A Method to Improve the Accessibility and Quality of a Brazilian Public Emergency Hospital Service

Francisco Sabbadini and Mário De Oliveira
Universidade Federal do Rio de Janeiro
Brazil

Antonio Gonçalves
Universidade Estácio de Sá
Brazil

Antonio Araujo Jr. and Jose Glenio Barros
Universidade do Estado do Rio de Janeiro
Brazil

Rosinei Batista
University of State of Rio de Janeiro
Brazil

Abstract

The accessibility to public hospitals in Brazil is a serious problem and there is much to do in order to improve the quality of care. The insufficient structure of the National Health Service imposes budget restrictions, a lack of control over the human and material resources and a high demand for an emergency which is in constant growth. The private care system is expensive and only a low percentage of the population can afford to pay for health care. The consequence is that most emergency hospitals are constantly overcrowded. Patients have to wait to start treatment and the delays can lead to increased risk of sequels or even represent the difference between life and death. Long waiting times causes discomfort for patients and put pressure on the medical staff. The reduction of waiting times for treatment is considered as an important measure to control costs and improve the effectiveness of the service. The objective of this paper is to show a method that aims to improve the access and the quality of the emergency admission system of a public municipal hospital in the State of Rio de Janeiro. The goal is to increase the number of real emergency patients successfully treated and to reduce the waiting time between the registration of the case and the beginning of the emergency care. The method is based on the authors’ experience in health services operational research, and combines the theory of constraints and discrete event simulation. It is expected that the method will contribute to administrative knowledge and the results can be implemented in other national public hospitals.

Keywords
Discrete event simulation, theory of constraints, emergency, services.

1 Introduction
The accessibility to public health services in Brazil is a critical issue and much remains to be done to improve the quality of the services. The insufficient structure of the public emergency services has contributed to its overload, resulting in one of the most problematic areas of the healthcare system. Most of the public hospitals are not able to cope with the increase of demand. Improvement of the quality of emergency services is considered as an important measure to control the social costs and increase the effectiveness of the medical care. The emergency hospital is a strategic unit where patients have to be promptly treated. Delays can lead to sequels or even represent the difference between life and death. The hospital is a complex structure that involves several strongly connected components. The economic component imposes budget restrictions and control over the human and material resources. The private system is usually connected with health insurance, but the majority of people cannot afford to pay for such treatment. The demographic growth and
The lack of affordable healthcare causes an impact on the demand for public healthcare services. It is very difficult to find an appropriate balance between the supply and demand in order to schedule services and the necessary resources. The emergency service process requires a method that effectively identifies the patients’ needs and offers the means to ease the access to the necessary resources.

There are a remarkable number of publications in the field of emergency management over the last two decades (Chaikken and Larson, 1998; Lagergren, 1998). Few studies take into consideration the value of the short-term prediction of the emergency demand and it is vital to improve the admission systems. Emergency arrivals are normally regarded as an unknown process (Lane, Monefield and Rosenhead, 2000). An extensive study of the emergency services in a major public hospital evaluates the flow of emergency patients in two phases (De Oliveira and Toscano, 2003): The pre-hospital phase which starts immediately after the emergency call, following the occurrence of an incident and finishes when the patient arrives at the hospital. The hospital phase comprises of the reception, the triage and medical emergency attention. It is argued that lack of information causes problems in accessing the hospital services in both phases. There has never been a more challenging time to discuss issues of services improvement using information, knowledge and technology. Accessibility and quality of health services was the subject of the conference. An earlier paper presented at this conference contributed for the reduction of the patient’s waiting time to start cancer treatment (Gonçalves and De Oliveira, 2004) and to develop an integrated information support system to help decision-making at the National Cancer Institute (Gonçalves, 2004).

