursing units), and (2) patients stabilized, treated, and discharged home without admission. Mayhew and Smith (2008) and Hopp and Lovejoy (2013) also recognized a key difference between these two groups. The latter authors note "Operational (ED) metrics can be divided into two categories: time and volume... The most basic time measure is LOS, which is usually measured separately for patients who are admitted to the hospital, patients who are kept for observation, and patients who are released." In order to effectively reduce ED diversion, the LOS of two basic patient groups should be quantitatively linked to ED diversion. Then the target LOS limits can be established based on ED patient flow analysis that significantly reduces or eliminates diversion.

Kolker (2008) provided a detailed analysis of the literature on ED LOS. One instructive article, Mayhew and Smith (2008), evaluates the consequences of a 4 h LOS limit mandated by the UK National Health Services (NHS) for the UK hospitals' Accident and Emergency Departments (A&ED). Because of significant difficulty to meet this standard, the target was later relaxed, allowing that not more than 2 % of patients could exceed 4 h LOS. However, Mayhew and Smith (2008) note that this relaxed standard was not sufficient to take the pressure of conformance from A&ED. These authors conclude "...a target should not only be demanding but that it should also fit with the grain of the work on the ground... Otherwise the target and how to achieve it becomes an end in itself." Further, "...the current target is so demanding that the integrity of reported performance is open to question." Another conclusion was "...the practicality of a single target fitting all A&ED will come under increasing strain." This work vividly illustrates the negative consequences of administratively mandated LOS targets that have not been based on the objective analysis of the patient flow and A&ED capabilities.

Another example of an administrative LOS target for ED department was the Position Statement on Emergency Department Overcrowding published by the Canadian Association of Emergency Physicians (CAEP 2007). The ED LOS benchmark suggested by CAEP was not to exceed 6 h in 95 % of cases for level 1, 2, and 3 patients. CAEP recommends the establishment of the national benchmark for total ED LOS that should be linked to objective ED performance.

Despite the considerable number of publications on ED patient flow and its variability (e.g., Carr and Roberts 2010; Jacobson et al. 2006), not much in the literature provides a practical solution for the target patient LOS: what it should be and how to establish it in order to reduce ED diversion to an acceptable low level, or to prevent diversion at all?

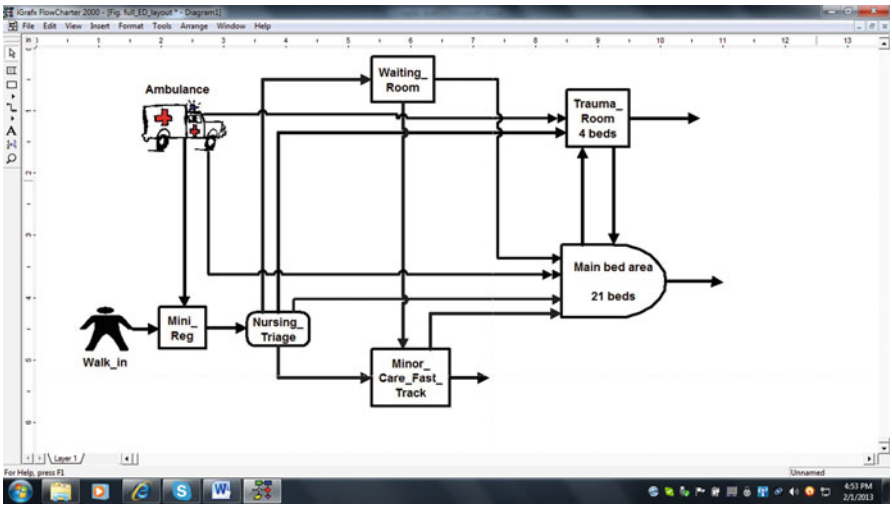

Fig. 2.2 ED structure of the study hospital. ED includes: mini-registration, nursing triage, waiting room, minor care/fast-track lane, trauma rooms, and the main patient bed area

To provide guidance on target LOS, the ED structure of the study hospital presented in Fig. 2.2 is analyzed (Kolker 2008). It includes a fast-track lane, minor care, trauma rooms, and the main patient beds area.

To focus on the effect of patient LOS on diversion for the entire ED, the detailed model layout was simplified in Fig. 2.3, keeping the model as simple as possible while capturing the objectives of the analysis (Law 2007).

Patients arrive into the ED by two modes of transportation: walk-in and ambulance. The week number, day of the week and arrival time characterize each patient in the arrival flow.

Discharged patients (released home or admitted as inpatients) moved out of the system according to their disposition routings. Patient flow "in and out" of the ED forms a dynamic supply and demand balance. The patient volume 8,411 for the 2-month period is included in the analysis. This patient volume is representative of subsequent months and years.

The critical element of the dynamics of the supply and demand balance is the time that patients spend in ED. This time was probabilistically fitted by continuous LOS distribution density functions, separately for admitted inpatients and discharged home patients. The ED length of stay distribution best fit for patients released home was Pearson 6 and for patients admitted to the hospital was log-logistic, as indicated in Fig. 2.4.

