Chapter 7
Simulation-Based Analysis of Patient Flow in Elective Surgery

Dario Antonelli, Giulia Bruno, and Teresa Taurino

Abstract The reduction of waiting lists and length of stay in hospitals, together with an efficient utilization of system capacity is the challenge facing healthcare systems today. In an elective surgery department, as operations can be scheduled in advance, this goal is generally achieved by maximizing the utilization index of the operation theatres. Nevertheless, operations are only one of the many activities performed during patient flow inside hospital and these activities interact with each other. The optimization of any single stage of the process is pointless without an efficient management of the entire routing from admission to dismissal. The paper presents a thorough analysis of the patient flow in an elective surgery ward using data gathered in a large hospital in Italy. Data, derived from log files and questionnaires, together with solutions proposed by healthcare managers, are considered. A model is then built and validated, its parameters are defined, and a variety of experiments are simulated in order to select the solution that improves the performance of the system. The solutions are discussed and refined in the light of corresponding production management approaches.

7.1 Introduction

Health-care resources are getting more and more expensive. The administrators of health-care facilities are constantly faced with the difficult task of balancing the achievement of quality standards of health with the appropriate allocation of resources [1]. Cutting the waiting lists and the length of stay in hospital is therefore an important managerial goal for modern healthcare systems because it increases the

D. Antonelli (✉) • G. Bruno • T. Taurino
Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino,
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
e-mail: dario.antonelli@polito.it; giulia.bruno@polito.it; teresa.taurino@polito.it

A. Matta et al. (eds.), *Proceedings of the International Conference on Health Care Systems Engineering*, Springer Proceedings in Mathematics & Statistics 61,
DOI 10.1007/978-3-319-01848-5_7,

perceived quality of care and frees resources [2]. In elective surgery departments, system administrators can maximize managerial performance parameters only partially, as not all arrivals can be scheduled. The external performance indexes are the waiting time and the waiting list which both impact on the perceived quality. The internal performance indexes are the throughput time (time from arrival to dismissal), bed occupancy, dismissal rate and resources utilization rate. Operations management techniques show the correlations among internal and external parameters [3]. However hospital manager often prefer to adopt a more intuitive approach, trying to get the full occupation of beds and the maximum utilization of every resource. To this aim, several different tactics have been adopted: use of priority levels in the discipline of the waiting list, scheduling of patient arrival, increasing the utilization of operation theatres by reducing idle times and the redesign of the procedures for patient accommodation on wards. It is worth noting that changing this procedure is yet another way to discipline the waiting list after the patient has been hospitalized.

In present study, Discrete Event Simulation (DES) is applied to simulate the effects of interventions on pre-hospitalization and on bed allocation for an elective surgery department in an Italian hospital. The main factors influencing patient flow are extracted and analyzed in order to find key solutions for the improvement of the system's performance. The results are discussed by using analogies with the PULL (demand driven) production processes.

7.2 Problem Description

The case study considered in this work regards a large hospital in Italy. Just one division is taken into account for the current analysis: an elective surgery department. In management terms, it is a process with scheduled arrivals.

In order to optimize system performances, strategies proposed by the hospital management were simulated through experiments and were compared, extending a method already applied in a former study [4].

Simulated experiments were conducted by following prescribed formal stages: system observation, data collection, model implementation, run and validation, output analysis.

Several practical issues arose during the experiment such as errors in the collected data, high variability of system parameters, and self-adjusting behavior of personnel.

It is important to bound the analyzed case study on the type of patient considered, the inpatient. An inpatient is "admitted" to the hospital and stays overnight or for an indeterminate time. An early selection of inpatients from the outpatients could considerably reduce the waiting time. Thus, as diagnostic is not an exact science, it is unavoidable that triage admits some outpatients, too.

An important performance index, directly perceived by every patient, is related to the length of the waiting time before hospitalization. In order to improve this issue, it is possible to adopt different tactics:

- Queue discipline based on priority rules (already adopted).
- Improve the scheduling of patient arrival.
- Increase the utilization of surgery rooms.
- Redesign the procedures for the accommodation in wards beds.

Also patient scheduling was adopted by many hospitals but not everywhere [5]. Scheduling is effective only when the scheduled system is deterministic or with low variability. This is not the case as recovery times display a variance equivalent to the mean times. Therefore the ward under analysis uses a flexible scheduling in which it is scheduled only the date from which the patient should be ready for hospitalization, with the results of the diagnostic exams. Starting from that date, the actual hospitalization will occur as soon as a bed is actually free.

