J Med Syst (2008) 32:389–401
DOI 10.1007/s10916-008-9144-x

ORIGINAL PAPER

Process Modeling of Emergency Department Patient Flow: Effect of Patient Length of Stay on ED Diversion

Alexander Kolker

Received: 23 December 2007 / Accepted: 13 March 2008 / Published online: 25 April 2008

Abstract A discreet event simulation methodology has been used to establish a quantitative relationship between Emergency Department (ED) performance characteristics, such as percent of time on ambulance diversion and the number of patients in queue in the waiting room, and the upper limits of patient length of stay (LOS). A simulation process model of ED patient flow has been developed that took into account a significant difference between LOS distributions of patients discharged home and patients admitted into the hospital. Using simulation model it has been identified that ED diversion could be negligible (less than ~0.5%) if patients discharged home stay in ED not more than 5 h, and patients admitted into the hospital stay in ED not more than 6 h Using full factorial design of experiments with two factors and the model's predicted percent diversion as a response function, other combinations of LOS upper limits have been determined that would result in low ED percent diversion as well. It has also been determined that if the number of patients exceeds 11 in queue in ED waiting room then the diversion percent is rapidly increasing.

Keywords ED diversion · Length of stay · Process model simulation · What-if scenarios · Design of experiments

Introduction

Emergency Department (ED) ambulance diversion due to 'no available beds' status has become a common problem in most major hospitals nationwide [1–3]. A diversion status due to 'no available ED beds' is usually declared when the ED census is close to or at the ED beds capacity limit. 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). Percent of time when ED is on diversion is one of the important ED patient performance metrics, along with the number of patients in queue in ED waiting room, or ED patient waiting time. ED diversion results in low quality of care, dissatisfaction of patients and staff, lost revenue for hospitals.

Patients' length of stay (LOS) in ED is one of most significant factors that affect ED diversion [4–8]. There are generally two major groups of patients with different LOS distributions: (1) patients admitted as inpatients into the hospital (OR, ICU, floor nursing units), and (2) patients stabilized, treated and discharged home.

A key difference between these two groups was also recognized in work [10].

In order to effectively attack the problem of ED diversion reduction the LOS of these two groups should be quantitatively linked to percent ED diversion. Then the target LOS limits should be established based on ED patient flow analysis.

A number of publications are available in which the importance of having ED LOS target was discussed: what it should be and how to establish it.

For example, the objective of work [4] was to study performance of Accident & Emergency department (A&ED) in the UK hospitals. The performance was measured as percentage of patients that exceeded established LOS target. Patient LOS was defined from the arrival time (registration) to discharge home or admission into hospital time. In 2002 the UK Department of Health established that the target for LOS should not exceed 4 h. In 2004 it was allowed that not more than 2% of patients could exceed 4 h LOS. However these targets have not

A. Kolker (✉)
Children's Hospital of Wisconsin,
Milwaukee, WI, USA
e-mail: alexanderkolker@yahoo.com

been linked quantitatively to A&ED diversion reduction. The authors [4] stated that 2% of patients that allowed exceeding 4 h target LOS were not sufficient enough to take the pressure of conformance from A&ED. The underlying logic of this study was to find out how the administratively established strict LOS target could be met. However, it was not addressed how to objectively establish the realistic LOS target in the first place.

Another example of the administratively suggested LOS target for ED department was Position Statement on Emergency Department Overcrowding published by the Canadian Association of Emergency Physicians (CAEP) [9]. The ED LOS benchmarks suggested by CAEP should have been not more than 6 h in 95% cases for levels 1, 2 and 3 patients. For levels 4 and 5 patients, LOS should not exceed 4 h in 95% cases. At the same time, CAEP recommends the establishment of national benchmark for total ED LOS that should be linked to objective ED performance [9].

An instructive article was recently published [10] in which 4 h LOS target in the UK hospitals A&ED was evaluated. One of the main conclusions of this work was '…that 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 that '…the practicality of a single target fitting all A&ED will come under increasing strain' [10] This work vividly illustrated the negative consequences of the administratively mandated LOS targets that have not been based on the objectives analysis of the patient flow and an A&ED capability to handle it.

A number of publications are available in which reduction of patient LOS in ED of large US hospitals is discussed. For example, using simulation of the ED operations the authors of work [6] showed that the hospital would not meet its goal for patient ED LOS. However there was no information in this work on what this goal was, and how it was established.

In work [5] it was reported that patient LOS in ED with a fast track lane was reduced by almost 25% for patients with low severity code. However there was no information whether such a reduction was enough to meet the ED performance goals, and what the LOS target was.

Similarly to work [5], it was concluded in work [7] that a fast track lane in ED would expedite non-critical patients through the system and shorten their LOS in ED. This would result in more patients being seen in the ED with shorter LOS. However there was also no specific information on LOS target, and how much LOS reduction would result in performance improvement.

In work [8] it was stated that '…the overall time patients spend in ED is longer than management would like it to be'. Here, again, no specific information was given on what this time was, and how much reduction of LOS was needed in order to get the desired benefit (which could be, e.g., zero or low single digits percent diversion, or acceptable patient waiting time in ED waiting room).

Thus, despite a considerable number of publications on the ED patient flow and its variability, there is not much in the literature that could help to answer a practically important question regarding 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.

Therefore, a methodology that could quantitatively link the patient LOS limits and ED performance metrics would have a considerable practical value.

Two main approaches are currently used to model patient flow: (1) queuing theory and (2) process model simulation. Both are based on principles of operations research. It is an area of applied mathematics developed to quantitatively analyze characteristics of the processes with a random demand for services and available capacity (resources) to provide those services.

Queuing theory uses closed mathematical formulas to describe a number of pre-determined simplified models of the real processes. Most widely used queuing models for which relatively simple closed analytical formulas have been developed were specified as $M/M/s/c$ type [11]. These models assume a queue that is served by s providers and c spaces in the system. It is assumed that arrivals into the queue form a Poisson process. The latter is, on definition, an ordinary stochastic process of independent events. Service time is assumed to follow an exponential distribution or, sometimes, uniform or Erlang distribution. (M stands for Markov since Poisson process is a particular case of a stochastic process with no 'after-effect' or no memory, known as continuous time Markov process).

However these assumptions are rarely valid for ED processes. For example, actual records indicate that several patients sometimes arrive in ED at the same time, and/or the probability of new patient arrivals could depend on the previous arrivals when ED is close to its capacity. These possibilities alone make the arrival process a non-ordinary one with after-effect, i.e. non-Poisson process for which queuing formulas are not valid.

Despite its limited applicability to actual hospital patient arrivals pattern, Poisson process (as well as an exponential service time) is widely used in operations research because of its mathematical convenience and an apparent analytical simplicity.

An example of using these two methodologies for the analysis of the performance of practically the same A&ED in the UK was presented in already referenced works [4] and [10].

A discreet event simulation model of patient process flow was developed in work [4]: patient was first registered, triaged, treated, sent to X-ray, re-evaluated and discharged. Each step has a different service time distribution (triangular, log-normal) depending on patients' triage category. The model's output was the total times of patients in A&ED and percentage of patients who exceeded the 4 h LOS target.

An approach based on queuing theory was used in work [10]. In order to make the queuing model tractable the authors made a significant simplification by presenting the workflow as a series of stages. The stages could include initial triage, diagnostic tests, treatment, and discharge. Some patients experienced only one stage while others more than one. However, '…what constitutes a "stage" is not always clear and can vary…and where one begins and ends may be blurred' [10]. The authors assumed a Poisson arrival and exponential service time but then used actual distribution service time for 'calibration' purposes. Moreover, the authors observed that exponential service time for the various stages '…could not be adequately represented by the assumption that the service time distribution parameter was the same for each stage'. In the end, all the required calibrations, adjustments, fitting to the actual data made the model to lose its main advantage as a queuing model: its analytical simplicity and transparency. On the other hand, all queuing formulas assumptions still remained. The authors themselves stated that '…we make no claim that this is best possible model of its type…' [10].

