Analysis of patient flows in elective surgery: modelling and optimisation of the hospitalisation process

Dario Antonelli, Giulia Bruno* and
Teresa Taurino

Department of Management and Production Engineering,
Politecnico di Torino,
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
Email: dario.antonelli@polito.it
Email: giulia.bruno@polito.it
Email: teresa.taurino@polito.it
*Corresponding author

Abstract: Rationalising costs while guaranteeing a good quality of service is a challenge that healthcare systems are currently facing. In elective surgery departments, as operations can be scheduled in advance, the goal is usually to maximise the utilisation index of the operating theatre. Nevertheless, the optimisation of a single stage of the process is pointless without an efficient management of the entire routing from income to dismissal. The paper presents discrete events simulation of the actual patient flows in elective surgery exploiting the recovery logs of a hospital department. A UML activity diagram of the surgery process together with the collected hospital data have been used to build a stochastic model of queuing network, identify its parameters and conduct different simulated experiments in order to select the solution that best optimises the performances of the system. The simulation results have shown that there is a large variation in waiting times in correspondence to small variations of the average value of the inter-arrival times. Therefore, solutions that optimise utilisation indexes of both beds and operating theatres should consider a concurrent effort to reduce the variance of admittance processes, otherwise the waiting times will lengthen beyond acceptable limits.

Keywords: healthcare systems; elective surgery; simulation; patient flow; unified modelling language; UML.

Reference to this paper should be made as follows: Antonelli, D., Bruno, G. and Taurino, T. (xxxx) ‘Analysis of patient flows in elective surgery: modelling and optimisation of the hospitalisation process’, *Int. J. Services and Operations Management*, Vol. X, No. Y, pp.000–000.

Biographical notes: Dario Antonelli is an Associate Professor of Technology and Production Systems at the Politecnico di Torino, Italy. His main research interests concern the development of strategies and methods for human-robot collaborative working cells and the finite elements simulation and optimisation of metalworking processes. He has been the team leader of several EC-funded research projects: ECHORD-FREE, AMICO, @CARE, ADIUVARE. He is the Director of the degree course of Engineering and Management at the Politecnico di Torino, and a member of the Scientific Committee of PRO-VE and VINORG.

Giulia Bruno is a Post-Doc Researcher at Politecnico di Torino from 2009. She holds a PhD in Information and System Engineering and a Master of Science in Computer Engineering from Politecnico di Torino. Her research activity is focused on knowledge management, data mining, simulation, and production system analysis. She is also working in collaboration with healthcare sanitary agencies to analyse patient flows. She was involved in several European projects in the context of product lifecycle management, healthcare systems and SME networks.

Teresa Taurino is a researcher in the Department of Management and Production Engineering at Politecnico di Torino in Italy. She holds a degree in Mathematical for the Engineering Sciences, a Master degree in Mathematical Engineering and a Doctorate in Production Engineering attended in Politecnico di Torino. Her main research area is on organisation and management of production systems and on methods for the performance evaluations. Her research is focused on the analysis of relations and peculiarity of network of industrial systems and healthcare services.

1 Introduction

Healthcare services have the troublesome goal of achieving high quality standards of cures despite increasing costs and escalating process complexity. Performance measures related with the goal are admission waiting list and length of stay (Haraden and Resar, 2004). They impact both on the perceived quality of care and on the efficiency of the management (Antonelli et al., 2012). Their optimisation could require actions at different levels: single ward, whole hospital, network of hospitals in a region.

The rules of operation management state that these measures, called external performance indexes, are tightly correlated with measures inherent to the organisation like throughput time (time from admission to dismissal), bed occupancy, dismissal rate and resource utilisation rate, called internal performance indexes (Young et al., 2004).

Therefore, it is necessary to coordinate the whole set of parameters in order to reduce the waiting list in hospitals. Conversely, hospital managers tend to adopt an intuitive approach addressing directly the length of the queue. To this aim, different tactics have been proposed (Wilson and Nguyen, 2004): priority levels in the queue of the admittance waiting list, optimised scheduling of patient arrivals, increased utilisation of operating theatres through a reduction of the idle times and redesigning of the procedures for patient accommodation in the ward beds.

Unfortunately, these tactics are ineffective as far as all the performance indexes are not addressed in a comprehensive way. Furthermore, often managers neglect to consider the fundamental role of variability of process times on the building up of the waiting queue.

Present study aims at showing the exactness of this statement in the case study of the optimisation of patient flow in an hospital’s surgery ward adopting elective admission. The optimisation technique makes use of patient flow simulation using queuing networks. Analysis of inpatient flow is less common in literature with respect to the outpatient flow, due to the difficulty to procure the data by direct observation: hospital stays for inpatients can be in the order of weeks or even months. Some data can be obtained in literature.

Recently, Armony et al. (2015) shared a dataset of patient flow data for a 1000-bed hospital along a time horizon of three years.

The data for present study were obtained on the field by using the recovery logs of an elective surgery in an Italian hospital. The availability of a rich set of data allowed to extract the main statistic time distributions of the different process along the course of treatment. Other authors (Schmidt et al., 2013; Pinto et al., 2014) assumed the time distributions, e.g., the exponential distribution. The approach is justified in these studies because the number of patients is very large as they consider a whole hospital or even several ones.

In present case study, the number of patient is relatively low: 800 patients per year, approximately two admissions per day. The method employed to assess the goodness of fit for the chosen statistic distributions is explained in Section 4. Furthermore, the correctness of the simulated model was tested in Section 4 against the measures taken from the actual surgery department.

Some improvements strategies proposed by hospital managers were experimented on the simulated model: interventions on pre-hospitalisation and selective bed allocation. As expected the strategies showed negligible effects as far as the variability of the process was not reduced. This result is discussed and justified in Section 5 by using analogies between elective admission wards and classic demand driven industrial processes.