The Theory of Constraints (TOC) has been designed to identify bottlenecks within commercial industry (Goldratt, 1990). There are a few implementations of this theory in the healthcare services (Kershaw, 2002; Rothstein, 2002). The author’s interest in discrete event simulation (DES) modelling started as early as 1980. It was consolidated through an interchange program with the computer assisted simulation modelling working group (Paul and Balmer, 1986). The relatively low cost of DES, the little risk of experiments with a model of the real system and the ability to keep up statistical fluctuations are some of the features that advocate the use of this technique. There have been former attempts to combine the principles of TOC and DES. This in turn generated a methodology called TOCANDES (Gonçalves, 2004; Sabbadini, 2005). This paper aims to develop the access and quality of a public emergency service within the State of Rio de Janeiro. The objective is to reduce the waiting times to start the emergency treatment and the increase the number of emergency cases effectively treated. The steps of the method are: Evaluate the flow of emergency patients and the operational routines, find interdependent processes and bottlenecks; a discrete-event simulation model is finally designed to assess alternative scenarios giving support to changes within the hospital admission system.

2 The Hospital

The Henrique Sergio Gregory Hospital (HSGH) is a 250-bed public emergency health unit. The services provided are: emergency, laboratory, outpatient care, general medicine and surgery. Due to its geographic location, its structure and the provision of emergency services, the hospital has become a reference in this type of service for the city of Resende and the neighbourhood. The emergency unit of the hospital is open 24 hours with an average daily demand of 192 patients. It is noticeable that the closure of some emergency facilities in the neighbourhood has caused an increase of the demand. The HSGH has received 83,683 patients which represents an increase of 30% in relation to the last year. The available resources are consolidated in Table 1. The emergency sector deals with the more complex cases and the other cases are absorbed by the outpatient clinics. Besides a nurse on 24-hour duty, there are two shifts between 8 am to 5 pm covered. The nurses check the evolution of the clinical cases and also help other cases if necessary. They are usually assisted by nursing technicians.

The emergency sector is constantly overcrowded. There are long waiting lines that cause discomfort to patients and put pressure on the medical team. Further to the increase of demand, the hospital management has observed changes in the population profile. It is noticed that the majority of the patients attending the emergency unit are not real emergencies. The distribution is: 90% are outpatients and 10% of emergency cases. The hospital data includes the access, the type of care, the resources required and the history of the treatment inside the hospital. The data analysis covers all patients. From this information it is possible to trace the flow of patients, which is made up from the triage to the treatment phases. All the work is carried out in loco with the supervision of the head of the medical emergency. The research results are soon after validated by other members of the clinical staff, involving both doctors and nurses. The goal of the emergency staff is to be able to speed up the evaluation, stabilize the health state and carry out the hospital admission, upon a medical decision. Technically, the emergency sector within a hospital can be described as a bottleneck (Rothstein, 2002). It is
argued that understanding the flow of the treatment, evaluating the constraints and managing the bottlenecks could provide possibilities to improve the quality of the service.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptionist</td>
<td>2</td>
</tr>
<tr>
<td>Nurse</td>
<td>1</td>
</tr>
<tr>
<td>Nursing Technician</td>
<td>5</td>
</tr>
<tr>
<td>Outpatient Care Physician</td>
<td>4</td>
</tr>
<tr>
<td>Emergency Care Physician</td>
<td>3</td>
</tr>
<tr>
<td>Outpatient Care Clinics</td>
<td>4</td>
</tr>
<tr>
<td>Medication and nursing room</td>
<td>1</td>
</tr>
<tr>
<td>Emergency room</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1 - Resources of emergency area.