Thus, process model simulation approach seems to be much more flexible and versatile [14]. It is free from assumptions of the particular type of the arrival process (Poisson or not), as well as the service time (exponential or not). The system structure (flow map) could be of any complexity, and custom action logic can be built in to mimic practically any features of the real system behavior.

Many currently available simulation software packages (ProcessModel, ProModel, Arena, Simula8, etc) [12] provide a user-friendly interface that makes the efforts of building a realistic simulation model not more demanding than the efforts needed to make simplifications, adjustments and calibrations to develop a rather complex but approximate queuing model.

Based on the performed literature review, the following objectives for this work have been established:

- develop an overall methodology to quantitatively link the patients' LOS limits and percent ED diversion (both for admitted and discharged home patients).
- identify the maximum LOS limits that will result in significant reduction or elimination of ED diversion.
- estimate the number of patients in ED waiting room that should not be exceeded in order to keep percent ED diversion on a low single digits level.

Based on an assessment of the capabilities of the two possible approaches (process model simulation or queuing theory approximation), a process model simulation methodology has been chosen to attain the above goals.

Method

Overall simulation methodology

Methodology used in this work was based on principles of Operations Research. It was implemented by building a model of ED patient flow using commercially available simulation software package (Process Model, Inc, Utah, version 5.2.0).

A process model is a computer model that mimics the dynamic behavior of a real process as it evolves with time in order to visualize and quantitatively analyze its performance. Typical applications include: staff and production scheduling, capacity planning, productivity improvement, cycle time and cost reduction, throughput capability, resources and activities utilization, bottleneck finding and analysis. Process model is the most effective tool to perform quantitative 'what-if' analysis, and play different scenarios of the process behavior as its conditions and variables change with time. This simulation capability allows to make experiments on the computer display, and to test different solutions (scenarios) for their effectiveness before going to the hospital floor for the actual implementation.

The basic elements (building blocks) of a process model are:

- Flow chart of the process, i.e. a diagram that depicts logical flow of a process from its inception to its completion
- Entities, i.e. items to be processed, e.g. patients, documents, customers, etc.
- Activities, i.e. tasks performed on entities, e.g. medical procedures, exams, documents' approval, customer check in, etc
- Resources, i.e. agents used to perform activities and move entities, e.g. service personnel, operators, equipment, nurses, physicians.
- Entity routings that define directions and logical conditions flow for entities

Typical information usually required to populate the model includes:

- Quantity of entities and their arrival time, e.g. periodic, random, scheduled, daily pattern, etc. There is no restriction on the arrival distribution, such as Poisson distribution, required by the closed analytical formulas of the queuing theory
- The time that the entities spend in the activities, i.e. service time. This is usually not a fixed time but a

statistical distribution. There is no restriction for a special exponential service time distribution required by the closed analytical formulas of the queuing theory

– The capacity of each activity, i.e. the max number of entities that can be processed concurrently in the activity.
– The size of input and output queues for the activities
– Resource assignments: their quantity and availability, and/or working shift schedule

Description of the ED patient flow model

The layout of the ED system is presented on Fig. 1. The entire ED system included a fast-track lane called Minor care (capacity five beds), as well as Trauma room (capacity four beds) and, a so-called arena area with the most patient beds (capacity 21 beds). Total ED capacity was 30 beds.

Because the objective of this work was simulating an effect of patient LOS on diversion for the entire ED, the layout could be significantly simplified. It is presented on Fig. 2.

There are two modes of transportation by which patients arrive into ED indicated on Fig. 2: walk-in and ambulance. When ED patients' census hits ED beds capacity limit (total 30 beds), an ambulance was bounced back (diverted), as it is indicated on Fig. 2. Ambulance diversion continued until the time when the ED census dropped below the capacity limit. An action logic code was developed and programmed into the model that tracked the percentage of time when the census was at the capacity limit. It was reported as percent diversion in the simulation output file.

All simulation runs start at week 1, Monday, at 12 A.M. (midnight). Because ED was not empty at this time, Monday midnight patients' census was used as the simulation initial condition on January 1, 2007: ED was pre-filled by 15 patients.

Each patient in the arrival flow was characterized by its week number, day of week, and admitting time on the record, as indicated on the panel on Fig. 2. The following descriptive attributes (also indicated on the panel on Fig. 2) were assigned to each patient on the arrival schedule to properly track each patient's routing and statistics in the simulation action logic:

– Mode of transportation: (1) walk-in, (2) ambulance.
– Disposition: (1) admitted as inpatient, (2) discharged home

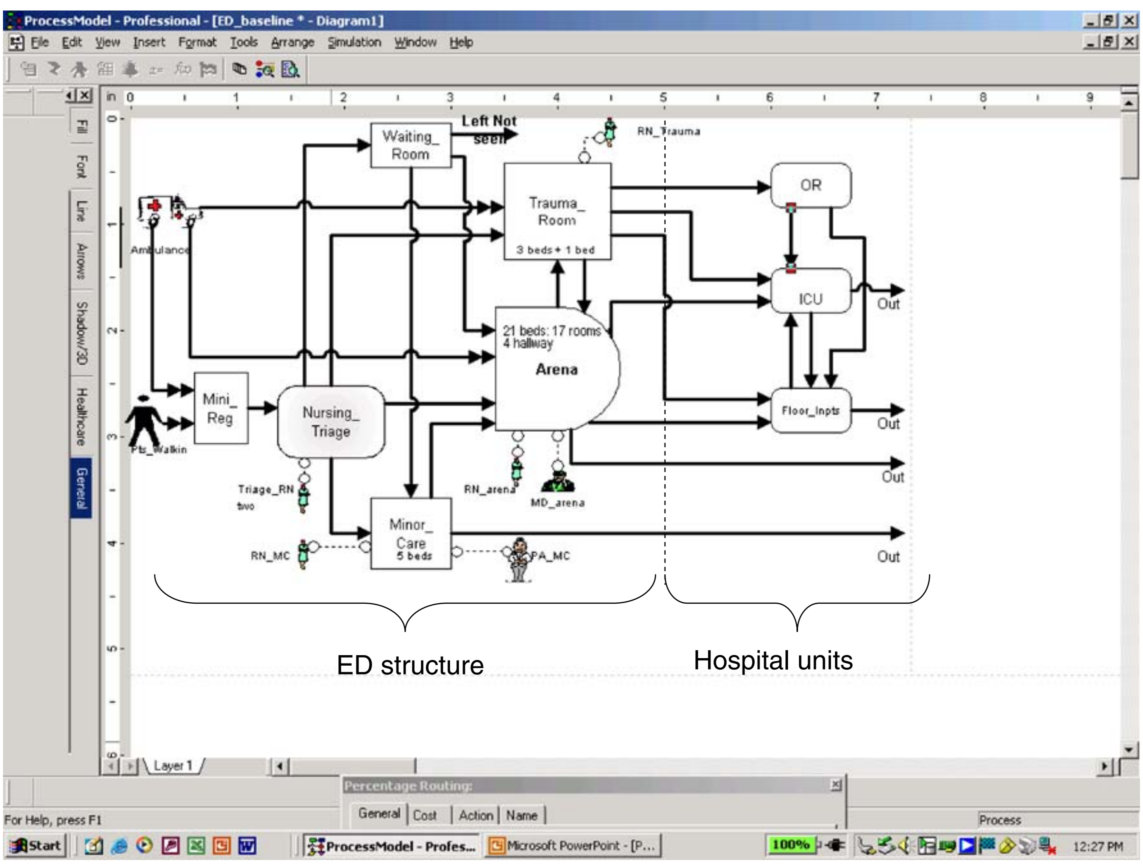


Fig. 1 General layout of ED structure and related hospital departments: *OR* Operating rooms, *ICU* intensive care units, *floor units* all other regular care hospital nursing units