Because these LOS distributions represent a combination of many different steps of the patient move through the entire ED, from registration to discharge, they are simply the best analytical fit used to represent actual patient LOS data. Random numbers drawn from these distributions were used to perform multiple replications

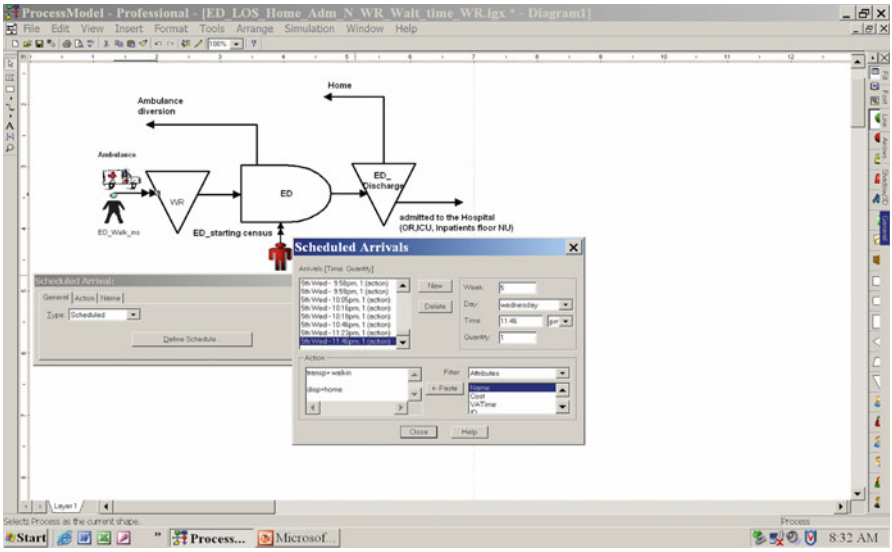

Fig. 2.3 Simplified ED structure included in the high-level hospital-wide simulation layout

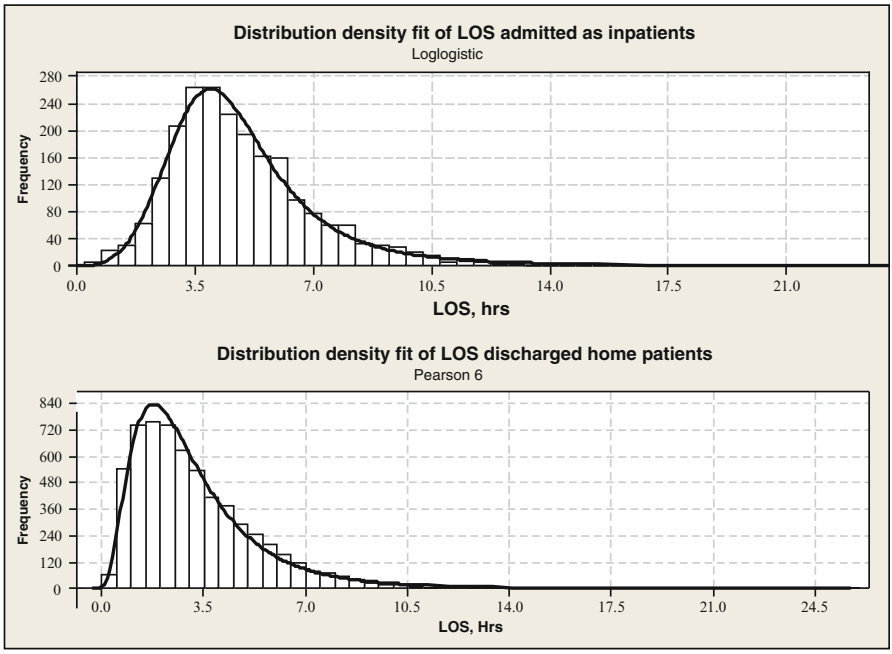

Fig. 2.4 Patient length of stay and best fit distribution density functions. *Top panel*: LOS for admitted inpatients. *Bottom panel*: LOS for discharged home patients

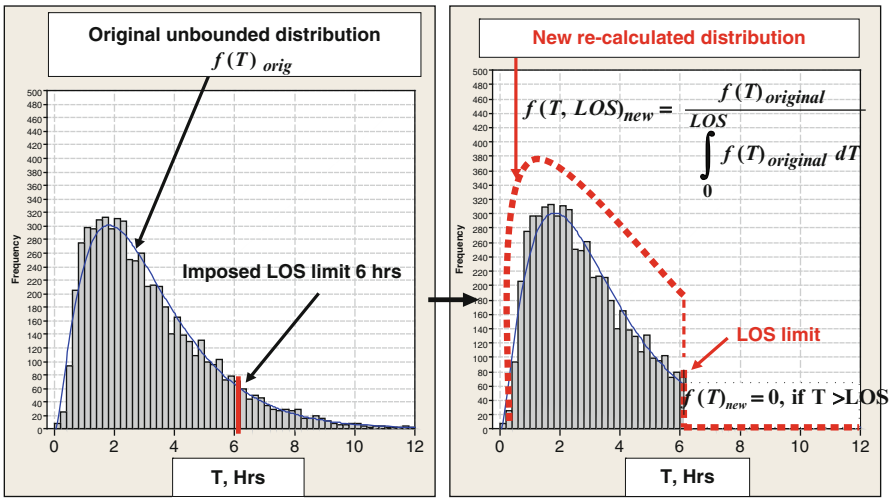

Fig. 2.5 Original LOS distribution density (*left panel*) and recalculated LOS distribution density with the imposed LOS limit (*right panel*)

using discrete event simulation (DES). Because the objective was to quantify the effect of the LOS limits (both for discharged home patients and admitted as inpatients) on the percent diversion, these limits were used as two variable simulation parameters. The original LOS distribution densities should be recalculated for each simulation scenario as functions of these parameters using the concept of conditional probability. Given the original LOS distribution density, $f(T)_{\text{orig}}$, and the limiting value, LOS_{lim}, the conditional LOS distribution density function of the new random variable restricted to LOS_{lim} is

$$f(T)_{\text{new}} = \frac{f(T)_{\text{orig}}}{\int_0^{\text{LOS}_{\text{lim}}} f(T)_{\text{orig}}\,\mathrm{d}T}, \qquad \text{if } T \text{ is less or equal to } \text{LOS}_{\text{lim}}$$

$$f(T)_{\text{new}} = 0, \qquad \text{if } T \text{ is greater than } \text{LOS}_{\text{lim}}$$

This is depicted in Fig. 2.5 (right panel, dotted bold line).