2 Related works

In order to simulate the system, it was necessary to develop a conceptual model of the elective surgery department. Presently, queuing theory and stochastic simulation are the most adopted methodological tools for the modelling of patient flow in hospitals (Cochran and Roche, 2008; Lamiri et al., 2008). Nevertheless, before building the queueing network, it is necessary to have a dynamic model of the process that represents the actual operations and the redesigned processes. In this study the Activity Diagram from unified modelling language (UML) was chosen (Fowler and Scott, 2000). Modelling was just the starting point of the optimisation procedure. Since healthcare refers to human beings, optimisation was performed with the help of simulated experiments.

UML – activity diagrams are suitable for describing system behaviour. However, the model itself is static, because UML does not provide the possibility of running the model. The designed UML diagram has to be transformed into a simulation model to be run within a specialised simulation environment (Teilans et al., 2008). Several works followed this procedure in the healthcare domain. For example, in Reindl et al. (2009), the process flow of cataract surgeries was modelled in a German eye hospital with an UML activity diagram and then a simulation model was produced to reduce the waiting time of the patients and increase the utilisation of the operating theatres. A methodology which combines the use of Petri Net and UML to model a business process as a system of discrete events was proposed by Pels and Goossenaerts (2007).

Regarding the stochastic simulation of the patient flow, there are several studies in literature, but many of them addresses different aspects of the problem or different problems. As an example, Schmidt et al. (2013) simulate the whole hospital, but in the Italian healthcare the patients are scheduled separately for every department. Pinto et al.

(2014) consider in their study all the public hospitals in the city of Belo Horizonte whose admission procedures are managed together by a Central. Meng et al. (2015) present a similar case study but they share emergency and elective surgery in the same department. Most of the authors use simulation as a support for the scheduling of patient admissions.

Few authors, like Montgomery and Davis (2013) try to employ simulation as a decision support tool for the redesign of admittance and recovery procedures. This is exactly what it is done in present study.

3 Method

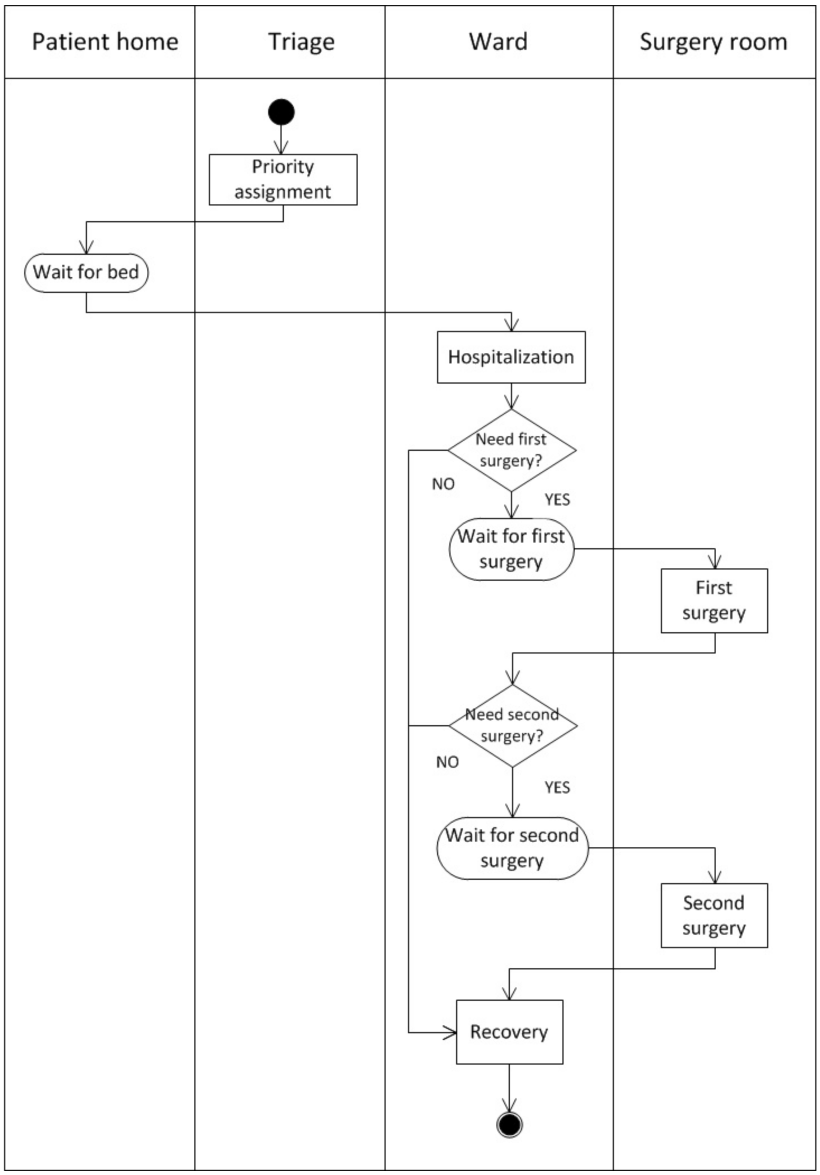
The case study considered in this work is the Oncology Department of a large hospital in North West Italy. In a former study (Antonelli and Taurino, 2010), the patient flow of the whole hospital was simulated, then in a following study (Antonelli et al., 2014), the department alone was simulated. The comparison with actual hospital data showed a good fit for the external indexes.