Alternatively pre-hospitalization analysis are a way to hospitalize patients just in time for the operation, saving beds [6].

Another improvement would be to cluster beds in two groups: standard stay patients and long terms patients. These latter delay the admission of new patients to surgery. The relative size of the two groups can be reallocated based on the demand [7]. Experiments on the actual patients are not advisable therefore it was decided to have recourse to simulated experiments.

Several approaches could be used to model and optimize patient flow: Markov and semi-Markov models, queuing theory, solved analytically or by discrete event simulation [8, 9]. Queuing theory models are usually based on some simple assumptions such as exponential inter-arrival and service time. However, for complex real-world systems, DES models are more flexible and adaptable [10, 11]. The model of the patient flow takes the form of a queuing network with G/G/m servers, there are m workstations in the server, the queue, intended as the waiting list, is virtually unlimited and the inter-arrival times and the process times follow a general distribution.

7.3 Elective Surgery Department

7.3.1 A Description of the Case Study

In the considered Elective Surgery Department, data were collected from different sources: the recovery logs (made anonymous) in the year 2008, integrated by inormation gathered through a questionnaire filled by hospital personnel.

When a surgery date is scheduled, the patient may be required to undertake pre-operatory analysis, such as laboratorial samples, cardiovascular and respiratory tests. Regular patients have a priority discipline for their waiting in queue. A triage is performed to assign a priority order to each single patient, with descending priorities A, B, C and D.

It is also important to state that patients ranked as B, C and D sometimes receive this designation because they must still undertake prior examinations before hospitalization that are mandatory for the surgical procedure. This forces them to wait longer before admission.

There is another category of patients, named urgent patients. Urgent patients arrive from other Wards as they have to submit to a surgical operation as a consequence of other diseases that were treated non surgically. They obviously don't have to undergo examinations as they are already hospitalized.

After entering the hospital, all patients are treated equally, disregarding their queuing priority and the surgeries follow the rule of First In First Out.

Whenever entering the hospital, a patient is allocated to a bed occupying this resource until the end of its recovery. It is clear that a patient only enters after there is a free bed.

During hospitalization, visits and examinations can be executed on patients (especially on patients A since the others had time to undergo examinations during the waiting). As a consequence of the analysis some patients are treated without recurring to surgery. Some patients may undergo complications during the surgery requiring, then, a second intervention that is executed as soon as possible. This is the only case in which the FIFO rule for the access to the operation theatre is not respected.

The previously described system was represented by means of a Process Flow Diagram that is reported in Fig. 7.1.

The process flow follows the vertical line from the top to the bottom, circles represent operations or activities, arrows represent transfers and the delay symbols represent waiting times according to the ASME (American Society Mechanical Engineering) symbology [12].

7.3.2 Data Collection

In the ward there are, totally, 24 beds, equally divided between the two genders. One of these beds is usually reserved for urgent patients. In the ward there is a single surgery room (also called operation theatre), and surgeries are only performed on Monday, Wednesday and Friday. According to the managerial staff, the estimated amount of surgeries performed in a week is 15.

The data collected from the log files of the ward cover the months of January and March 2008 for a total of 112 patients (i.e., patients that enter the hospital in those 2 months). From the logs it's possible to gather the percentage of patient types that

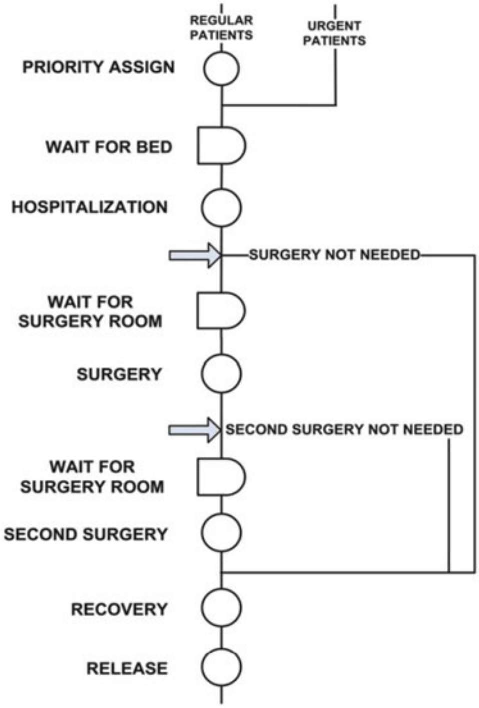

Fig. 7.1 Process Flow Diagram of the elective surgery department

Table 7.1 Basic statistics about the patients

Number of patients	Patient type	Percentage (%)
112	All	100
65	Regular A	58.0
32	Regular B	28.6
2	Regular C	1.79
1	Regular D	0.89
12	Urgent	10.7
112	First surgery	100
9	Second surgery	8.03
1	Third surgery	0.89
109	Released	97.3
2	Deceased	1.79
1	Transferred	0.89