The conditional distribution density is a function of both the original distribution density and the simulation parameter LOS_{lim} (upper integration limits of the denominator integrals). These denominator integrals were first calculated and then approximated by these third order polynomials:

For discharged home patients:

If $\text{LOS}_{\text{lim}} \leq 10$ h, then

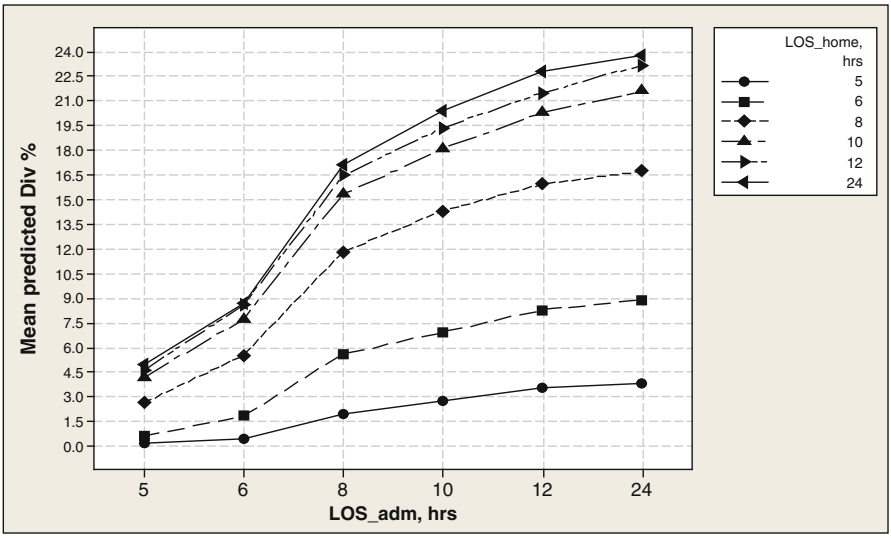

Fig. 2.6 Summary of the percentage diversion as a function of two parameters $\mathrm{LOS}_{\mathrm{lim}}$(home) and $\mathrm{LOS}_{\mathrm{lim}}$(adm)

$$\int_0^{\mathrm{LOS}_{\mathrm{lim}}} f(T)_{\mathrm{orig}}\,\mathrm{d}T = -0.2909 + 0.4013 \times \mathrm{LOS}_{\mathrm{lim}} - 0.04326 \times \mathrm{LOS}_{\mathrm{lim}}^2 + 0.001599 \times \mathrm{LOS}_{\mathrm{lim}}^3$$

else, the integral is approximately equal to 0.997.

For patients admitted into the hospital as inpatients:

If $\mathrm{LOS}_{\mathrm{lim}} \leq 10$ h, then

$$\int_0^{\mathrm{LOS}_{\mathrm{lim}}} f(T)_{\mathrm{orig}}\,\mathrm{d}T = -0.7451 + 0.3738 \times \mathrm{LOS}_{\mathrm{lim}} - 0.02188 \times \mathrm{LOS}_{\mathrm{lim}}^2 + 0.000157 \times \mathrm{LOS}_{\mathrm{lim}}^3$$

else, the integral is approximately equal to 0.994.

The model's adequacy was checked by running the simulation of the original baseline patients' arrival. The model's predicted percent diversion (~23.7 %) and the reported percent diversion (22.5 %) are close (in the range of a few percentage points). Thus, the model captures dynamic characteristics of the ED patients' flow adequately enough to mimic the system's behavior. A summary of results is presented in the plot Fig. 2.6.

It follows from this plot that several combinations of parameters LOS_{lim}(home) and LOS_{lim}(adm) would result in a low percent diversion. For example, if LOS_{lim}(home) is 5 h (low curve) then LOS_{lim}(adm) could be about 6 h with practically negligible diversion. Notice that Clifford et al. (2008) established the goal for ED LOS 6 h for inpatients to eliminate ambulance diversion and this metric is considered exceptional if less than 5 % of patients exceed this limit. Any other combination of LOS_{lim}(home) and LOS_{lim}(adm) could be taken from the graph to estimate a corresponding expected percent diversion. Thus, simulation helped to establish a quantitative link between an expected percent diversion and the limiting values of LOS. It has also suggested reasonable targets for the upper limits LOS_{lim}(home) and LOS_{lim}(adm).

Analysis of the actual LOS pattern in the study hospital indicated that a significant percentage of ED patients stayed much longer than the LOS targets required for low or no ambulance diversion. For example, ~24 % patients of a study hospital exceeded LOS_{lim}(adm) of 6 h, and stayed up to 24 h; ~17 % of patients exceeded LOS_{lim}(home) of 5 h, and also stayed up to 24 h (Fig. 2.4). These long LOS values were a root cause of ED closure and ambulance diversion.