### 3 The Constraints Evaluation

A constraint is anything that limits the system’s performance. The TOC concepts have been originally designed to investigate productivity improvements in the manufacturing area (Goldratt and Cox, 1992). The first premise is the existence of a managerial constraint. The argument is that the administrative scheme is a process or a series of processes, in which inputs are turned into desired outputs, like a chain. The capacity of the system is limited by its constraints. A comprehensive review of the visibly benefits resulting from implementation of the TOC philosophy and practice can be seen elsewhere (Watson et al, 2006). A survey of viewpoints and applications is carried out to identify journal articles and conference proceedings between 1980 and 1995 (Rahman, 1998). It is noticed that most of the applications are concentrated on accounting, production, purchase, quality, administration and education. The author suggests further investigation in the services sector. The applications are usually focused on operational performance together with reduction of order-to-delivery lead time and manufacturing cycle time or increase in throughput/revenue and performance to uphold gains for several organizations (Mabin and Balderstone, 2000). An interesting study demonstrates the benefits of TOC and its impact on the NHS waiting lists and patient throughput (Lubitsh, Dolly and Valentine, 2005). The theory holds that the only way to create ongoing improvements is to release the weakest links of the chain. These can be summarized in the five steps below (Goldratt, 1990; Knight, 2000; Burton and Aron, 2002):

1. Identify the system’s constraint(s);
2. Decide how to exploit the system’s constraint(s);
3. subordinate/synchronize everything else to the above decisions;
4. Elevate the system’s constraint(s);
5. If in the above steps the constraint has shifted, go back to step 1

This theory views the organization as a chain of interdependent processes. The performance of each process is dependent upon the previous one. The capacity of the system is determined by the capacity of the parts or -more precisely, by the capacity of the bottleneck. The expansion of the process is limited by the bottlenecks. Improvements can be obtained by simply managing and synchronizing the activities that need to be accomplished before the bottleneck – the scarcest resource – to make sure that there is no ‘starvation’ of the bottleneck (Knight, 2000). Figure 1, shows an example the identification of a constraint in sequence of processes related to the hospital admission system.

The example shows a sequence of processes that takes place in the emergency patients flow. These processes are interdependent; hence the performance of one depends on the result of the previous. It can be seen that the consultation process is a constraint because of its lower capacity. Therefore, if the service capacity is limited, it is necessary to synchronize and elevate the constraints to be able to improve the flow. Figure 2 shows details of the patient flow.

The patient comes from the outside world and arrives into the reception area, and following this they go trough the triage where a nurse evaluates the health risk. A decision should be taken at this stage: If that is the case of a real emergency, the patient is promptly directed to the appropriate sector. Otherwise, a patient record is opened at the reception. After reception, the patient waits to see a doctor in the consulting room. It is necessary to manage the constraints to improve the performance. The basic flow of patients is composed by the triage, consultation, diagnosis and treatment. These pro-
cesses are interdependent; hence the performance of a particular process depends on previous one. It is argued that patients run through a chain of processes.

![Figure 1 - Constraint identification](image1)

Figure 1 - Constraint identification

![Figure 2 - Patient Flow - HSGH](image2)

Figure 2 – Patient Flow - HSGH

Figure 3 shows that consultation is a constraint. The analysis of the flow reveals an unbalance between demand and supply. This causes the formation of queues that contribute to increase waiting time to start treatment. It is noticeable that managers have evolved their capabilities in medical procedures, by using new concepts and tools. Now, they are able to evaluate to a greater level of important details of the patients’ flow within the hospital, associated to the clinical and administrative information. Some pieces of information on the follow-up of the emergencies are essential to ensure that the patient receives an efficient and clinically correct referral by the physician in charge of the treatment. This system also guarantees that the research protocols for complex cases are addressed. Finally, the use of the electronic records and flow evaluation helps to verify whether the appointments and exam schedules follow the appropriate sequencing.

The risk classification is sometimes carried out by the receptionist when the nurse is not available. The lack of a proper evaluation causes problems as the decision is sporadically made by an unqualified person. This setback was overlooked by the hospital administration and only doctors were able to identify the problem. There are other restrictions caused by
the number of consulting rooms, the right configuration of the medical staff and the excessive number of patients that are not real emergencies. There are unwanted situations where patients only require laboratory exams. Because of the overload, emergency doctors are constantly called to the outpatient sector. This situation cause unnecessary overcrowding that interferes with the performance of the system.