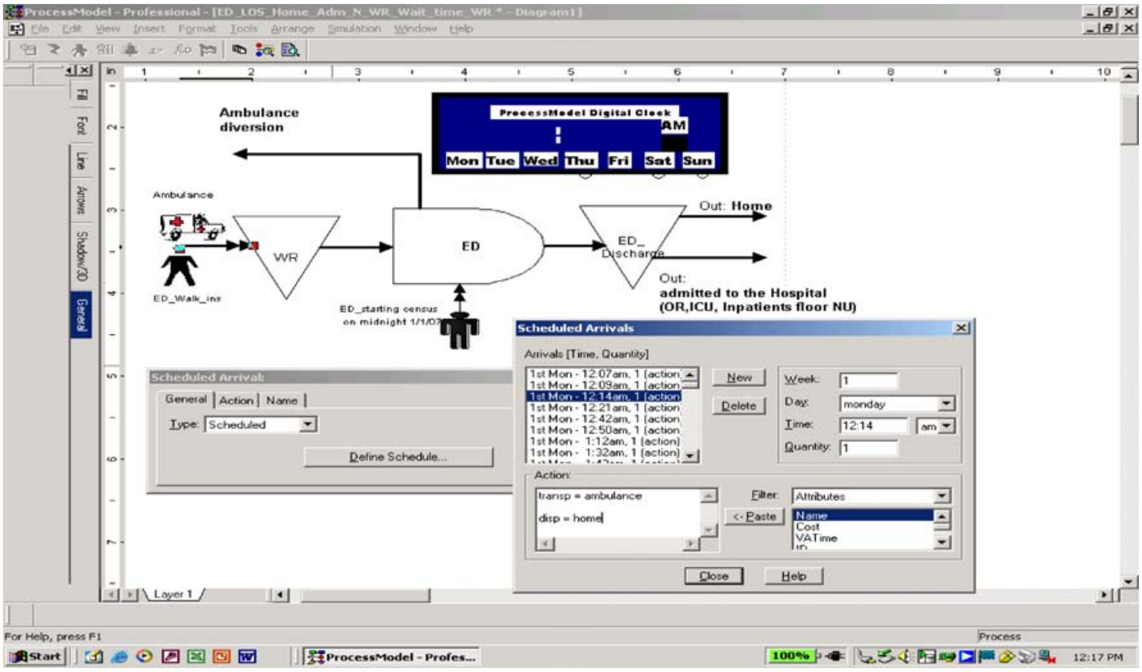


Fig. 2 Simplified ED structure used to simulate limiting length of stay for patients discharged home and patients admitted to the hospital

Arrived patients take available free beds reducing ED free capacity.

Discharged patients (released home or admitted as inpatients) moved out of the simulation system according to their disposition conditional routings. The patients' flow 'in and out' of the ED formed a non-steady-state (dynamic) supply and demand balance.

The number of patients included into ED simulation model is presented in Table 1. Total number of patients included in the simulation was 8,411 for the 2-month period from January 1 to February 28, 2007. This number of patients was representative enough to make results valid for subsequent months and years (in work [10] 3 months 2002 data-base was used to calibrate the queuing model; however the total number of patients was not given).

Overall simulation approach and LOS distribution density functions

The critical element of the dynamics of the supply and demand balance was the time that the patients spent in ED.

Table 1 Number of patients included in the simulation: discharged home and admitted as inpatients into the hospital. Total 8,411

	Admitted as inpatients	Discharged home
Jan-07	1,133	3,255
Feb-07	1,052	2,971
Subtotal	2,185	6,226

This time was fitted by a continuous LOS distribution density functions, separately for admitted as inpatients and discharged home patients (Fig. 3).

The best fit was performed using the Stat:Fit module built in the Process Model simulation package.

It was identified that the best fit distribution for admitted inpatients LOS was log-logistic, while the best fit distribution for LOS of patients discharged home was Pearson 6. These distributions were built into the simulation action logic.

The log-logistic and Pearson 6 distributions are bounded on the lower side; they commonly used to model the output of complex processes, such as business failure, product cycle time, etc [13]. Because these LOS distributions represent a combination of many different steps of the patient move through the entire ED process from registration to discharge (including both value-added and non-value-added steps and delays), there is no their simple interpretation: these are simply a best analytical fit used to represent actual patient LOS data.

Random numbers drawn from these distributions were used to perform multiple replications in each simulation run. It was identified in 'cold' runs that about 100 replications were needed for each simulation in order to get a stable outcome.

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, the LOS limits were used as two independent simulation parameters.

An overall simulation approach was based on a full factorial design of experiments (DOE) with two factors

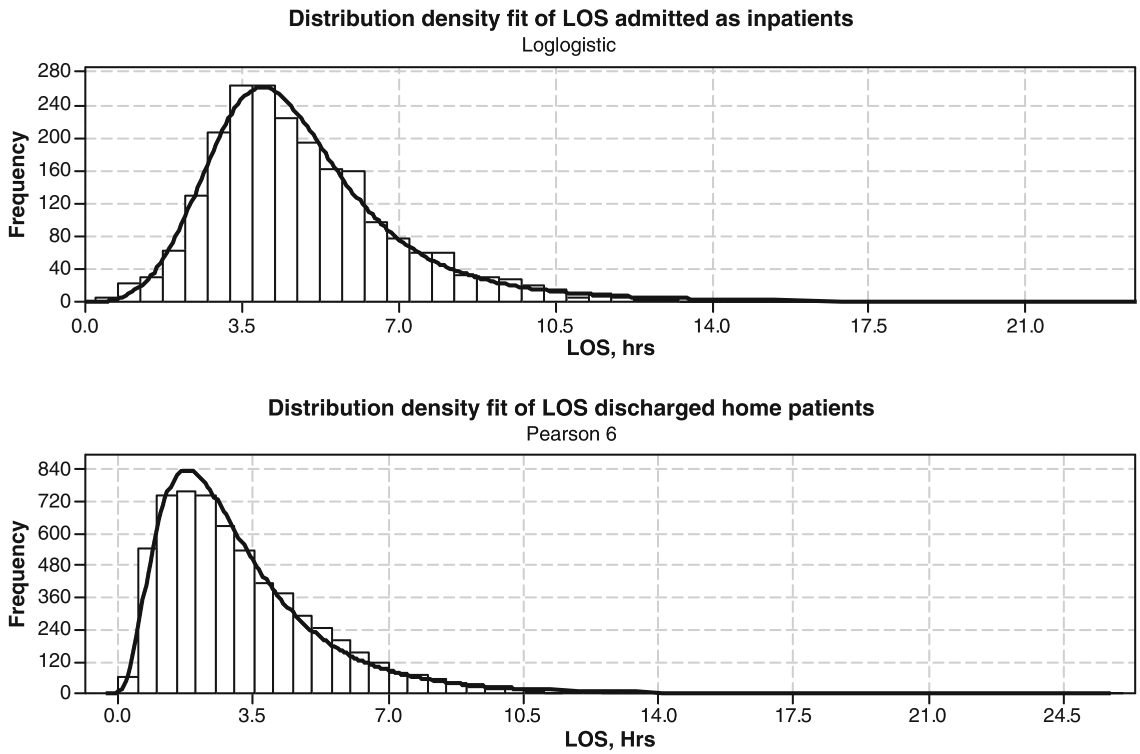


Fig. 3 Histogram of patients' length of stay and best fit distribution density functions. *Top panel* LOS for admitted as inpatients. *Bottom panel* LOS for discharged home

(parameters) at six levels each imposed on the original (baseline) LOS distribution functions. Response function was the simulated percent diversion. Imposing LOS limits (parameters) on original (baseline) LOS distribution functions means that no drawn random LOS value higher than the given limiting value was allowed in the simulation run. Therefore the original LOS distribution densities should have recalculated for each simulation run as functions of the LOS limits (parameters).