Established LOS_{lim} targets could be used to better manage a daily patient flow. The actual current LOS is being tracked down and known for each individual patient. If the current LOS for the particular patient is close to the target LOS_{lim} a corrective action should be implemented to expedite a move of this patient. Multiple factors could contribute to the looming delay over the target LOS, such as delayed lab results or X-ray/CT; consulting physician is not available; no beds are downstream on hospital floor (ICU) for admitted patients, etc. Analysis and prioritizing the contributing factors to the over-the-target LOS is an important task. Notice that the average LOS that is frequently reported as one of the ED patient flow performance metric is not adequate to manage daily patient flow.

In order to calculate the average LOS, the data should be collected retrospectively for at least a few dozen patients. Therefore, it would be too late to make corrective actions to expedite a move of the particular patient if the average LOS becomes unusually high (whatever "high" means). In contrast, if the established upper limiting LOS targets were not exceeded for the great majority of patients, it would guarantee a low ED percent diversion, and the average LOS would be much lower than the upper limiting LOS lim. Marshall et al. (2005) and de Bruin et al. (2007) also discussed the shortcomings of reporting LOS only as averages (the flaw of averages) for the skewed (long tailed) data (as wells as Costa et al. (2003) and Savage (2009).

Emergency Departments of different hospitals differ by their structure, patient mix, LOS distribution, and bed capacity. However, the overall simulation methodology presented here will be valid regardless of the particular hospital ED.

4 Intensive Care Unit (ICU) as a Separate Subsystem: ICU Diversion

An Intensive Care Unit (ICU) is often needed for patient care. Demand for ICU beds comes from emergency, add-on and elective surgeries. Emergency and add-on surgeries are random and cannot be scheduled in advance. Elective surgeries are scheduled ahead of time. However, they are often scheduled for the daily block-time driven mostly by physician priorities. (Daily block time is the time in the operating room that is allocated to the surgeon or the group of surgeons on particular days of the week to perform a particular type of surgical service.) Usually elective surgery scheduling does not take into account the competing demand for ICU beds from the emergency and add-on cases.

Because of the limited capacity of ICU beds, a mismatch between bed availability and the flow of unscheduled patients can result in the Emergency Department (ED) diversion. This is an example of a system disconnect caused by the interdependent and competing demands among patient flows in a complex system: the upstream problem (ED closure) is created by the downstream problem (no ICU beds).

Usually two types of variability affect the system's patient flow: natural process flow variability and scheduled (artificial) flow variability (Litvak et al. 2001; Haraden et al. 2003). Patients can be admitted into an ICU from the Emergency Department (ED), other local area hospitals, inpatient nursing units, and/or operating rooms (OR). Patients admitted into ICU from ED, other local area hospitals, and inpatient nursing units are primary contributors to the natural random flow variability because the timing of these admissions is not scheduled in advance and is unpredictable.

Admissions into ICU from the OR include emergency, add-on, and elective surgeries. Elective surgeries are defined as surgeries that could be delayed safely for the patient by at least 24 h (or usually much longer). Emergency and add-on surgeries also contribute to the natural process flow variability. Because this type of variability is statistically random, it is beyond hospital control. It cannot be eliminated (or even much reduced). However, some statistical characteristics can be modeled based on data over a long period of time.

Elective surgeries that require postoperative admission into ICU contribute to the scheduled (artificial) flow variability. Elective surgery scheduling is driven by individual priorities of the surgeons and their availability, which reflects other commitments (teaching, research, etc.). This variability is usually within the hospital control, and it can be reduced or eliminated with proper management of the scheduling system. It is possible to manage the scheduling of the elective cases in a way to smooth (or to daily load level) overall patient flow variability. A daily load leveling would reduce the chances of excessive peak demand for the system's capacity and, consequently, would reduce diversion. There are quite a few publications in which the issues of smoothing surgical schedules and ICU patient flow are discussed. Kolker (2009) provided a detailed analysis of the literature.