4 Methodology
The method requires the evaluation of the flow of emergency patients, the study of the operational routines, the identification of interdependent processes, the detection of bottlenecks and the elaboration of a simulation model that represents the main features of the emergency admission system. The practical application requires:
1. Set the objectives of the study;
2. Formulate the Problem;
3. Build the model;
4. Estimate parameters;
5. Run experiments with the model;
6. Validate the model;
7. Implement the results.

The simulation experiment aims at evaluate scenarios identifying bottlenecks and to weigh up effects of changes on the present configuration of resources and patient flow on systems performance. The model covers both the outpatient and the emergency sectors.

4.1 Objectives of the study
The setting of the objectives was fully discussed with the hospital management. This interaction was important to define the course of actions. It is established that the main line is to evaluate the attendance capacity and the provision of the resources. The capacity planning is essential to assign and utilize human and material resources. The hospital management requires a trustworthy instrument to understand, evaluate scenarios and implement suitable changes. The flow of patients and the reception process should be studied in order to define, among other things, the configuration and number of staff required to improve the system Therefore, it is necessary to examine both the supply and the demand. The objective of the model is to improve the quality of the services. The quality measures defined by the hospital administration are the
increase of the number of emergency patients served and the reduction of the time gap between registration and the effective beginning of the treatment.

4.2 Formulation

The formulation of the problem is the first and the most important step of the simulation experiment. To start the formulation one shall define the elements of the model. The most important elements are the entities, the activities and the queues. The entities are the system's components, that is, people or objects that are able to change the state of the system in time. Examples of entities of different classes are the patient, the staff (receptionist, nurse and doctor), beds, and other physical components (mechanical, logical). The entities have an attribute, which identify its state. The physical components have variables, which identify its operational state. The activities are actions which represents functions and services executed within the system where two or more entities are used simultaneously for a certain period of time. The activity is an active state, which involves the cooperation of entities from different classes. The duration of one activity could always be determined when it starts. It is independent of the future status of other entities, which are not participating in that particular activity. For example the activity "reception" requires a patient coming from the outside world, and an idle receptionist. The queues are passive states where the entities should wait until all the necessary conditions for beginning the next activity are achieved. The time spent in the passive state depends on the future state of other entities because a particular activity can only start if all the entities involved are available. For example, the activity “triage” can only start if there is a patient and a nurse available.

An entity oriented cycle diagram is used to represent the relationship between entities, activities and queues involved in the hospital admission process. The formulation approach adopted here helps the user create the life cycle of all entities and provides tools to check the integrity of the formulation (De Oliveira, 1998). There are three basic modules: Entity, Activity and Queue. The entity module creates the entities and their life cycle. The life cycle of the patient is a sequence of queues and activities. The activity module keeps track of all information provided by the entity module in a way that the user is able to check if the formulation is all right. It displays, for every activity, the entities involved, which queue they come from and to which queue they go after the activity finishes. The queue module is designed to look at the queues belonging to the life cycle of the entities involved. One is able to set up the queue disciplines, the initial conditions, and to make options for the output. This module completes the formulation of the problem.

4.3 Modelling

The simulation process is described as follows. There is a real world problem. This problem is formulated as a logical model. Logical models can be activity cycle diagrams, flow charts, block diagrams etc. There are a variety of ways to represent the logic of a formulated problem. The next step is to convert the logical model into a computer model. This computer model is verified and tested to see whether it is doing what the analyst wants. The model is used as an operational model to produce some results, or some conclusions, or for implementation after the operational model has been validated against the real world. The conceptual model is formulated from the analysis of the patient flow. The model includes both the outpatient and the emergency sectors. It considers the human and material resources related to each process. The processes are the registration, the triage, the outpatient care and the emergency care.