One might be tempted to assume that if a drawn random LOS number was higher than the given LOS limit value this number should be made equal to the LOS limit. However such an approach would result in a highly skewed simulation output because a lot of LOS numbers would be concentrated at the LOS limit value.

Instead, a concept of conditional distribution density function should be used. If a random LOS number was in the interval from 0 to $\mathrm{LOS_{lim}}$ this number was used to run a simulation replication. However if a random LOS number was outside the interval from 0 to $\mathrm{LOS_{lim}}$ this number was not used, and the next random number was generated until it was in the given interval. This procedure generated a new restricted random variable that is conditional to being in the interval from 0 to $\mathrm{LOS_{lim}}$.

Given the original LOS distribution density, $f(T)_{orig}$, and the limiting value, LOS_{limit}, the conditional LOS distribution density function of the new restricted random variable, $f(T)_{\mathrm{new}}$ will be (Fig. 4)

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

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

The conditional distribution density $f(T)_{\mathrm{new}}$ is a function of both original distribution density and the simulation parameter $\mathrm{LOS_{lim}}$ (upper integration limits of the denominator integrals).

These denominator integrals were preliminary calculated and then approximated by third order polynomials which were built in the simulation action logic:

For discharged home patients:if $\mathrm{LOS_{lim}} \leq 10$ h

$$\int_0^{LOS_{\mathrm{lim}}} f(T)_{orig} dT = -0.2909 + 0.4013 * LOS_{\mathrm{lim}} - 0.04326 * LOS_{\mathrm{lim}}^2 + 0.001599 * LOS_{\mathrm{lim}}^3$$

else the integral is approximately equal to 0.997.

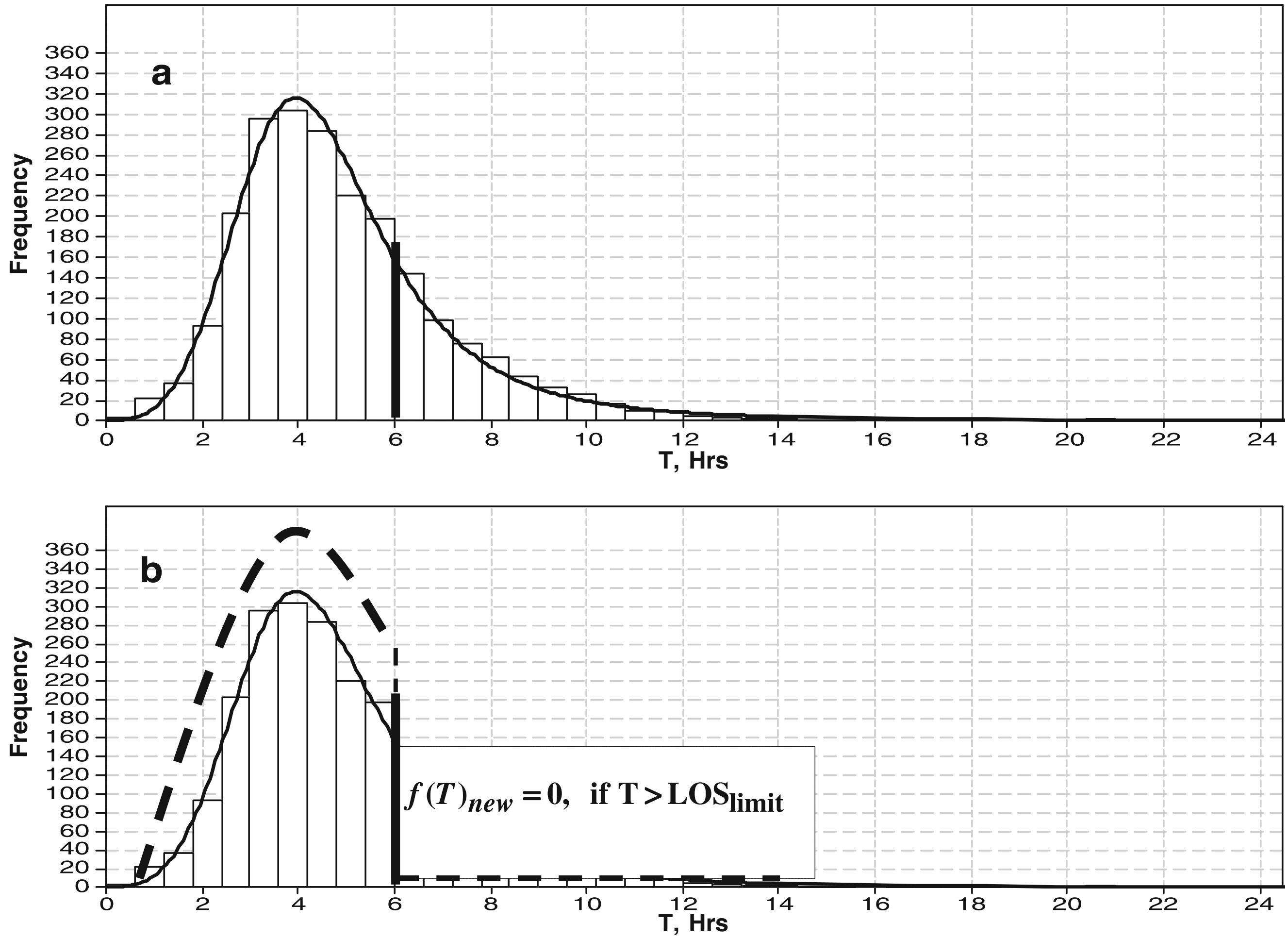


Fig. 4 LOS distribution density function of the new restricted random variable and imposed LOS limit: **a** Thin solid line-original LOS distribution (*top panel*). Bold vertical line-example of imposed LOS limit, 6 h; **b** re-calculated restricted LOS distribution: bold dotted line (*bottom panel*)

For patients admitted into hospital as inpatients: if $LOS_{lim} \leq 10$ h

$$\int_0^{LOS_{lim}} f(T)_{orig} dT = -0.7451 + 0.3738 \times LOS_{lim} - 0.02188 \times LOS_{lim}^2 + 0.000157 \times LOS_{lim}^3$$

else the integral is approximately equal to 0.994.

Results and discussion

Baseline simulation and an evaluation of the model adequacy

Before using the model to play 'what-if' scenarios the model should have evaluated for how adequately it mimics the real process behavior. An adequacy check was performed by running the simulation of the original baseline patients' arrival. The model's predicted percent diversion (~23.7%) was compared with the reported percent diversion (21.5%). The later was reported by the Emergency Department (ED) as the percent of time when the ED was closed to the Emergency Management Transportation (EMT).

It should be noted that the data base for arrived patients and their LOS used in simulation and the ED closure reporting data are not equivalent. Therefore some discrepancies between the simulated and reported percent diversion were expected. However EMT percent diversion and independently simulated percent diversion are close enough (in the range of a few percentage points). Thus, it could be concluded that the model captures dynamic characteristics of the ED patients' flow adequately enough to mimic the system's behavior, and to compare alternatives ('what-if' scenarios).