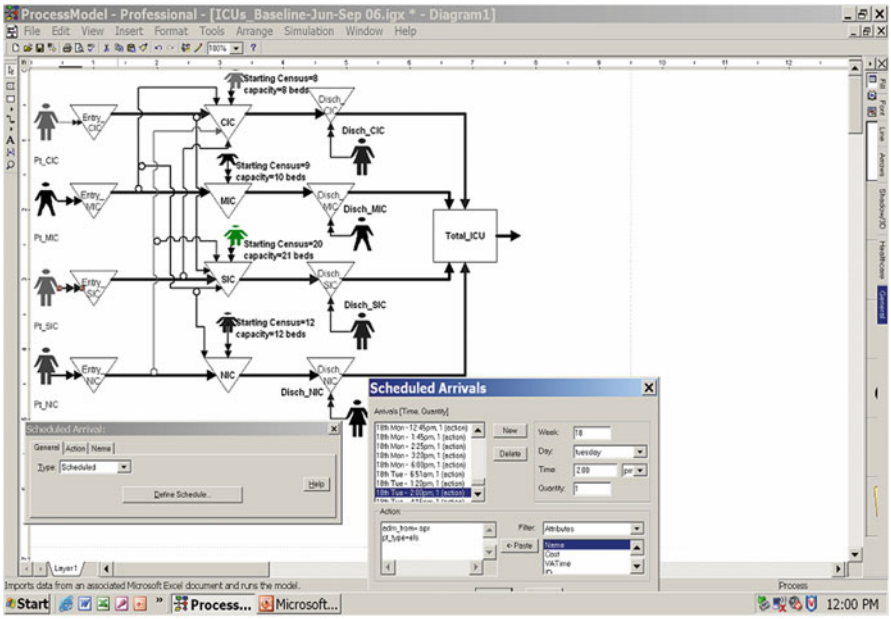

Fig. 2.7 Layout of the ICU patient flow model

Layout of the ICU model of the study hospital is represented in Fig. 2.7. The entire ICU system includes four specialized ICU units: Cardio (CIC), bed capacity is 8; Medical (MIC), bed capacity is 10; Surgical (SIC), bed capacity is 19; and Neurological (NIC), bed capacity is 12. The total ICU bed capacity is 49. Patients admitted into each ICU unit form an arrival flow. The week number, the day of the week, and the admitting time characterize each patient in the arrival flow. Each discharged patient is also characterized by the week number, the day of the week, and the discharge time.

Patient flow "in and out" forms a dynamic supply and demand balance (supply of ICU beds and patient demand for them). ICU length of stay is assumed to be in the range from 1 day to 3 days, with 1.5 days most likely, represented by a triangle distribution. If there is no free bed at the time of admission in the particular primary ICU unit, then the patient is moved into another ICU unit using alternate type routings (depicted by the thin lines between the units, Fig. 2.7). Patient moves followed the following hospital's rules to deal with the excess capacity of the particular ICU units: (1) if no beds are available in CIC then move to SIC; (2) if no beds are available in MIC then move to CIC else move to SIC else move to NIC; (3) if no beds are available in NIC then move to CIC else SIC.

When the patient census of the ICU system hit its bed capacity limit, then an ICU diversion is declared due to "no ICU beds." In the study hospital the number of elective cases was about 21 % of all ICU admissions.

The model adequacy check was performed by comparing the predicted percent diversion for the different time periods and the actual percent diversion. It could be concluded (Kolker 2009) that the model captures dynamic characteristics of the ICU patient flow adequately (within 1–2 % from the actually reported values) to mimic the system's behavior and to compare alternative ("what-if") scenarios.

5 OR as a Separate Subsystem

An OR suite has six interchangeable operating rooms used both for ED emergency and scheduled surgeries. There are two general surgery operating rooms, one operating room each for trauma, cardiovascular, orthopedic, and neurosurgery. The operating rooms are interchangeable, so if the primary surgical room is busy, then the patient can be moved into another room if it is available. Emergency cases have higher priority than scheduled ones. Typically four OR cases are scheduled three times a week on Monday, Tuesday and Thursday at 6 am, 9 am, 12 pm and 3 pm. Usually there are no scheduled surgeries on Wednesday, Friday and weekends because surgeons have other commitments, such as teaching, research, manuscripts preparation, consulting, training, etc. However, more elective cases are occasionally added if needed and can be included in the simulation model. This artificial scheduling variability illustrates an observation (McManus et al. 2003) that "...variability is particularly high among patients undergoing scheduled surgical procedures, with variability of scheduled admissions exceeding that of emergencies." Further, "One result of this variability is a widely ranging demand for critical care services (ICU) that, in units operating at high capacity, frequently responsible for patients being placed off-service or denied access altogether."

Scheduled cases form a separate OR admissions flow, as indicated on the diagram Fig. 2.1. Elective surgery duration depends on surgical service type, such as general surgery, orthopedics, neurosurgery, etc. For model simplicity, elective surgery duration was weighted by each service percentage, and the best statistical distribution fit was identified (inverse Gaussian in this case). Emergency surgery duration was best fit by Pearson 6 statistical distribution.

About 40 % of postsurgical patients are admitted from OR into ICU (direct ICU admission), while 60 % are admitted into inpatient nursing units (NU).

6 Inpatient Nursing Units as a Separate Subsystem

Total inpatient nursing unit (NU) bed capacity was 360 beds. Patient length of stay (LOS) in inpatient NU was assumed to be in the range from 2 days to 10 days, with the most likely of 5 days, represented by a triangle distribution. Simulated census for a typical week is represented in Fig. 2.8. It is clear that the bed capacity limits are consistently hit on a daily basis, usually at the middle of the day, except on weekends.