Present study aims at applying the simulation as a decision support tool to assess the strategies proposed by the hospital management. The simulated experiments were exploited to compare different admission process redesign scenarios. It is important to connect the analysed case study to the type of patient being considered, that is the inpatient. An inpatient is 'admitted' to the hospital and stays overnight or for an indeterminate time. An early selection of inpatients from outpatients could considerably reduce the waiting time. Thus, as diagnostics is not an exact science, the triage group unavoidably also admits some outpatients. An important performance index, which is perceived directly by each patient, is related to the length of the waiting time before hospitalisation. In order to improve this index, it is possible to adopt different tactics, such as a queue discipline based on priority rules (which has already been adopted in most hospitals), improvement of the scheduling of patient arrival, increased utilisation of operating theatres, and the redesign of accommodation procedures for the department beds.

Patient scheduling has also been adopted by many hospitals, but not everywhere (Gupta and Denton, 2008). Scheduling is only effective when the scheduled system is deterministic or has limited variability. This is not the case here, as RTs display an equivalent or larger variance than the mean times. The analysed department uses a flexible scheduling in which only the date from which the patient should be ready for hospitalisation is scheduled, owning to the necessary medical reports. Starting from that date, the actual hospitalisation will take place as soon as a bed is actually free.

Alternatively, pre-hospitalisation analysis is a way of hospitalising patients just in time for the operation, thus saving beds (Qi et al., 2006). Another improvement would be to cluster beds into two groups: standard stay patients and long-term patients. The latter option delays the admission of new patients to surgery. The relative size of the two groups can be reallocated on the basis of the demand (Akkerman and Knip, 2004). Experiments on actual patients are not advisable, and it was therefore decided to resort to simulated experiments. The modelling can be considered under two different perspectives: clinical and operational (Cote 2000). Following the clinical perspective the patient flow represents the progression of a patient's health status. In the operational perspective the movement of patient through a set of location is considered. We focus on the first perspective that requires routinely collected data and is less costly (Marshall et al., 2005). Several approaches could be used to model and optimise patient flows: Markov and semi-Markov models, the queuing theory, solved analytically or by DES (Xiong et al., 1994; Vissers, 1998). As far as waiting time reductions are concerned, DES appears both more flexible and adaptable (Davies and Davies, 1994; Koo et al., 2010).

Figure 1 UML activity diagram of the elective surgery department



3.1 Model of the elective surgery department

The process representing the behaviour of the elective surgery department was identified after several questionnaires had been distributed and interview had been conducted with healthcare staff as well as from the analysis of anonymous log data which reported the times of admission, surgery and discharge of patients. There are 24 beds in the department, divided equally between the two genders. There is only one operating theatre in the department and surgery is performed on Monday, Wednesday and Friday. According to the managerial staff, an average of 15 operations is performed per week.

When an operation is scheduled, the patient may be required to undertake pre-operatory analysis. Regular patients have priority in the queue. A triage member of staff assigns a priority order to each patient, with priorities of A, B, C and D in descending order. The differences among patients belonging to the different groups are related to the severity of the disease evaluated by the triage process. Patients ranked B, C or D are allowed to carry out further examinations before hospitalisation (pre-hospitalisation).

Another category of patients exists, named urgent patients. They arrive from other Departments and their operations a result of other diseases or illnesses. They obviously do not have to undertake examinations as they are already hospitalised. Regardless of the priority assignment, after entering the hospital, all patients are treated equally. Whenever a patient enters the hospital, he/she is allocated a bed that they will occupy until they are dismissed.

Some patients could be cured without the necessity of recurring to surgery. Some patients may suffer from complications during surgery and require a second operation, which is carried out as soon as possible. The previously described system is represented by means of the UML activity diagram [Fowler and Scott, (2009), Chapter 9] in Figure 1. The simplicity of the diagram is the result of a great deal of filtering work in which all the processes that have been considered important for the medical treatment, but of no influence on the management performances have been dropped.

Figure 2 Distribution of the types of patient types (see online version for colours)

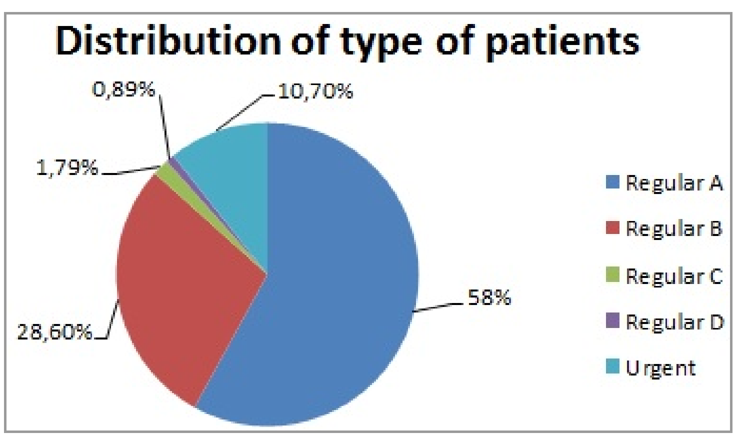
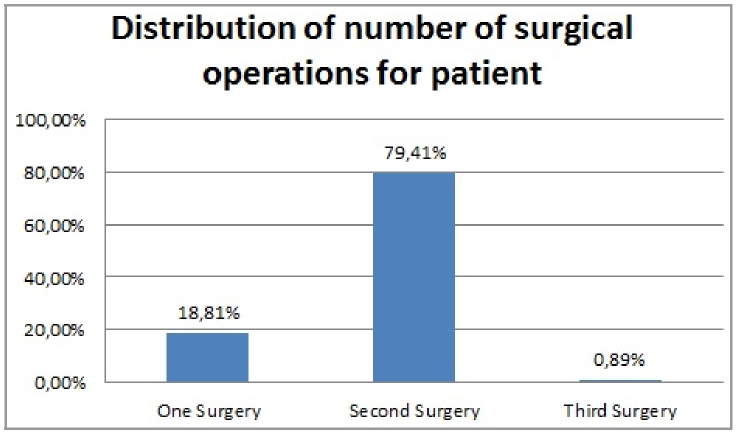