Along with the percent diversion calculation, a plot of ED census as a function of time (hours per week) was also simulated. This instructive plot is presented on Fig. 5. The plot visualizes the timing when the ED census hits the

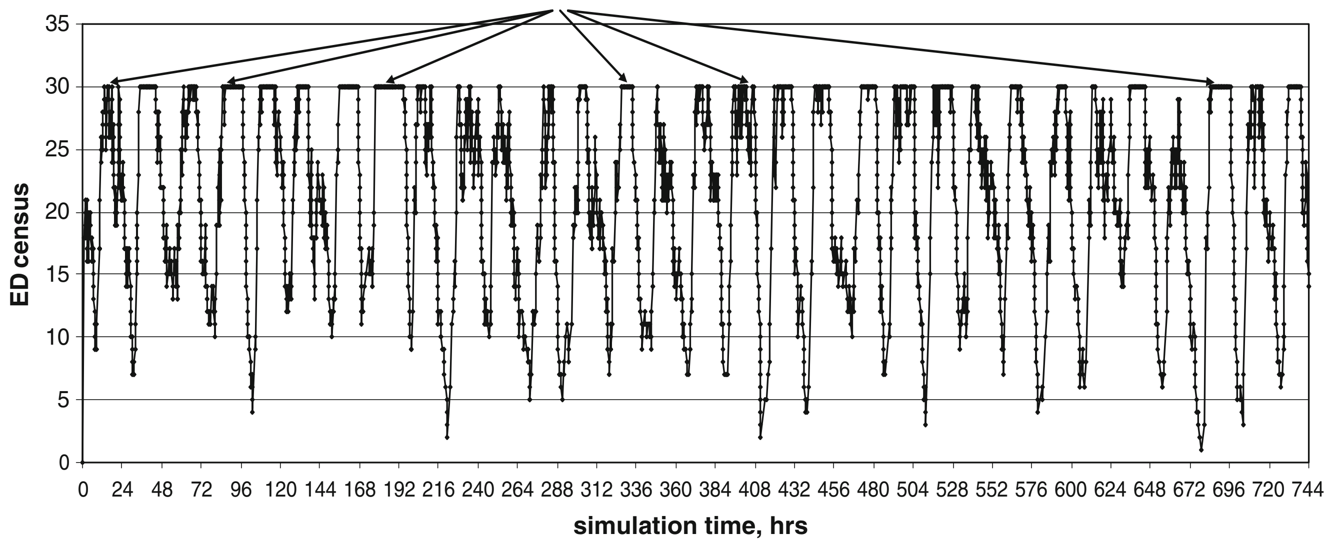


Fig. 5 Baseline simulated ED census. ED capacity is 30 beds. Diversion is 23.7%

capacity limit, and therefore ED diversion had to be declared. The plot also illustrates that at some periods of time (mostly late night time) the ED was actually underutilized having a low census. Therefore, once the model was checked for its adequacy, it was used with enough confidence to simulate 'what-if' scenarios.

Simulation scenarios: Phase 1

In Phase 1 of the simulation, a full factorial computer design of experiments (DOE) was performed with two factors: LOS_{lim} (home) for discharged home patients and LOS_{lim} (adm) for patients admitted into hospital. Each factor had six levels. Simulated percent diversion was a response function. The factors levels and simulated percent diversion are given in random order in Table 2.

A three-dimensional surface plot of percent diversion as a function of two parameters LOS_{lim} (home) and LOS_{lim} (adm) is presented in Fig. 6. The response surface is highly curved (non-planar). Therefore, it is highly inaccurate to make a simple linear projection for percent diversion for different values LOS_{lim} (home) and LOS_{lim} (adm). This plot indicates that a low single digit diversion corresponds to LOS_{lim} values in the range of 5 h to 7 h.

In order to get a more detailed picture, a cross-sectional plot was generated (Fig. 7). It follows from this plot that several combinations of parameters LOS_{lim} (home) and LOS_{lim} (adm) would result in low single digit percent diversion. The best combination LOS_{lim} (home) of 5 h and LOS_{lim} (adm) of 5 h resulted in almost negligible diversion ~0.13%.

However, if a little higher diversion is acceptable, these LOS limits could be relaxed (Fig. 7). For example, if LOS_{lim} (home) stays at 5 h (low curve) then LOS_{lim} (adm) could be extended up to 10 h with the diversion still about 3%. A similar diversion level of less than 3% could be achieved if LOS_{lim} (home) level was increased to 6 h (second low curve) while LOS_{lim} (adm) was also kept at 6 h level. Any other combination of LOS_{lim} (home) and LOS_{lim} (adm) could be used in order 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 the reasonable targets for the upper limits LOS_{lim} (home) and LOS_{lim} (adm).

Analysis of the current LOS pattern indicated that a significant percentage of ED patients stayed much longer than the LOS targets suggested by the simulation. For example, ~24% patients exceeded LOS_{lim} (adm) of 6 h, and ~17% of patients exceeded LOS_{lim} (home) of 5 h. These long over-targets LOS for a significant percentage of patients were a root cause of ED closure and ambulance diversion. (Compare these data to the UK government Department of Health benchmarks for A&ED: not more than 2% of patients are allowed to exceed the LOS limit of only 4 h [4, 10].

The established LOS_{lim} targets could be used to better manage a daily patient flow. The actual current LOS is being tracked down for each individual patient. If the current LOS for the particular patient at the moment is close to the target limiting 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 downstream on hospital floor (ICU) for admitted patients, etc. Analysis and prioritizing of the contributing factors to the over the target LOS_{lim} is being conducted. Results will be presented elsewhere.

Table 2 Results of full-factorial computer DOE: Two factors: LOS_{lim} (home) and LOS_{lim} (adm)

LOS_{lim} (home) (h)	LOS_{lim} (adm) (h)	ED Diversion (%)
6	12	8.28
12	5	4.61
24	5	4.93
6	8	5.58
24	8	17.17
10	6	7.72
8	8	11.82
6	24	8.91
12	8	16.48
12	12	21.52
6	6	1.81
24	12	22.82
8	12	15.95
24	24	23.77
12	10	19.38
10	10	18.09
5	6	0.42
10	12	20.32
5	10	2.78
24	6	8.65
10	5	4.19
8	6	5.53
5	8	1.9
8	5	2.62
8	10	14.31
6	10	6.92
10	24	21.62
24	10	20.42
5	24	3.8
12	24	23.21
6	5	0.6
5	5	0.13
10	8	15.32
8	24	16.73
12	6	8.59
5	12	3.5

Each factor has six levels. Response function is simulated percent diversion.

Notice that an average LOS that is frequently reported as one of the ED flow performance metrics could not be used as a useful metrics to manage a daily flow. In order to calculate an average LOS, data should be collected retrospectively for at least a few dozens 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_{lim} 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}.

The shortcomings of reporting LOS only as averages (flaw of averages) for the skewed (long tailed data) were also discussed in works [10, 14] and [15].