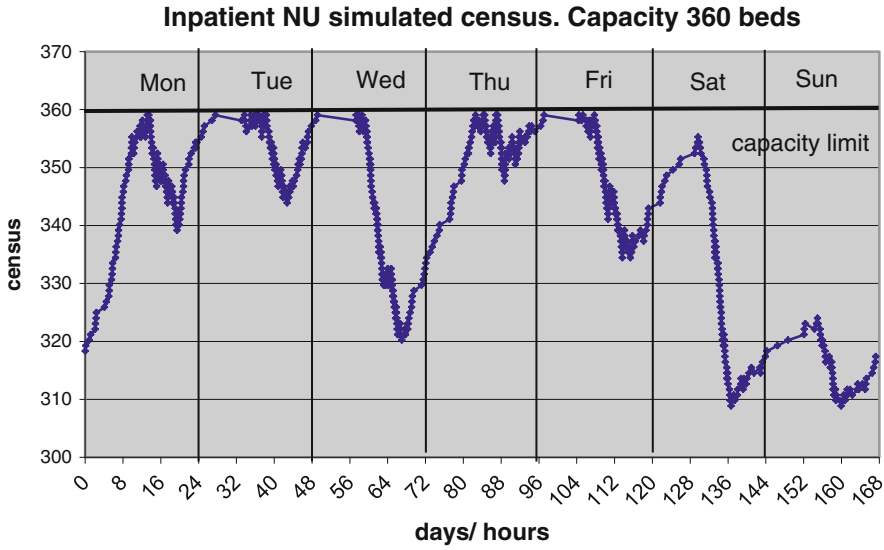

Fig. 2.8 Inpatient nursing units NU. Simulated census of a typical week

7 Reconnecting Separate Units: The Entire Hospital System

As was discussed in the introduction, large complex hospital systems and multi-facility clinics are usually analyzed as deconstructed smaller subsystems or units. Most published simulation models focus on the separate analysis of these individual units. However, according to the principles of complex systems analysis, these separate subsystems (units) should be reconnected back in a way that captures the most important interdependency between them. Simulation models that capture interaction of major units in a hospital, and the information that is obtained from analysis of the system responses as a whole can be invaluable to hospital planners and administrators.

This section illustrates a practical application of this system-engineering principle. High-level simulation models of the separate main hospitals units, i.e., ED, ICU, OR, and inpatient NU patient flow, have been described in the previous sections. These units are not stand-alone systems but they are closely interdependent, as indicated in Fig. 2.1.

The output of the ED model for patients admitted into the hospital (ED discharge) now becomes an ICU, OR and NU input through ED disposition. In our case study, about 70 % of admitted ED patients are taken into operating rooms (OR) for emergency surgery; about 20 % of admitted ED patients move directly into ICU; and about 10 % of patients admitted from ED are taken into combined inpatient nursing units.

At the simulation start on week 1, at the Monday midnight all units are empty, while in reality they are not. We are interested in this analysis in a long-term steady-state period rather than a transient period. Therefore, at the simulation start, the empty units should be prefilled to the typical midnight census values, which are 15, 46 and 350 patients for the ED, ICU and NU, respectively.

A summary of simulations for the various performance metrics is shown in Table 2.1.

Eight performance metrics (95 % Confidence Intervals-CI) are indicated in column 1. Baseline metrics that correspond to patient ED LOS up to 24 h are presented in column 2. It was demonstrated in Sect. 3 that the ambulance diversion for stand-alone ED becomes very low if improvement efforts reduced LOS for patients admitted into the hospital to less than 5 h and LOS for released home patients to less than 6 h (from ED registration to ED discharge). However, because of interdependency of the ED and the downstream units, four out of eight metrics became much worse (columns 3 and 4). The ED bottleneck just moved downstream into the OR and ICU because of their inability to handle the increased patient volume from ED. Thus, aggressive process improvement in one subsystem (ED) resulted in a worse situation in other interrelated subsystems (OR and ICU). ED improvement is not necessarily translated into the goal of increasing the throughput of the entire hospital system. It turns out that patient flow is a property of the entire hospital system rather than the property of the separate departments/units. A detailed analysis is required of the overall hospital system patient flow and the interdependency of subsystems/units in order to establish the system's weak link and the right units for process improvement projects priority.

If, instead of too aggressive ED LOS reduction, a less aggressive improvement is implemented, e.g., ED LOS is not more than 9 h for patients admitted to the hospital, then none of the eight metrics become much worse than the baseline state (columns 5 and 6). While in this case ED performance is not as good as it could be, it is still better than it was at the baseline level. At the same time, a less aggressive local ED improvement does not make the ICU, OR, and NU much worse. In other words, the less aggressive ED improvement is better aligned with the ability of the downstream units to handle the increased patient volume.

Thus, from the entire hospital system standpoint, the primary focus of process improvement should be on the ICU because it has the highest percent of patients waiting for admission more than 1 h and the highest diversion, followed by the NU and ED. If process improvement aimed at reducing patient LOS starts in the upstream unit—ED without addressing first capacity to handle increased patient flow of the downstream units—ICU and NU—it will only result in more patients that are formally discharged from ED but boarded there waiting for admission to ICU and NU, as indicated by the increased diversion and waiting time for latter units in Table 2.1. Otherwise, even if the ED makes significant progress in its patient LOS reduction program based on formal discharge time, this progress will not translate into improvement of the overall hospital-wide patient flow. Of course, many other scenarios could be analyzed using the simulation model to find out how to improve the entire hospital-wide patient flow rather than that for each separate

Table 2.1 Summary of simulation results for the hospital system patient flow performance metrics

	1	2	3	4	5	6
	Performance metrics	Baseline state	Aggressive ED improvement: admitted LOS 5 h; released home LOS 6 h	Better or worse than baseline?	Less aggressive ED improvement: admitted LOS 9 h; released home LOS 10 h	Better or worse than baseline?
1	95 % CI of ED diversion	23.6–23.9 %	3.0–3.1 %	Much better	20.9–21.1 %	Better
2	95 % CI of the percentage of patients left not seen	7.7–8.1 %	0 %	Much better	4.8–5.0 %	Better
3	95 % CI of the percentage of patients waiting admission to OR from ED longer than 1 h	0.05–0.1 %	0.4–0.7 %	Much worse	0.1–0.3 %	A little bit worse
4	95 % CI of OR diversion	0.7–0.8 %	1.9–2 %	Much worse	0.9–1.1 %	A little bit worse
5	95 % CI of the percentage of patients waiting admission to ICU from ED longer than 1 h	25.4–28.2 %	33.4–36.6 %	Much worse	28.3–31.3 %	A little bit worse
6	99 % CI of ICU diversion	16.2–17.9 %	23.4–25.5 %	Much worse	18.9–20.7 %	A little bit worse
7	95 % CI of the percentage of patients waiting admission to NU from ED longer than 1 h	29.5–31.4 %	29.4–31.3 %	Not much different	28.9–30.8 %	Not much different
8	95 % CI of NU diversion	8.5–8.6 %	9.0–9.2 %	Slightly worse	8.7–8.8 %	Not much different

local subsystem/unit. This illustrates one of the fundamental principles of system analysis.