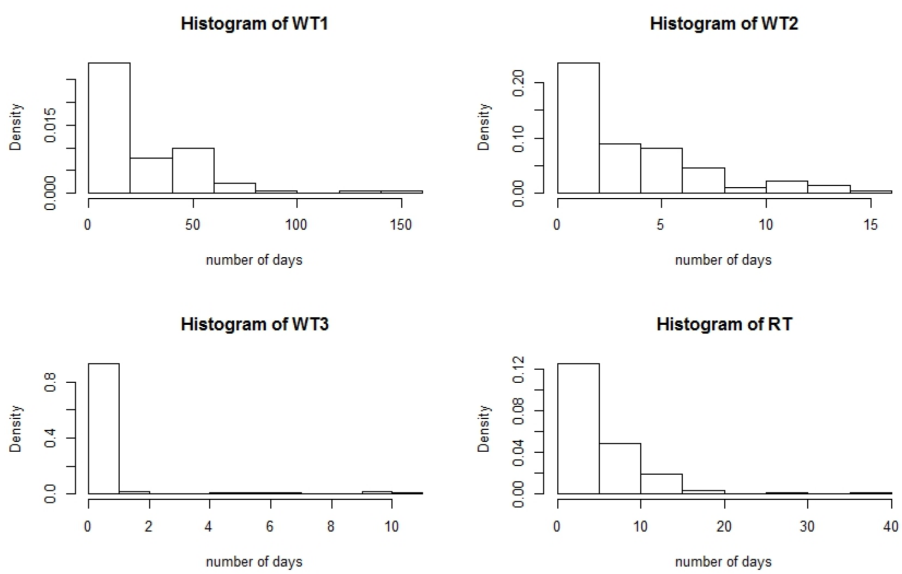
Figure 3 Distribution of the number of operations per patient (see online version for colours)



3.2 Data collection and cleaning

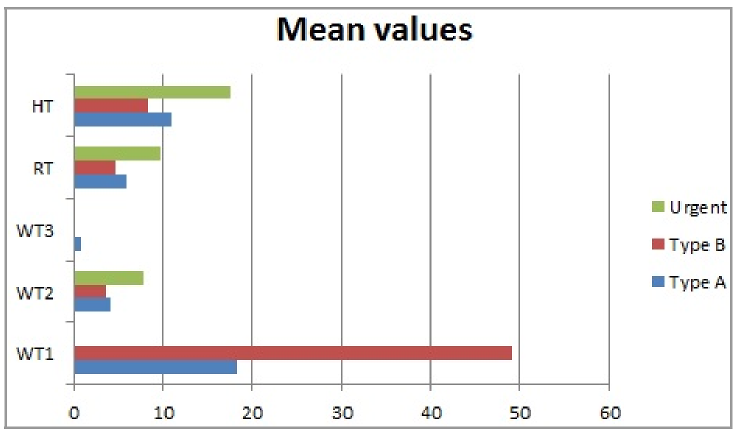
The data collected from the department recovery logs (made anonymously) covered the months of January to March 2008 and pertain to a total of 111 patients. From the logs, it was possible to ascertain the number of patients of different types that entered the hospital (Figure 2), the number of operations they needed (Figure 3) and the way the patients exited from the hospital.

Figure 4 Histograms of the analysed times: waiting time before hospitalisation (WT1), waiting time from hospitalisation to the first operation (WT2), waiting time from the first to the second operation (WT3), RT



In order to focus on the main factors, the model was simplified by dropping some less frequent occurrences, such as the presence of type C and D patients or patients who had to undergo a third operation. From the log data, it was possible to establish the times the patients spent in the hospital. The following times were considered in particular: the waiting time before hospitalisation (WT1), the waiting time from hospitalisation to the first operation (WT2), the waiting time from the first to the second operation (WT3), the recovery time (RT), and the total time spent in the hospital (HT, which is equal to the sum of the three previous times). The histograms that represent the distribution of the collected times for all of the 111 patients are reported in Figure 4.

Figure 5 Mean values of the analysed times: waiting time before hospitalisation (WT1), waiting time from hospitalisation to the first operation (WT2), waiting time from the first to the second operation (WT3), RT, total time spent in the hospital (HT) (see online version for colours)



A further analysis of these data was conducted in order to establish the different behaviour of the patients as a function of their type. The mean values of times for each type are reported in Figure 5. As far as WT1 is concerned, a notable difference can be observed between the three categories. In fact, urgent patients do not usually wait before they enter the hospital, while the patients in group A on average wait 18 days and the patients in group B on average wait 49 days. The waiting for surgery (WT2) is also quite different from category to category, and, interestingly, the patients in group B usually wait less than the others (only 3.6 days on average), while urgent patients wait more than twice the time of the patients in group B. This is due to the fact that patients in group B can carry out all the necessary examinations in advance and can be operated as soon as they are hospitalised. Not enough cases were available to differentiate between the three categories for WT3. The differences for RT are similar to those of WT2.

In Table 1, the confidence intervals for the mean values of the characteristic times of the process are reported.