4.4 Parameters estimation

The data collection is performed in stages. The first stage is the elaboration and the analysis of the treatment flow. Data collection, statistics and analysis of the flow made for each type of patient. The results are validated by members of the hospital administration. The second stage is the assemblage of the necessary information for the definition of the conceptual model. That is: service times, rate of service for outpatient and emergency patients (in the reception, the general clinic, the paediatrics, the orthopaedics, the surgical clinic and the emergency room). The third stage is the collection of data referring to typical days in the emergency unit operation. The patient arrival process is compiled from the service report, as well as the service time in the reception, in the general clinic, the paediatrics, the orthopaedics, the surgical clinic and the emergency room. Table 2 shows the average service time and probability distributions for some of the services.

<table>
<thead>
<tr>
<th>Specialty</th>
<th>Average Time (minutes)</th>
<th>Service-time</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Medicine</td>
<td>4.18</td>
<td>Erlang</td>
<td>ER(4.18;1.00)</td>
</tr>
<tr>
<td>Paediatrics</td>
<td>7.27</td>
<td>Lognormal</td>
<td>L(7.27;2.56)</td>
</tr>
<tr>
<td>Orthopaedics</td>
<td>9.9</td>
<td>Lognormal</td>
<td>L(9.90;3.52)</td>
</tr>
<tr>
<td>Surgery</td>
<td>9.9</td>
<td>Lognormal</td>
<td>L(9.90;3.52)</td>
</tr>
</tbody>
</table>
These probability distributions are adjusted to the data, according to Qui-Square, Kolmogorov-Smirnov and Anderson Darling tests. The result suggests that the arrival of patients occurs according to a Poisson process. The time interval between the arrivals is estimated as 3 minutes for the outpatients and of 30 minutes for the emergencies. The service time for the paediatrics, orthopaedics and surgical clinics are best adjusted to a lognormal distribution and the Erlang distribution is adjusted for the general clinic. It is often difficult to try and collect a reliable number of data points when several patient profiles and many caregivers are involved. When this is the case, an estimate of the distribution should be made. The estimation is based upon the doctors’ experience. It is established that the service time varies between 20 and 40 minutes, 30 minutes being the most common. The triangular distribution for the doctors’ activity time in the emergency room is estimated as: minimum of 20, maximum of 40 and the most frequent value equals 30 minutes.

4.5 Experiments with the model

There are several ways to execute a simulation. The approach to be adopted depends on the nature of the problem. The most common approaches to build a DES are: events, activities and processes. All approaches have in common the fact they produce programs with a three level structure: The executive level is a control program, responsible for sequencing the operations. The Operation level has statements that describe the interactions that make up the model. The third level includes details of the system, sampling, statistics and reports.

In the event approach, the executive produces a “diary” into which future event notices are written. The time scan determines the time of the next event, moves the simulation clock and produces a current event list. The next step is to carry on the event execution. In this approach the analyst is required to state all possible outcomes from each event. The second level produces a set of “event routines” which describes the operations that could follow from a particular change in the system. The activity approach concentrates on the interactions of the various classes of entity. The basic building block is the activity. The test head are the conditions that must be satisfied and the actions are the operations which constitute the activity. The executive makes a repeated activity scan, checks the attributes of the permanent entities and indicate, when each activity is due to change state. The process approach concentrates in the sequence of operations through which an entity faces during its life within the system. Its progress continues until the entity is “delayed”. The activity is like an unconditional delay and the queues are like a conditional delay. The level 2 consists of a set of processes. There is a reactivation point and the entity is held until reactivation occurs, due either to the end of a scheduled delay or to favourable conditions within the simulation.