In order to improve ICU throughput performance, more rigorous ICU admission and discharge criteria could be applied. If, for example, ICU admission volume is reduced to 15 % of the total ED disposition patient volume, then the simulation model indicates that the percentage of patients waiting more than 1 h will be about 21 % (down from about 28 to 31 %), and ICU diversion will be about 11 % (instead of 19–20 %).

Another option is reducing the maximum ICU length of stay from 3 days (72 h) to, for example, 2.75 days (66 h) instead of limiting the ICU admission volume. In this case, the percentage of ICU patients waiting more than 1 h will be about 22 % and ICU diversion will be about 13 %. These ICU performance metrics are very close to the above with the reduced admission volume.

Of course, a combination of the above scenarios is possible for further improvement. Many other scenarios could also be modeled to find out how to improve the entire hospital system patient flow rather than each separate hospital department.

8 Effect of Reduced Avoidable 30 Day Readmission Rate

It was already mentioned that nearly one-fifth of patients discharged from a hospital return within 30 days in the USA (MPAC 2007). Identifying and reducing avoidable readmissions will improve patient safety, enhance quality of care, and lower health care spending. That is why policymakers, consumers, hospital leaders and the medical community are focused increasingly on readmissions to hospitals. Most recently, in the Patient Protection and Affordable Care Act (ACA), the US Congress enacted the Hospital Readmissions Reduction Program (HRRP) under which Medicare will penalize hospitals for higher-than-expected rates of readmissions, beginning in 2013. Some hospitals are moving forward with efforts to reduce readmissions and improve quality of care. For example, Metro Health Hospital in Wyoming initiated its Congestive Heart Failure (CHF) readmissions program and cut its avoidable CHF readmission rate to 7.4 % (AHA 2011).

A thirty-day readmission was simulated here as feedback loops of the discharged patients with random uniformly distributed delay in the range from 1 to 30 days (Fig. 2.1). Suppose, for example, that the study hospital analyzed in this chapter cut its total avoidable 30 days readmission rate to about 10 % (including 2 % ICU readmission rate and 8 % inpatient NU readmission rate). Simulation modeling with this lower readmission rate indicated that the ICU performance would markedly improve: the ICU percentage of patients waiting more than 1 h dropped to about 17 % (down from about 28 to 31 %) and ICU diversion is down to 6 % (rather than 19–21 %). Thus, reduction of the avoidable readmission rate not only reduces the monetary penalty but also significantly improves performance characteristics.

9 Conclusions

Analysis of a complex system is usually incomplete and can be misleading without taking into account subsystems' interdependencies. The insight that systems behave differently than a combination of their stand-alone independent components is a fundamental management principle.

It was demonstrated in this chapter that the performance of a hospital-wide system could inadvertently be jeopardized because locally oriented improvement in one process or department worsens performance of the overall system. It may be said "Curing the Process May Kill the System" (Kamanth et al. 2011). It was quantitatively demonstrated using simulation modeling that aggressive process improvements implemented in the ED to reduce patient length of stay (good for the ED) can result in increasing ICU and operating room wait time and percent diversion (bad for the ICU and OR). Thus, improvements in an upstream subsystem may worsen performance of the downstream units and the overall system—at least for some performance measures. Therefore, improvement of the upstream units should be aligned with the ability of the downstream units to handle the increased patient volume. The ability of system analysis and simulation modeling methodology to incorporate a broader system-thinking approach is one of its advantages over some local process-specific improvement methods, such as plan-do-study-act (PDSA) learning cycles (Kamanth et al. 2011).

References

AHA. (2011). *Trend watch. Examining the drivers of readmissions and reducing unnecessary readmissions for better patient care*. Washington, DC: American Hospital Association (AHA). September 2011.

Blasak, R., Armel, W., Starks, D., & Hayduk, M. (2003). The use of simulation to evaluate hospital operations between the ED and medical telemetry unit. In S. Chick et al. (Eds.), *Proceedings of the 2003 winter simulation conference* (pp. 1887–1893). Washington, DC: IEEE.

CAEP. (2007). Canadian Association of Emergency Physicians. Position statement on emergency department overcrowding, February issue. www.CAEP.ca

Carr, S., & Roberts, S. (2010). Computer simulation in healthcare. Chapter 14. In Y. Yih (Ed.), *Handbook of healthcare delivery systems*. Boca Raton, FL: CRC Press.

Cho, S. J., Jeong, J., Han, S., Yeom, S., Park, S. W., Kim, H. H., et al. (2011). Decreased emergency department length of stay by application of a computerized consultation management system. *Academic Emergency Medicine, 18*(4), 398–402. doi:10.1111/j.1553-2712.2011.01039.x.