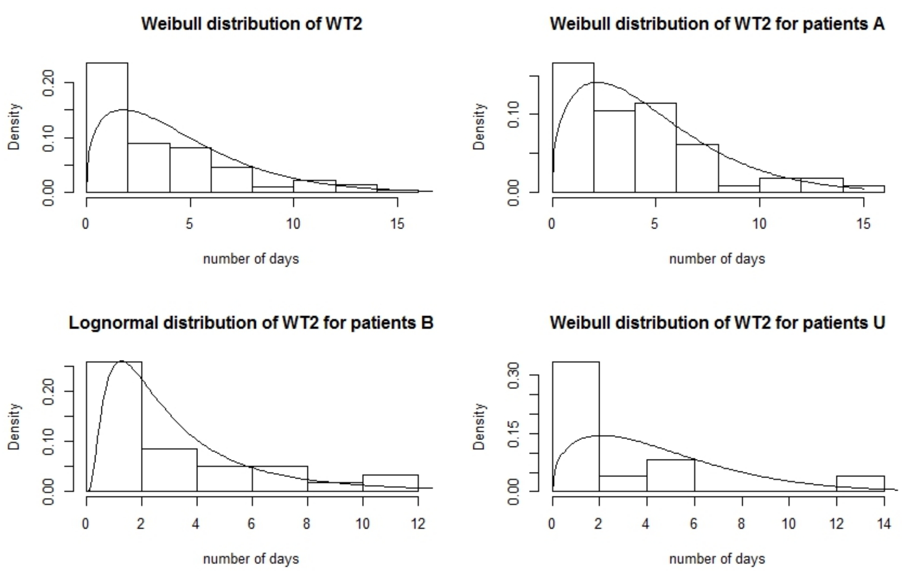
Table 1 Mean values and confidence intervals

	Type A		*Type B*		*Urgent*	
	MV	*0.95 CI*	*MV*	*0.95 CI*	*MV*	*0.95 CI*
WT1	18.45	(13.51–23.39)	49.17	(38.31–60.03)	-	-
WT2	4.13	(3.33–4.92)	3.60	(2.4–4.8)	7.82	(0.51–15.13)
WT3	0.46	(0–1.36)	–	–	-	-
RT	5.95	(4.63–7.28)	4.72	(3.87–5.56)	9.82	(5.26–14.37)
HT	10.80	(9.41–12.18)	8.31	(6.52–10.11)	17.64	(7.02–28.25)

4 Results

The process described in Figure 1 has been converted into a discrete event model and has been simulated with the Rockwell Arena software (Kelton et al., 2010). The experimental data were exploited to find the best estimates for the data distribution (Kleijnen, 2005). The Kolmogorov-Smirnov (K-S) test was applied to find the best distribution (Smirnov, 1948). As the number of patients is sufficiently large, the best probability distribution function for the arrival rate resulted to be the Exponential one, with a mean value of 0.53.

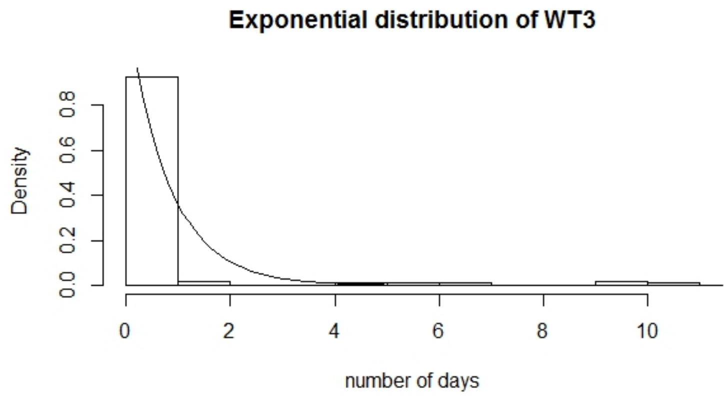
Figure 6 Waiting time from hospitalisation to the first operation (WT2) for all the patients and for the different types of patients



In the simulation, priority levels were randomly assigned to patients to reflect the actual proportions (58% of type A patients, 31% of type B patients and 11% of Urgent patients). Once the patient was assigned a type, he/she entered a queue that represented the waiting until a bed was empty (WT1). The queue discipline is the Lowest Attribute Value. A patient spends time to enter the operating theatre (WT2). The distributions that best fit the delays for surgery, according to the patient type, are Weibull and Lognormal, as reported in Figure 6.

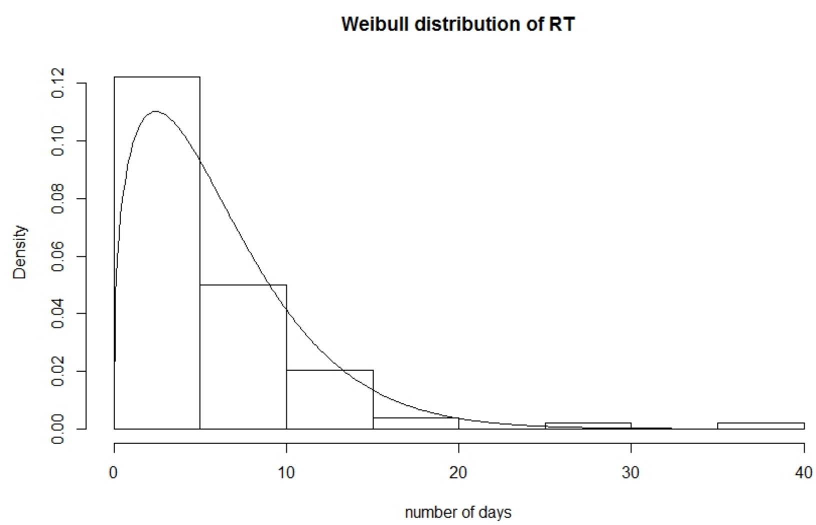
Some patients (7% of the cases) were then sent for a second operation. The data distributions of the waiting times are shown in Figure 7.

Figure 7 Waiting times from the first to the second operation (WT3)



Finally, all the patients undergo a recovery period before leaving the hospital. Since it has emerged, from discussions with domain experts, that the distribution of the RT is independent of the patient type, all the values were considered to obtain an estimation of the distribution. The representations of the real distribution and the theoretical distribution are shown in Figure 8.