arrived at the hospital, the number of operation they needed and the way patients got out the hospital. All these data are reported in Table 7.1. Since there are only three patients that belong to patient types C and D, and thus the number is not significant to model their distribution, they are considered as assigned to class B. The only patient that needed the third surgery was not modeled as a case a part from the others, but was included in the patients that needed two surgeries. An outlier

Table 7.2 Means and standard deviation of time spent by patients (in days)

Time	Mean	S.D.
WT1	26.2	35.8
WT2	4.33	4.86
WT3	0.46	1.93
RT	5.95	5.23
HT	10.7	8.46

Table 7.3 Means and standard deviation of time spent by patients (in days)

	Type A		Type B		Urgent	
Times	Mean	S.D.	Mean	S.D.	Mean	S.D.
WT1	18.2	23.8	49.2	46.9	–	–
WT2	4.14	3.64	3.6	3.49	7.82	10.9
WT3	0.78	2.48	–	–	–	–
RT	5.86	5.25	4.71	4.09	9.82	6.78
HT	10.9	7.00	8.31	6.78	17.6	15.8

patient that presents a waiting time before the second surgery of 89 days has been considered as an error of data entry and removed since it seems unfeasible that a patient remains for such a long time in the considered ward. The total number of considered patients is thus 111.

From the log data it's also possible to gather the times patients spent in the process. Particularly, the following times are analyzed: the waiting time before hospitalization (WT1), the waiting time from hospitalization to first surgery (WT2), the waiting time from first until second surgery (WT3), the recovery time (RT), and the total time spent inside the hospital (HT, equal to the sum of the three previous times). The derived mean and standard deviation of such times for the considered 111 patients are reported in Table 7.2.

A further analysis of these data is done in order to see differences in behaviors of patients based on their typology. The mean values and standard deviations of times for each type of patients are reported in Table 7.3.

Regarding WT1, there is a strong difference among the behavior of patients belonging to the three categories. As a matter of fact, urgent patients usually do not wait until entering the hospital, while patients A wait on average 18 days and patients B wait on average 49 days. Also the waiting for surgery (WT2) is quite different among the categories, and interestingly patients B usually wait less than the others (only 3.6 days on average), while urgent patients wait more than twice the time of patients B. This can be due to the fact that patients B have a long wait time before the hospitalization, which they can use to perform some examinations, thus saving time for when they are inside the hospital. On the contrary, urgent patients come directly to the hospital, and thus probably have to perform other kinds of analysis before being allowed to the operation.

For WT3 we do not have enough cases to differentiate among the three categories, having only 11 patients needing the second surgery, all of type A. The differences for RT are similar to the ones of WT2.

7.4 Simulation of the Process

7.4.1 Process Workflow

The process described in Fig. 7.1 has been modeled using the Rockwell Arena software [13] to perform the simulation. To assign the distribution of waiting times and patients' arrivals, data obtained by the hospitals were exploited to find the expression that best estimates data distributions. The Kolmogorov-Smirnov (K-S) test [14] is applied to select the best distribution.

The best probability distribution function that fits the arrival rate of patients is the Exponential distribution with mean equal to 0.53.

In the simulation, a priority level is randomly assigned to each patient in order to reflect proportions found in data (i.e., 58% of type A, 31% of type B and 11% of Urgent). Once the patient is assigned a type, it enters in a queue representing the waiting until there is an empty bed (WT1). The queue is of Lowest Attribute Value type, i.e., the precedence is given to patients with the lowest priority value, according to the real procedure in which patients of type A (i.e., priority level 2) have the precedence over patients of type B (i.e., priority level 3), and urgent patients (i.e., priority level 1) have the precedence over both of them.