Simulation scenarios: Phase 2

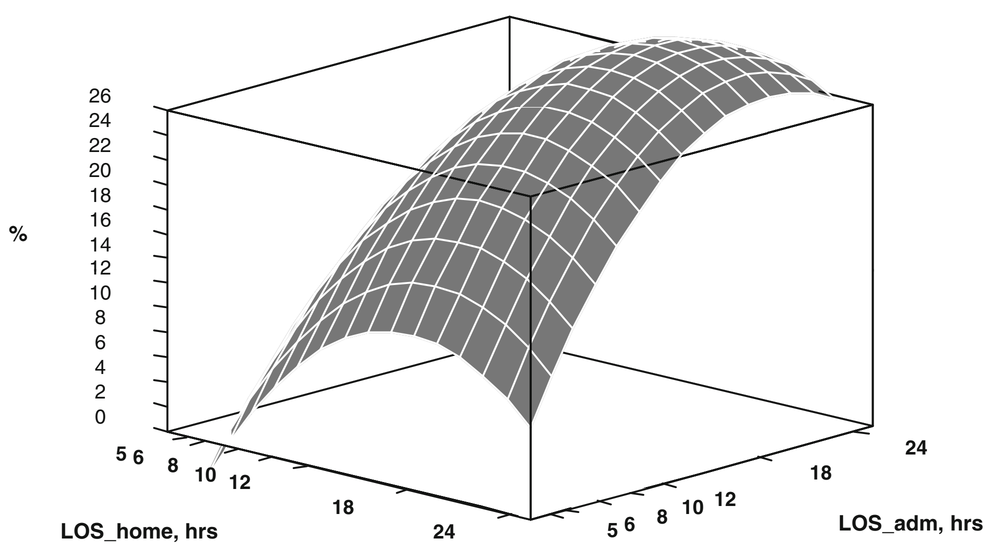
The ED closure criterion due to 'no ED beds' is sometimes considered 'too late.' Therefore, an alternative closure criterion would be useful to indicate that there would be the need to close soon. Such a convenient alternative closure criterion for ED could be the number of patients that are already in the ED waiting room, i.e. the size of patients' queue.

A question for a simulation scenario was: how many patients should be in the ED waiting room in order to have a corresponding low projected percent of diversion?

A plot of the baseline simulation number of waiting patients is presented on Fig. 8. The plot shows the largest peak of 34 patients in the queue on Monday of the second week. (First week ends at 168 h, Monday of the second week ends at 192 h).

The corresponding plot of simulated waiting time for those admitted into the hospital as inpatients is presented in Fig. 9. It is apparent that there is a close match between the timing location of the peaks of the number of patients in the

Fig. 6 3-dimensional surface plot representing simulated percent diversion as a function of two parameters LOS_{lim} (home) and LOS_{lim} (adm)



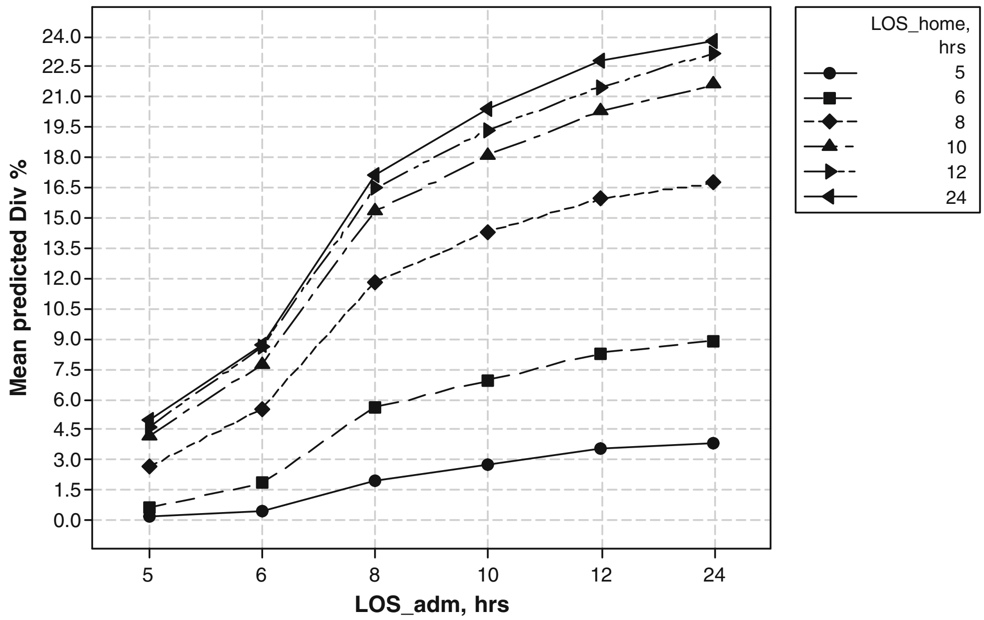


Fig. 7 Cross-sectional plot representing simulated percent diversion as a function of two parameters LOS_{lim} (home) and LOS_{lim} (adm)

queue and the peaks of their waiting time. For example, the timing of the peak of the number of waiting patients (34 patients) matches exactly the timing of the peak for the longest waiting time (about 3.5 h).

Similar plots have been generated, for example, for LOS_{lim} (home) of 5 h and LOS_{lim} (adm) of 6 h. These plots are presented in Figs. 10 and 11. It is clearly seen a dramatic difference in the number of waiting patients and their waiting times compared to the baseline simulation: the peak of seven patients in ED waiting room happens to be once (vs. peak of 34 patients for baseline), and the corresponding waiting time peak does not exceed about 0.6 h (~35 to 40 min) vs. 3.5 h peak for baseline.

An overall plot that represents the diversion percent and the corresponding number of patients in the queue in ED waiting room is given on Fig. 12. It follows from this plot that the low single digits percent diversion, for example, about 3% would correspond to 11 patients in ED waiting room. Any other number of waiting patients and corresponding percent diversion can be found from this plot.

Conclusions

This work has demonstrated the value and the power of predictive simulation methodology. By simulating different

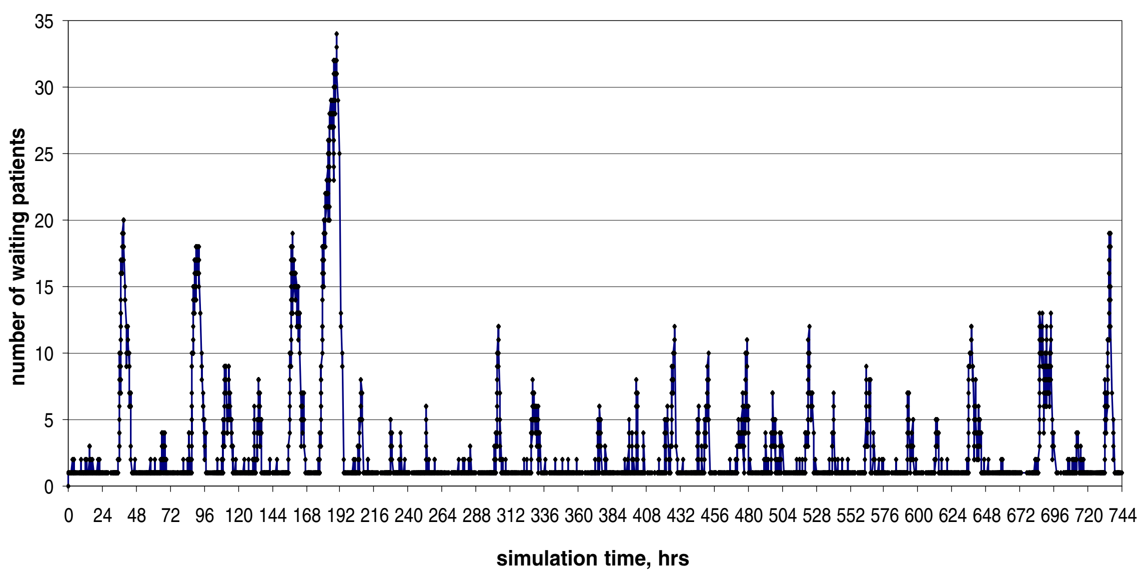


Fig. 8 Simulated number of patients in the queue in ED waiting room: baseline. Simulation time is plotted in 1 day increments (every 24 h). Each week corresponds to 168 h