Figure 8 Recovery time



4.1 Simulation parameters

A simulation of the process workflow was carried out from the data in Table 2. The Warm up time was chosen using the Welch method. The most significant results, in terms of the average values of WT1, WT2. WT3, RT, HT, the number of patients waiting in the queue, the bed utilisation rate and the number of occupied beds are reported in Table 3 where they are compared with real average values.

Table 2 Simulation parameters

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

Table 3 Average values of the obtained results for the standard case

Field	*Real average value*	*Standard simulation average value*
WT1	26.16	1.44
WT2	4.33	4.42
WT3	0.46	0.43
RT	5.95	5.93
HT	10.74	10.78
Waiting patients	-	2.71
Bed utilisation rate	0.85	0.88
Busy beds	20.4	20.28

All of the values obtained in the simulation are coherent with the real data, except waiting time WT1, which was significantly lower in the simulation. This is due to the fact that when a patient asks for a schedule, a delay of two weeks is added due to hospital procedures.

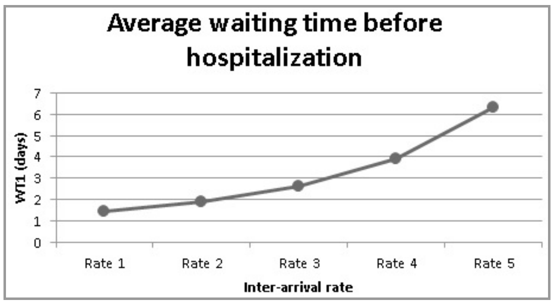
Table 4 Average values of the obtained results for the cases with more patients

Inter-arrival mean value (day/patient)	*Rate 10.53*	*Rate 2 0.52*	*Rate 3 0.51*	*Rate 40.50*	*Rate 50.49*
Number of simulated patients	6.887	7.019	7.157	7.300	7.449
WT1	1.44	1.87	2.64	3.92	6.30
WT2	4.42	4.42	4.41	4.42	4.42
WT3	0.43	0.43	0.43	0.43	0.43
RT	5.93	5.94	5.94	5.93	5.94
HT	10.78	10.79	10.78	10.79	10.79
Waiting patients	2.71	3.62	5.2	7.85	12.87
Bed utilisation rate	0.88	0.90	0.92	0.94	0,96
Busy beds	20.28	20.72	21.16	21.6	22.01

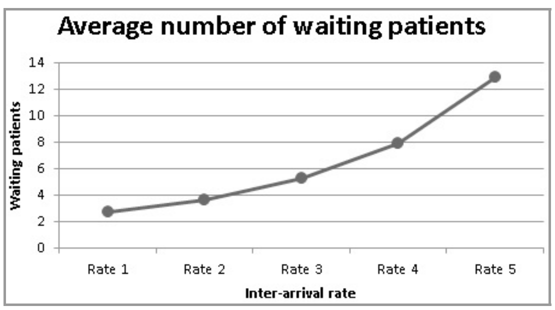
4.2 Proposals for improvement

In the simulation of the department, the bed utilisation rate was, on average, less than 90%, and the utilisation rate was in the 0.88 ± 0.03 ranges. The objective stated by the healthcare managers is to reach a 0.95 utilisation rate. Thus, the inter-arrival mean value was constantly decremented from the actual value (i.e., 0.53 days) in order to find the value to reach the objective of bed occupation. The results of the set of simulations are reported in Table 4.

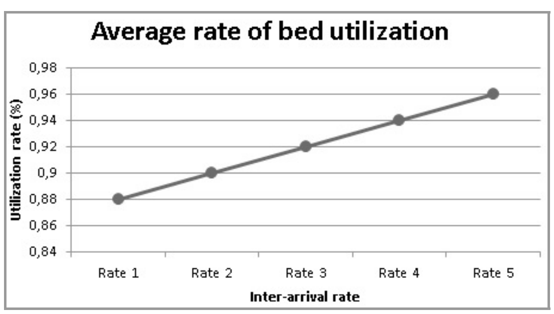
Figure 9 Variation of, (a) waiting time before hospitalisation (b) number of waiting patients (c) bed utilisation rate for different values of patient inter-arrival time



(a)



(b)



(c)

As expected, the decreasing of inter-arrival time caused a sudden increase in the waiting times. The simulation results show that the average time spent waiting for a bed (WT1) increased to a great extent from 1.44 to 6.30 days, as reported in Figure 9(a). Figure 9(b) and Figure 9(c) show that by increasing the frequency of patient arrival, the average number of patients waiting for a bed increases from 3 to almost 13, with a trend similar to WT1, while the bed utilisation average rate increases linearly.

The problem emerging from these results is to find a way to meet the objectives of managers without affecting the performances of the system. By considering the department equivalent to a production line, buffers are not allowed (i.e., patients cannot be hospitalised without available beds). Therefore, the department corresponds to a constant work in process (CONWIP) system: the admittance of a new patient is based on the system status (availability of beds). CONWIP systems suffer from variability and, unfortunately, the present case suffers from a high variability. In industrial management, if a system displays a high variability it can be buffered by increasing the capacity, the work in process (WIP) or the waiting time. In the hospital environment, increasing the capacity (beds) has a direct cost. An increase in WIP is not feasible because it would correspond to adopting an office-based type of surgery that would exclude inpatients. The last possible way is to increase the total cycle time, which is totally in contrast with the managers' objective.

Another way of improving the system, with no additional costs, is to address efforts directly to the reduction of variability of the waiting times before surgery (WT2), for example by reducing the number of examinations done during the hospitalisation period by increasing the pre-hospitalisation activities. This procedure involves a reorganisation of the healthcare process and can be done by enforcing the pre-hospitalisation process. In fact, carrying out some tests before entering the hospital can reduce the waiting time inside the hospital.