Since all patients went through surgery, all of them spend some time waiting for the surgery room (WT2). The distributions that best fit the delays for surgery depending on patient type are reported in Table 7.4 (first column). Then, a decisional process sends some patients (7% of cases) to the second operation, represented by the delay process in which a patient waits for the second operation. The data distribution follows the expression $0.5 + 11$*BETA(0.802, 0.757). Finally, all patients perform a recovery step before leaving the hospital. From interviews to domain experts it appears that the distribution of recovery time is independent from the patient type. Therefore we put together all the values to provide an estimation of the distribution; the retrieved expression is $-0.5 +$ GAMM(3.11, 2.07). Table 7.4 reports the time distributions adopted in our simulation.

7.4.2 Simulation parameters

A simulation of the workflow of the process was executed with parameters' values reported in Table 7.5. The Warm up period was chosen using the Welch method.

Table 7.4 Process times distribution for each patient type

Patient type	WT2 distribution	WT3 distribution	RT distribution
Type A	−0.5 + WEIB(5.02, 1.3)	0.5 + 11*BETA (0.802, 0.757)	−0.5 + GAMM (3.11, 2.07)
Type B	−0.5 + LOGN(4.22, 4.5)	"	"
Urgent	−0.5 + WEIB(6.96, 0.76)	"	"

Table 7.5 Simulation parameters

Parameter	Value
Number of replications	100
Warm-up period	730 days
Replication length	3,650 days

Table 7.6 Average values of obtained results for the standard case

Field	Real average value	Standard simulation average value 95% confidence interval
WT1	26.16	1.44 ± 0.11
WT2	4.33	4.41 ± 0.01
WT3	0.46	0.43 ± 0.005
RT	5.95	5.93 ± 0.01
HT	10.74	10.77 ± 0.13
Waiting patients	–	2.71 ± 0.22
Bed utilization rate	0.85	0.88 ± 0.001
Busy bed	20.4	20.28 ± 0.07

The obtained results in term of 95% confidence intervals for average values of WT1, WT2. WT3, RT, HT, number of waiting patients in the queue, bed utilization rate and number of busy beds are reported in Table 7.6 compared with real average values. The half-widths of the confidence intervals for average values suggest that an acceptable level of convergence is reached after 100 replications.

All of the values obtained in the simulation are coherent with the real data, except the waiting time WT1, that in the simulation is significantly lower. This is due to the fact that when a patient asks for a schedule, the admission date is not calculated by analyzing the current waiting list only, but adding a further 2 weeks to the date in order to allow the patient to perform some pre-operation exams. Thus, the length of this time depends not only from current resources or from organization of the ward, but also from the management rules of patients.

7.5 Proposal of Improvement

In the simulation of the ward, i.e., in the current state, the bed utilization rate is, on average, less than 90%; particularly, the utilization rate is in the range 0.88 ± 0.03. The objective of the ward's managers is to increase the utilization rate of beds to a value close to 0.95. Thus, we performed a simulation by changing the arrival rate to reach the desired utilization rate, in order to evaluate effects on waiting queues. As can be seen from Table 7.7, this change causes a sudden increase of waiting times.

The desired utilization rate is reached by decreasing the average time between arrivals from a value of 0.53 days (less than two patients a day) to a value of 0.49 days (more than two patients a day). Simulation results show that the average time spent to wait for a bed (WT1) strongly increases from 0.84 to 6.30 days with an average number of waiting patients of almost 13. In the last column of Table 7.7

Table 7.7 Average values of obtained results for the case with more patients

Field	Average value	Range average
WT1	6.30	(2.40, 16.31)
WT2	4.42	(4.27, 4.57)
WT3	0.43	(0.36, 0.49)
RT	5.94	(5.81, 6.05)
HT	10.78	
Waiting patients	12.87	(4.77, 34.30)
Bed utilization rate	0.96	(0.92, 0.99)
Busy bed	22.01	(21.24, 22.69)

range values for average obtained in replications are given. Thus, the problem becomes how to meet the manager objective without having such a worsening of performance.

The main point is that, if the ward is considered equivalent to a production line, buffers are not allowed (i.e., patients cannot be hospitalized without available beds). Therefore the ward corresponds to a pull system: the admittance of a new patient is based on the system status (availability of beds). Pull systems suffer from variability and unfortunately present case has high variability, as can be seen in Table 7.3. In industrial management, if a system displays an high variability it can be buffered by increasing the capacity, the WIP or the waiting time. Increasing the capacity (beds) has a direct cost. WIP increasing in this case is unfeasible because it corresponds to adopt an office-based surgery that has been excluded a priori for inpatients. The last way is by increasing the total cycle time that is the exact opposite to the objective of ward's managers.