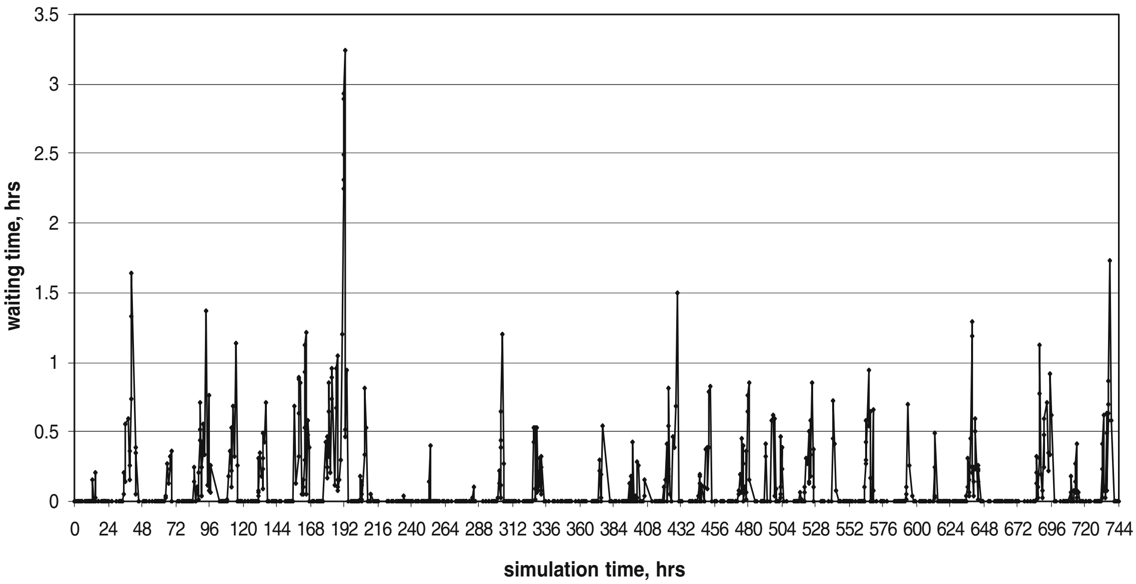


Fig. 9 Baseline simulated waiting time for admitted inpatients. Simulation time is plotted in 1 day increments (every 24 hrs). Each week corresponds to 168 hrs

scenarios of ED patient flow it has been identified that ED diversion could likely be negligible (less than ~0.5%) if patients discharged home stay in ED not more than 5 h, and patients admitted into the hospital stay not more than 6 h.

Currently ~17% of patients discharged home stayed above this limit, up to 24 h; ~24% of admitted patients stayed above this limit, up to 24 h.

This long LOS for large percent of patients results in ED closure/diversion.

Using full factorial design of experiments (DOE) with two factors and the simulated percent diversion as a response function, a quantitative link has been established between an expected percent diversion and the limiting values of LOS. Any other combination of LOS_{lim} (home) and LOS_{lim} (adm) could be used in order to estimate a corresponding expected percent diversion.

It has also been determined that the number of patients 11 or less in queue in ED waiting room corresponds to diversion less than ~3%. This could be used as an alternative ED diversion closure criterion.

Results obtained in this work might look as strictly related to the data input and to the particular LOS distribution functions, and therefore could not be generalized to other ED.

It should be noted that the main point of this work was to develop an overall methodology of the using discreet event simulation in order to establish a quantitative relationship between ED performance characteristics (percent diversion) and the target upper limits of patient length of stay.

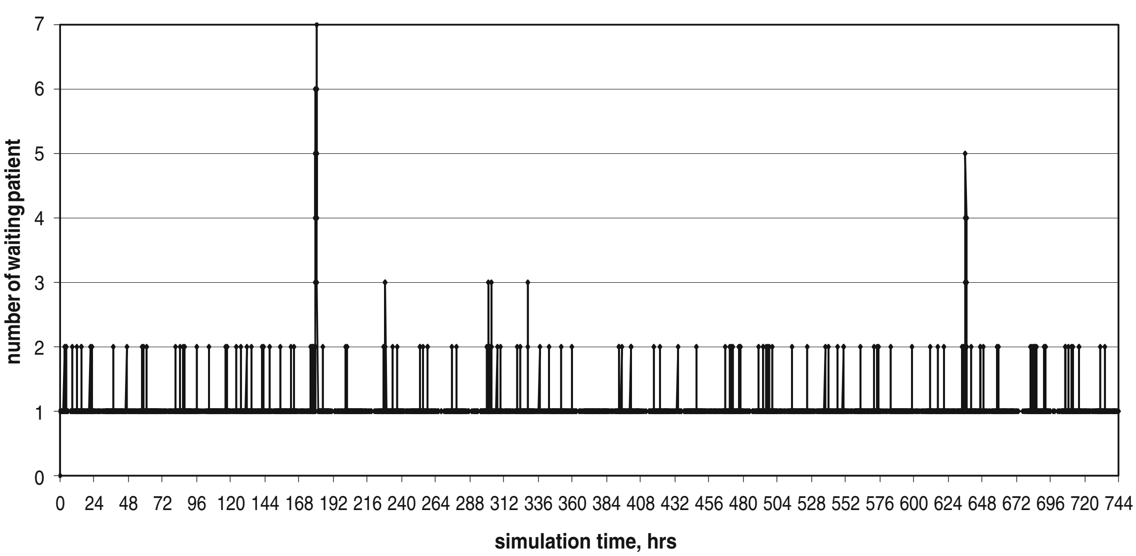


Fig. 10 Simulated number of patients in the queue in the ED waiting room. $Loss_{lim}$ (home) is 5 h and Los_{lim} (adm) is 6 h

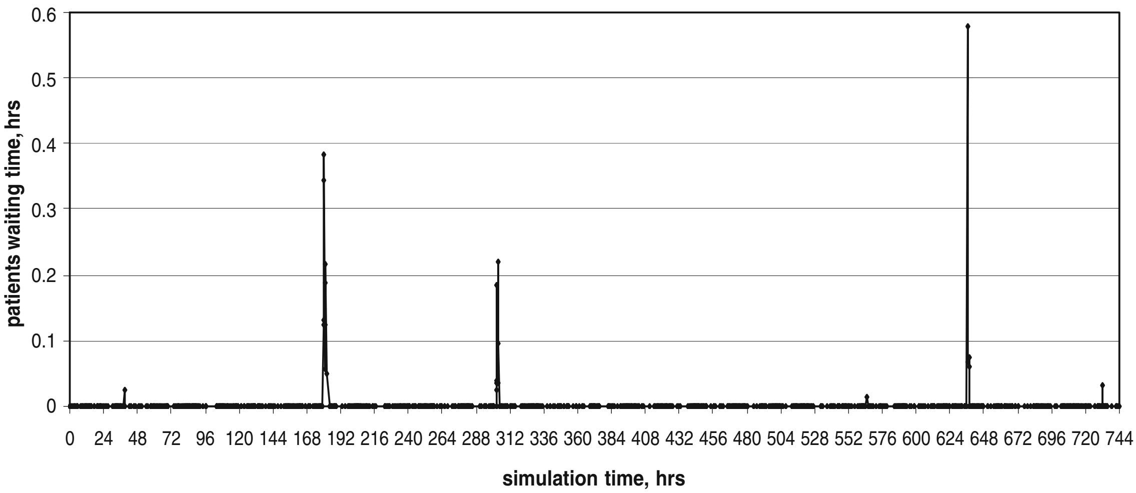


Fig. 11 Simulated waiting time in the queue in ED waiting room (for those admitted subsequently into hospital as inpatients): LOS_{lim} (home) is 5 h; LOS_{lim} (adm) is 6 h

An application of this methodology was illustrated using the ED data from one specific institution that was a large 450+ beds primary teaching hospital with level 1 trauma center and diverse patient population.