Clifford, J., Gaehde, S., Marinello, J., Andrews, M., & Stephens, C. (2008). Improving inpatient and emergency department flow for veterans. *Improvement report*. Institute for Healthcare Improvement. Retrieved from http://www.IHI.org/ihi

Cochran, J., & Bharti, A. (2006). A Multi-staged stochastic methodology for whole hospital bed planning under peak loading. *International Journal of Industrial and System Engineering, 1*, 8–36.

Costa, A., Ridley, S., Shahani, A., Harper, P., De Senna, V., & Nielsen, M. (2003). Mathematical modeling and simulation for planning critical care capacity. *Anesthesia, 58*, 320–327.

De Bruin, A., van Rossum, A., Visser, M., & Koole, G. (2007). Modeling the emergency cardiac in-patient flow: An application of queuing theory. *Health Care Management Science, 10*, 125–137.

Goldratt, E., & Cox, J. (2004). *The goal* (3rd ed.). Great Barrington, MA: North River Press.

Gunal, M., & Pidd, M. (2006). Understanding accident and emergency department performance using simulation. In L. Perrone et al. (Eds.), *Proceedings of the 2006 winter simulation conference* (pp. 446–452). Washington, DC: IEEE.

Haraden, C., Nolan, T., & Litvak, E. (2003). Optimizing patient flow: Moving patients smoothly through acute care setting. Institute for Healthcare Improvement Innovation Series 2003. White Papers 2: Cambridge, MA.

Hopp, W., & Lovejoy, W. (2013). *Hospital operations: Principles of high efficiency health care*. Fontana, CA: FT Press, Upper Saddle River, NJ.

Jacobson, H., Hall, S., & Swisher, J. (2006). Discreet-event simulation of health care systems. In R. Hall (Ed.), *Patient flow: Reducing delay in healthcare delivery* (pp. 210–252). New York, NY: Springer.

Jencks, S., Williams, M., & Coleman, E. (2009). Re-hospitalizations among patients in medicare fee for service program. *New England Journal of Medicine, 360*, 1418–1428.

Kamanth, J., Osborn, J., Roger, V., & Rohleder, T. (2011). Highlights from the third annual mayo clinic conference on systems engineering and operations research in health care. *Mayo Clinic Proceedings, 86*(8), 781–786.

Kolker, A. (2008). Process modeling of emergency department patient flow: Effect of patient length of stay on ED diversion. *Journal of Medical Systems, 32*(5), 389–401.

Kolker, A. (2009). Process modeling of ICU patient flow: Effect of daily load leveling of elective surgeries on ICU diversion. *Journal of Medical Systems, 33*(1), 27–40.

Kolker, A. (2012). *Healthcare management engineering: What does this fancy term really mean? Springer_Briefs series in healthcare management & economics*. New York, NY: Springer. 122.

Law, A. (2007). *Simulation modeling and analysis* (4th ed.). New-York: McGraw-Hill.

Lefcowitz, M. (2007, February 26). Why does process improvement fail? Builder-AU by Developers for developers. Retrieved from www.builderau.com.au/strategy/projectmanagement/

Litvak, E., Long, M., Cooper, A., & McManus, M. (2001). Emergency department diversion: Causes and solutions. *Academic Emergency Medicine, 8*, 1108–1110.

Marshall, A., Vasilakis, C., & El-Darzi, E. (2005). Length of stay-based patient flow models: Recent developments and future directions. *Health Care Management Science, 8*, 213–220.

Mayhew, L., & Smith, D. (2008). Using queuing theory to analyze the Government's 4-h completion time target in Accident and Emergency departments. *Health Care Management Science, 11*, 11–21.

McManus, M., Long, M., Cooper, A., Mandel, J., Berwick, D., Pagano, M., et al. (2003). Variability in surgical caseload and access to intensive care services. *Anesthesiology, 98*(6), 1491–1496.

Miller, M., Ferrin, D., & Szymanski, J. (2003). Simulating Six Sigma Improvement Ideas for a Hospital Emergency Department. In S. Chick et al. (Eds.), *Proceedings of the 2003 winter simulation conference* (pp. 1926–1929). Washington, DC: IEEE.

MPAC (Medicare Payment Advisory Commission). (2007). Payment policy for inpatient readmissions. *Report to the congress: Promoting greater efficiency in Medicare*. Washington, DC.

Oredsson, S., Jonsson, H., Rognes, J., Lind, L., Göransson, K., Ehrenberg, A., et al. (2011). A systematic review of triage-related interventions to improve patient flow in emergency departments. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 19*, 43. doi:10.1186/1757-7241-19-43.

Reid, P., Compton, W., Grossman, J., & Fanjiang, G. (Eds.). (2006). Building a better system: A new engineering healthcare partnership. Washington, DC: National Academy of Engineering and Institute of Medicine. The National Academy Press

Savage, S. (2009). *The flaw of averages*. Hoboken, NJ: Wiley. 392.

Simon, S., & Armel, W. (2003). The use of simulation to reduce the length of stay in an emergency department. In S. Chick et al. (Eds.), *Proceedings of the 2003 winter simulation conference* (pp. 1907–1911). Washington, DC: IEEE.

Wang, J. (2012). Reducing length of stay in emergency department: A simulation study at a community hospital. *IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans, 42*(6), C1–1309.

http://www.springer.com/978-1-4614-9512-3

Patient Flow
Reducing Delay in Healthcare Delivery
Hall, R. (Ed.)
2013, XII, 553 p. 107 illus.,
ISBN: 978-1-4614-9512-3