Table 5 Average values of the obtained results for the considered case after re-organisation

Field	*Average value*	*Range*
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 utilisation rate	0.92	(0.89, 0.96)
Busy beds	21.26	(20.56, 22.01)

In order to simulate this scenario, all the patients undergo a pre-hospitalisation period while presently, pre-hospitalisation is assumed only for B, C and D patients. The distribution used to model the waiting time before surgery has to be changed. As the experiment is only simulated, there are not real data to fit with the distribution. The solution was obtained by using the waiting time before surgery of B patients also for A patients. As a consequence, the variance in the waiting time before the first operation (WT2) was reduced. Table 4 reports the results obtained with this simulation, which show a consistent reduction in the patients' waiting time and in the queue length.

Comparisons of the characteristic times and of the key performance indicators of the simulations are graphically provided in Figure 10 and Figure 11, respectively.

Figure 10 Comparison of the characteristic times (see online version for colours)

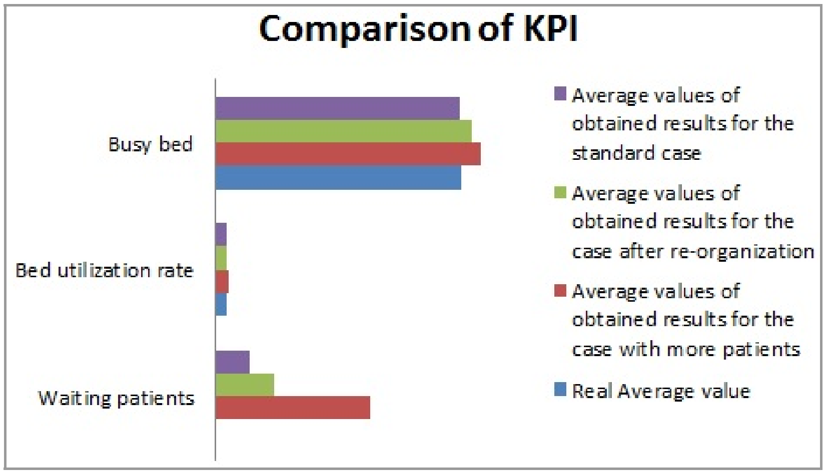
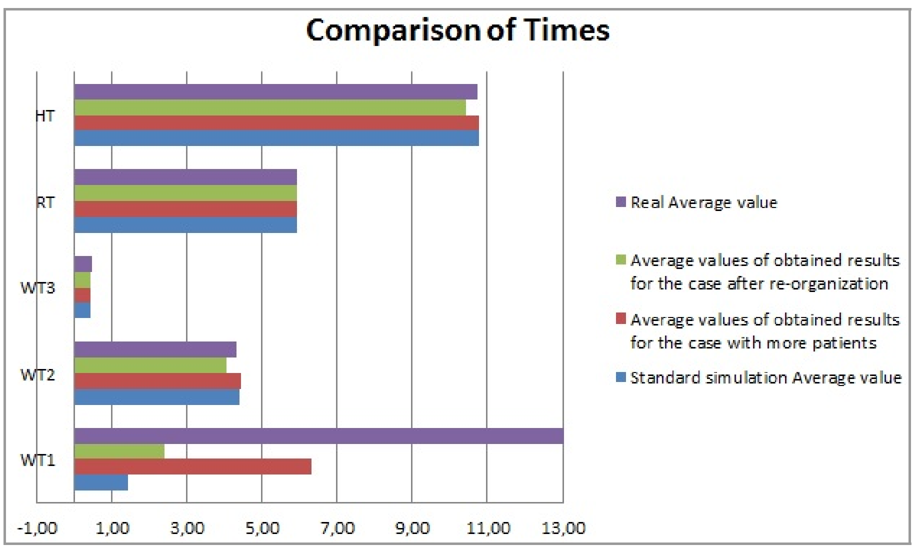


Figure 11 Comparison of the KPIs (see online version for colours)



5 Discussion

An attempt has here been made to analyse patient flows in a hospital through the operations management methods. In this context, a patient corresponds to a job, and his/her flow corresponds to process routing, while the department and the operating theatre correspond to workstations. The stations the patient has to pass through are hospitalisation and surgery. Recovery uses the same resources (beds) as hospitalisation, and as it does not have an associated queue, it can therefore be treated as a non-pre-emptive outage time that has to be added to the hospitalisation time in order to establish the effective process time the patient stays in hospital. Buffers are not allowed in the equivalent production line (patients cannot be hospitalised if no beds are available). The waiting list corresponds to an unlimited raw material inventory. The patient schedule corresponds to a CONWIP system: the release of jobs is based on the system status (available beds).

The advantages of CONWIP are that there is a limited WIP and cycle time compared to the raw process time (the sum of the process times along the line). The disadvantages are a reduced throughput and a transfer of the waiting line from inside to the outside the system. If the hospital behaves like a CONWIP, a number of flaws concerning common management strategies applied in hospitals can be observed.

One of the main management objectives is to maximise the utilisation of the operating theatre. When there is full utilisation of the beds (and beds represent the bottleneck of the system), increasing the utilisation of the operating theatre is impossible because the bottleneck controls the arrival rate. A side effect is the increase in the utilisation of the stations and the consequent increase of the length of the waiting list. The throughput of CONWIP is reduced, with respect to the system capacity, if there is variability. If the variability is very high, the throughput is reduced in an unacceptable manner. Variability in surgery is very large and it can be due above all to natural process variability, which is not modifiable. The only way to improve the system, with no extra cost, is to directly address efforts towards a reduction in variability. In the present situation, variability is mainly due to natural causes – the evolution of a patient's health status, which cannot be reduced directly but can be controlled indirectly.