To improve the system with no additional costs, another way exists: by addressing efforts directly to the reduction of variability on waiting times before surgery (WT2), for example by reducing the number of exams done during the hospitalization by increasing the pre-hospitalization activities. This operation involves a reorganization of the admission and recovery process and can be done by reinforcing a pre-hospitalization process. Infact, trying to anticipate some examinations before the admission to the ward can reduce the waiting time inside the hospital.

To simulate this scenario, the variance of waiting time before the first surgery (WT2) has been reduced by considering a process organization that admits exclusively Urgent patients and patients with a pre-hospitalization period (B patients) and by considering waiting time before the surgery equal to real average values. Table 7.8 reports results obtained by this simulation, which shows a consistent reduction of patients' waiting time and of queue length.

7.6 Conclusion

In this work an engineering approach is used to provide a process parameterization in order to reach managerial objectives of beds utilization. A simulation of the new process is done to test proposed parameters. Simulation results shows that small

Table 7.8 Average values of obtained results for the case after re-organization

Field	Average value	Range average
WT1	2.39	(1.05, 5.72)
WT2	4.06	(4.02, 4.09)
WT3	0.43	(0.36, 0.49)
RT	5.94	(5.82, 6.05)
HT	10.44	
Waiting patients	4.88	(2.09, 12.02)
Bed utilization rate	0.92	(0.89, 0.96)
Busy bed	21.26	(20.56, 22.01)

variation on the average value of inter-arrival times cause significant variations on waiting times. So a solution to find a compromise between bed utilization and waiting times is provided and simulated. The idea is to reduce variance on waiting times before surgery with a reorganization of the admittance process, placing more emphasis on the pre-hospitalization phase. On the basis of his/her objectives and requirements, the healthcare manager is provided with better guidance for an informed choice.

Acknowledgements The authors would like to thank Prof. Baudolino Mussa (University of Torino Medical School, Italy) for his support and fruitful hints.

References

1. Culyer, J.G., Cullis, J.G.: Some economics of hospital waiting lists in the NHS. J. Soc. Policy **5**(3), 239–64 (1976)
2. Antonelli, D., Bellomo, D., Bruno, G., Villa, A.: Evaluating collaboration effectiveness of patient-to-doctor interaction in a healthcare territorial network. In: PRO-VE 2012: Collaborative Networks in the Internet of Services, pp. 128–136. ISBN 9783642327742 (2012)
3. Young, T., Brailsford, S., Connell, C., Davies, R., Harper, P., Klein, J.H.: Using industrial processes to improve patient care. Br. Med. J. (BMJ) **328**, 162–164 (2004)
4. Antonelli, D., Taurino, T.: Application of a patient flow model to a surgery department. In: 2010 IEEE Workshop on Health Care Management, Venezia, 18–20 Febbraio 2010
5. Gupta, D., Denton, B.: Appointment scheduling in health care: challenges and opportunities. IIE Trans. **40**(9), 800–819 (2008)
6. Qi, E., Xu, G., Huo, Y., Xu, X.: Study of hospital management based on hospitalization process improvement. Proc. IEEE IEEM 74–78 (2006)
7. Akkerman, R., Knip, M.: Reallocation of beds to reduce waiting time for cardiac surgery. Health Care Manag. Sci. **7**, 119–126 (2004)
8. Vissers, J.: Patient flow-based allocation of inpatient resources: a case study. Eur. J. Oper. Res. **105**, 356–370 (1998). Elsevier Science
9. Xiong, H.H., Zhou, M.C., Manikopoulos, C.N.: Modeling and performance analysis of medical services systems Using Petri Nets. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 2339–2342 (1994)
10. Davies, R., Davies, H.T.O.: Modeling patient flows and resource provision in health systems. Omega Int. J. Manag. Sci. **22**, 123–131 (1994)

11. Koo, P.-H., Jang, J., Ielsen, K.B., Kolker, A.: Simulation-based patient flow analysis in an endoscopy unit, Health Care Management (WHCM). In: 2010 Institute of Electrical and Electronics Engineers (IEEE) Workshop on, pp.1–6, 18–20 Feb 2010
12. American Society of Mechanical Engineers: ASME Standard Operation and Flow Process Charts. New York (1947)
13. Kelton, W.D., et al.: Simulation with Arena. McGraw-Hill Professional, New York, NT, US (2006)
14. Smirnov, N.S.: Tables for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. **19**, 279 (1948)