ED of other hospitals differ by their patient mix, their LOS and bed capacity. However the overall simulation methodology presented in this paper will be the same regardless of a particular hospital ED.

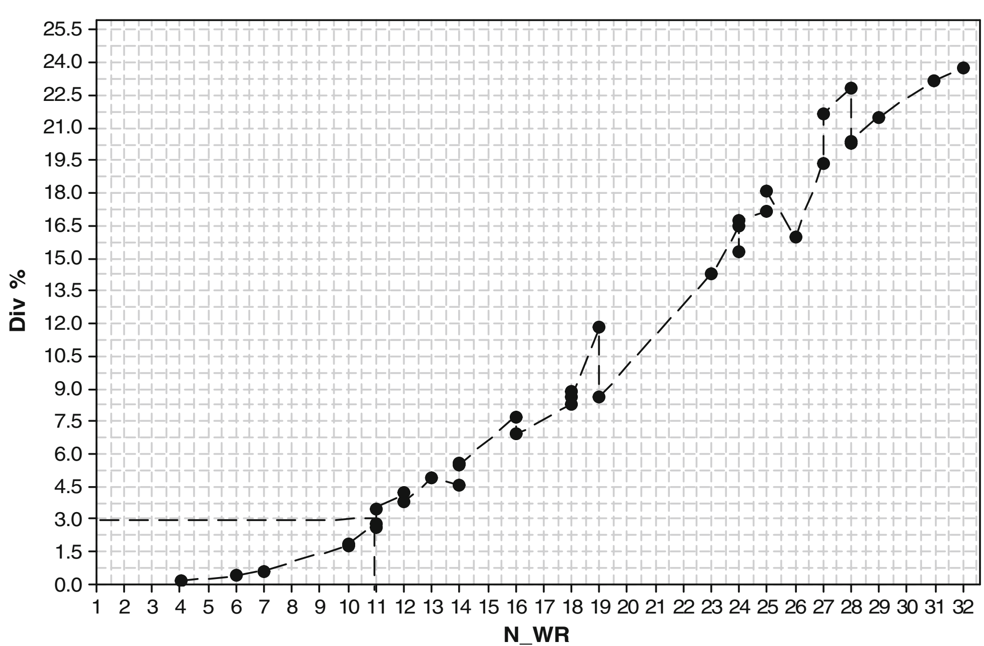
In order to find out the target LOS limits for any other ED their actual LOS data could easily be fit by a continuous distribution using readily available statistical software packages, such as Minitab, SPSS, StatFit, and many others. Such an operation is a few clicks away, and it takes a few minutes to complete. These distributions could then be plugged into the simulation module along with the actual ED bed capacity, and the simulation should be run to get an expected percent of time when ED would be on diversion.

Such a general methodology would practically be more useful for other ED than some pre-determined generalized 'one size fits all' target values.

The negative consequences of the 'one size fits all' approach were vividly summarized in work [10]: '…the practicality of a single target fitting all A&ED will come under increasing strain'.

A process model simulation methodology could also be used to analyze other 'what-if' scenarios. For example, the staffing problem: what resources (the number of doctors, nurses, technicians) would be needed to achieve and

Fig. 12 DOE summary: simulated percent diversion and corresponding number of patients in ED waiting room, *N_WR*. In order to get a low single digit percent diversion (about 3%) the number of patients in ED waiting room should not exceed 11



maintain the established LOS targets? What should be their shift allocation during a day of the week, and/or for different days of the week? How to best match staff schedule and short term fluctuations of the patient flow? Results of these advanced applications will be presented elsewhere.

Analysis of patient flow is an example of the general dynamic supply and demand problem. There are three basic components that should be accounted for in such problems: (1) the number of patients (or, generally, any items) entering the system at any point of time, (2) the number of patients (or any items) leaving the system after spending some time in it, (3) capacity of the system which limits the flow of items through the system. All three components affect the flow of patients (items) that the system can handle. A lack of the proper balance between these components results in the system's over-flow and its closure. Process model simulation methodology provides invaluable means for analyzing and managing the proper balance.

References

1. Litvak, E., Long, M., Cooper, A., and McManus, M., Emergency department diversion: Causes and solutions. *Acad. Emerg. Med.* 8:111108–1110, 2001.
2. Haraden, C., Nolan, T., and Litvak, E., *Optimizing Patient Flow: Moving Patients Smoothly Through Acute Care Setting*. Institute for Healthcare Improvement Innovation Series 2003. White papers 2, Cambridge, MA, 2003.
3. Litvak, E., and Long, M., Cost and quality under managed care: Irreconcilable differences? *Am. J. Manage. Care*. 6:3305–312, 2000.
4. Gunal, M., and Pidd, M., *Understanding Accident and Emergency Department Performance using Simulation*. Proceedings of the 2006 Winter Simulation Conference. Ed. Perrone, L, et al. 2006 IEEE, p. 446–452.
5. Garcia, M., Centeno, M., Rivera, C., and DeCario, N., *Reducing Time in an Emergency Room via a Fast-track.* Proceedings of the 1995 Winter Simulation Conference. Ed.: Alexopoulos C., et al. 1995 IEEE, p.1048–1053.
6. Miller, M., Ferrin, D., and Szymanski, J., *Simulating Six Sigma Improvement Ideas for a Hospital Emergency Department*. Proceedings of the 2003 Winter Simulation Conference. Ed. Chick, S, et al. 2003 IEEE, p. 1926–1929.
7. Simon, S., and Armel, W., *The Use of Simulation to Reduce the Length of Stay in an Emergency Department.* Proceedings of the 2003 Winter Simulation Conference. Ed. Chick, S, et al. 2003 IEEE, p. 1907–1911.
8. Blasak, R., Armel, W., Starks, D., and Hayduk, M., *The Use of Simulation to Evaluate Hospital Operations between the ED and Medical Telemetry Unit*. Proceedings of the 2003 Winter Simulation Conference. Ed. Chick, S, et al. 2003 IEEE, p. 1887–1893.
9. Canadian Association of Emergency Physicians. Position Statement on Emergency Department Overcrowding, February 2007. www.CAEP.ca.
10. Mayhew, L., and Smith, D., Using queuing theory to analyze the Government's 4-h completion time target in accident and emergency departments. Health Care Manage. Sci. 11:11–21, 2008.
11. Hall, R., *Queuing Methods for Service and Manufacturing*. Prentice Hall, Englewood Cliffs, NJ, USA, 1990.
12. Swain, J., *Biennial Survey of discreet-event simulation software tools*. OR/MS Today, v.34, N 5, October, 2007. The Institute for Operations Research and the Management Science. Lionheart Publishing, Inc. http://www.lionhrtpub.com.
13. Johnson, N., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distributions*, vol 2. John Wiley & Sons, New York, 1995.
14. Marshall, A., Vasilakis, C., and El-Darzi, E., Length of stay-based patient flow models: recent developments and future directions. *Health Care Manage. Sci*. 8:213–220, 2005.
15. de Bruin, A., van Rossum, A., Visser, M., and Koole, G., Modeling the emergency cardiac in-patient flow: an application of queuing theory. *Health Care Manage. Sci*. 10:125–137, 2007.