The simulator adopted here uses the process approach that suits the available data and reflects the patient flow. The model covers a running period that starts at 7am and ends at 11pm. The model simulates the flow of patients through the emergency hospital. It examines individual patients as they arrive and pass through the system. The arrival cycles, priority rules and type of service are included to provide the necessary detail to reflect real-life processes. Alternative scenarios can be compared and modified. The knowledge gained from each scenario can be evaluated against set criteria, the status quo, and each other to determine the best option. The focus of this study is the reduction of the patient’s waiting time between reception and effective treatment. The target is increase the capacity to achieve quality. Simulation is a powerful tool to investigate several “what- if” scenarios. The visual representation the entities of the hospital environment has helped the medical staff to understand their role in the process. Figure 4 gives an example of a computer screen of the proposed simulation model.
4.6 Validation
The validation is the process of ensuring that the model is an accurate representation of the current system and performs in the same way. This process is achieved through the calibration of the model comparing the current model to system performance and identifying failures. The developed model is validated by the hospital management and by the medical staff involved in the research. The model is an accurate representation of the real system. Several simulation runs are carried out. The animation component allows a good understanding of the experiment. The participation of the hospital staff is fundamental to ensure consistency and reliability of the project. The verification is the phase of ensuring that the simulation model is correct and performs accordingly. It can be performed by running the simulation and monitoring its operation. The interaction between the users and the analyst speeds up the verification and increases the reliability of the model. The following procedures have been adopted: The logical sequence and the entities scripts are compared with the flow of patients in order to confirm its correct representation in the computational model; The tracking of the model allows monitoring the system’s performance event after event, guaranteeing the accurate representation of the real system; The experiment is carried out several times. The reports are compared to reality and analyzed until the model is finally accepted by the users.

4.7 Implementation
After the validation and the authentication of the model, an experiment is performed to evaluate the effect of the changes in the admission system after the implementation of new features. The first experiment is the evaluation of a new triage system. The parameters taken into consideration are the waiting time for emergency patients and the average number of patients waiting for service in the emergency room. The experiments cover a period of 16 hours (from 7am to 11 pm). Each experiment is carried out 100 times in order to eliminate distortions and to guarantee the accuracy of the results. Figure 5 shows the average waiting time that an emergency patient has to wait to start treatment in the current system and after the introduction of the new scheme.

It is clear that the introduction of more qualified people in the triage reduces considerably the average waiting time. This measure points to the value of a risk classification scheme. It is interesting to reinforce that this was not detected before this experiment. The rise of this constraint restriction has caused a significant time saving and a reduction of the risk involved with the emergency patients. Another important point is the influence of this action in the performance of the treatment flow. Figure 6 shows the relative size of the waiting line before and after the changes. The inclusion of the specialized triage people cause a reduction of the relative number of patients waiting for emergency service.
Figure 5 – Average waiting time (Emergencies)

Figure 6 – Relative number of patients waiting

Figure 7 shows the effect of the increase of the number of qualified nurses in the emergency sector. It can be a considerable reduction of the average waiting time as the number of nurses increases.

Figure 7 – Average waiting time
5 Conclusion
This method improves the access and quality of public emergency hospitals in Brazil. It is noticeable that the hospital managers have been able to take strategic decisions based on the results of the model. Some of the problems found here are common to other hospitals. One of the rewards of this approach is the possibility to foresee the consequences of operational changes before its implementation in the real world. It offers logical procedures that encompass important features of the emergency service. The management of the critical constraints observed in this particular hospital has contributed to improve the performance of the admission system. The improved scheme of risk classification with the presence of skilled professionals represents a paradigm break in Brazilian hospitals. The added information generated by the model leads to improvements in the value of the service. The reduction of the patients’ waiting time from the time of registration and the beginning of the emergency treatment is a very significant achievement and has important implications for the patients and medical staff. The assistance level rises and the patients’ risk declines. The prediction of the possible changes enables the hospital management to evaluate the whole medical care process. In practice, the changes in the system are dependent on where and when changes have to be made. The principles proposed here can be used as a course of action, which leads to improvements. It is expected that the knowledge acquired here can be used by other public emergency hospitals.

6 References