In our work, we have shown that simulation is an effective decision support tool to test different scenarios of the reorganisation of hospital processes. Furthermore, the significant role of variability for improving the process is shown. Two suggested improvements were experimented. The separation of beds during admission based on the intensity of cure appears to be the best proposal for the department reorganisation.

References

Akkerman, R. and Knip, M. (2004) 'Reallocation of beds to reduce waiting time for cardiac surgery', *Health Care Management Science*, Vol. 7, No. 2, pp.119–126.

Antonelli, D. and Taurino, T. (2010) 'Application of a patient flow model to a surgery department', *IEEE Workshop on Health Care Management*, pp.1–6.

Antonelli, D., Bellomo, D., Bruno, G. and Villa, A. (2012) 'Evaluating collaboration effectiveness of patient-to-doctor interaction in a healthcare territorial network', *Collaborative Networks in the Internet of Services*, pp.128–136.

Antonelli, D., Bruno, G. and Taurino, T. (2014) 'Simulation-based optimization of patient flow in elective surgery', *Proceedings in Mathematics & Statistics*, Vol. 61, pp.87–97, Springer.

Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y.N., Tseytlin, Y. and Yom-Tov, G.B. (2015) 'On patient flow in hospitals: a data-based queueing-science perspective', *Stochastic Systems*, Vol. 5, No. 1, pp.146–194.

Cochran, J.K. and Roche, K. (2008) 'A queuing-based decision support methodology to estimate hospital inpatient bed demand', *J. Oper. Res. Soc.*, Vol. 59, No. 11, pp.1471–1482.

Davies, R. and Davies, H.T.O. (1994) 'Modeling patient flows and resource provision in health systems', *Omega: The International Journal of Management Science*, Vol. 22, No. 2, pp.123–131.

Fowler, M. and Scott, K. (2000) *UML Distilled*, Addison-Wesley, Boston, MA, USA.

Gupta, D. and Denton, B. (2008) 'Appointment scheduling in health care: challenges and opportunities', *IIE Transactions*, Vol. 40, No. 9, pp.800–819.

Haraden, C. and Resar, R. (2004) 'Patient flow in hospitals: understanding and controlling it better', *Frontiers of Health Services Management*, Vol. 20, No. 4, p.3.

Kelton, W.D. et al. (2010) *Simulation with Arena*, McGraw-Hill Professional, Boston, MA, USA.

Kleijnen, J. (2005) 'An overview of the design and analysis of simulation experiments for sensitivity analysis', *European Journal of Operational Research*, Vol. 164, No. 2, pp.287–300.

Koo, P-H., Jang, J., Nielsen, K.B. and Kolker, A. (2010) 'Simulation-based patient flow analysis in an endoscopy unit', *IEEE Workshop on Health Care Management*, pp.1–6.

Lamiri, M., Xie, X., Dolgui, A. and Grimaud, F. (2008) 'A stochastic model for operating room planning with elective and emergency demand for surgery', *Eur. J. Oper. Res.*, Vol. 185, No. 3, pp.1026–1037.

Marshall, A., Vasilakis, C. and El-Darzi, E. (2005) 'Length of stay-based patient flow models: recent developments and future directions', *Health Care Management Science*, Vol. 8, No. 3, pp.213–220.

Meng, F., Qi, J., Zhang, M., Ang, J., Chu, S. and Sim, M. (2015) 'A robust optimization model for managing elective admission in a public hospital', *Operations Research*, Vol. 63, No. 6, pp.1452–1467.

Montgomery, J.B. and Davis, K. (2013) 'The hospital patient flow model: a simulation decision support tool', *Conference Proceedings*.

Pels, H.J. and Goossenaerts, J. (2007) 'A conceptual modeling technique for discrete event simulation of operational processes', *IFIP International Federation for Information Processing*, Vol. 246, pp.305–312.

Pinto, L.R., de Campos, F.C.C., Perpétuo, I.H.O. and Ribeiro, Y.C.N.M.B. (2014) 'Analysis of hospital bed capacity via queuing theory and simulation', *Proceedings of the Winter Simulation Conference 2014*, December, pp.1281–1292, IEEE.

Qi, E., Xu, G., Huo, Y. and Xu, X. (2006) 'Study of hospital management based on hospitalization process improvement', *IEEM*, pp.74–78, IEEE.

Reindl, S., Monch, L., Monch, M. and Scheider, A. (2009) 'Modeling and simulation of cataract surgery processes', *Winter Simulation Conference*, pp.1937–1945.

Schmidt, R., Geisler, S. and Spreckelsen, C. (2013) 'Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources', *BMC Medical Informatics and Decision Making*, Vol. 13, No. 1, p.1.

Smirnov, N.S. (1948) 'Tables for estimating the goodness of fit of empirical distributions', *Annals of Mathematical Statistics*, Vol. 19, No. 2, pp.279–281.

Teilans, A., Kleins, A., Merkuryev, Y. and Grinbergs, A. (2008) 'Design of UML models and their simulation using ARENA', *WSEAS Trans. on Computer Research*, Vol. 3, No. 1, pp.67–73.

Vissers, J. (1998) 'Patient flow-based allocation of inpatient resources: a case study', *European Journal of Operational Research*, Vol. 105, No. 2, pp.356–370.

Wilson, M.J. and Nguyen, K. (2004) *Bursting at the Seams: Improving Patient Flow to Help America's Emergency Departments*.

Xiong, H.H., Zhou, M.C. and Manikopoulos, C.N. (1994) 'Modeling and performance analysis of medical services systems using Petri nets', *IEEE International Conference on Systems, Man and Cybernetics*, pp.2339–2342.

Young, T., Brailsford, S., Connell, C., Davies, R., Harper, P. and Klein, J.H. (2004) 'Using industrial processes to improve patient care', *The British Medical Journal*, Vol. 328, No. 7432, pp.162–164.