Evaluation of the Efficiency of Intensive Care Units Using Queueing Modelling: A Case Study in Kyiv Hospitals

Hanna Livinska and Antonina Pererva

Taras Shevchenko National University of Kyiv, Volodymyrska St, 64/13, Kyiv, 01601, Ukraine

Abstract

This work aims to show queueing modelling as an extremely useful tool for investigating, analyzing and designing real-life systems, especially for the systems that cannot be observed during long time. Given the main properties and characteristics of system operation, queueing modelling and appropriate system simulation provide possibilities to control and operate system performance measures. It allows to determine the optimal number of beds, resources, and personnel, and also to evaluate the quality of patient care. Maintenance system cost optimization problems can be formulated and solved to find the optimal balance between average workload and loss probability. In the modern medical environment, queueing models should be used to analyze and plan the work of hospitals, clinics, and other medical institutions. They would help establish optimal modes of operation, distribute workload, ensure minimum waiting time for patients and maximum efficiency of resource using.

Data from Kyiv hospitals are used to demonstrate possibilities of managing an intensive care unit. Simulation results provide us evaluation of crucial operational characteristics, which are probability of failure of a patient and workload of a ward, of such a unit for different values of the system parameters.

Keywords [1]

Queuing modeling, failure probability, emergency block

1. Introduction

Health care systems are something that everyone deals with, and their effective functioning is extremely important. In any medical process, there is a demand side (patients) and a supply side (hospital resources such as surgeons, nurses, operating rooms, waiting rooms, laboratories, etc.). Both supply and demand are inherently stochastic, so, the need for resources is largely unplanned. As a result, there is a constant mismatch between treatment demand and available capacity. However, timely care is very important. In hospital systems, the waiting time (if there is a queue) to receive attention or the probability of failure (without a queue) are key elements of measuring the quality of service. Thus, the reduction of these elements is an essential factor in the management of these systems. The main objective of this work is to identify the factors that influence patient waiting time or the probability of patient rejection, to point out levers for improvement and to analyze trade-offs.

Analytical tools derived from queueing theory can be used to obtain the above-mentioned properties. Appropriate queueing models allow us to understand the existing relationships between each of the elements of the system. Different health care units can be represented as a queueing system or a queueing network. Analyzing queuing and failure rates can significantly improve medical outcomes, patient satisfaction, and cost-effectiveness of health care. Such modelling is particularly useful to simulate and investigate various scenarios such as mass

Information Technology and Implementation (IT&I-2023), November 20-21, 2023, Kyiv, Ukraine
EMAIL: hanna.livinska@knu.ua (H. Livinska); tont0n@knu.ua (A. Pererva)
ORCID: 0000-0001-9676-7932 (H. Livinska)

epidemics, emergencies, and medical crises. This helps to understand the possible consequences and to develop effective management strategies in case of emergencies. In addition to its relevance during the COVID-19 pandemic and martial law, queueing theory is proving useful in dealing with other emergencies that require quick and efficient problem solving. This theory helps to improve production processes and optimize the distribution of goods during such crises. The first steps in the use of queueing models in medicine can be attributed to the middle of the 20th century, when the theory of queues and methods of mathematical modeling began to develop. During this period, the first mathematical models were created for the analysis and optimization of on-call systems in hospitals and medical institutions ([1], [2], etc.). Later, a number of works appeared that consider various models in healthcare: models with bulking ([1-3], etc.), variable arrival rate ([4], etc.), priority queueing discipline ([5], etc.), blocking ([5], [6], etc.), etc. Simulation ([7], [8], etc.) is one more successful approach for solving healthcare systems management problems. Such simulations helps manage optimal bed capacities, given the data from certain hospitals. A number of modern works are devoted to the construction of models of medical units based on the data of hospitals in different countries ([8-11], etc.).

In general, queueing modelling and appropriate simulation play an important role in healthcare operations, helping to improve efficiency, manage patient flow, and address critical situations by providing analytical tools and information for decision making. However, it should be noted that the use of queueing models in healthcare is not widespread. This work examines the intensive care unit, that is available in most hospitals. It provides intensive care (treatment and observation) for people who are seriously ill or who are in an unstable condition. People in intensive care need constant medical support. There are several problems associated with the intensive care unit: shortage of beds, lack of trained personnel of the intensive care unit, costs. Emergency care is more expensive than other types of health care because of higher needs for personnel, specialized equipment, and therapeutic interventions. The approach based on mathematical queueing modeling and simulation is used. First, we consider how certain health care configuration affect patient care delays and the use of health care resources. When modeling the intensive care unit, we focus on one of the key factors of the system's operation that is the probability of blocking. Second, analyzing system parameters based on data from the intensive care unit of Oleksandrivsky hospital in Kyiv, we can determine the loss probability for the Oleksandrivsky hospital and for all together hospitals in Kyiv in 2021, that was the COVID year.

This paper is organized as follows. In Section 2, we give some ideas of queueing modelling as an analytical tool for systems research and describe a critical care unit in a hospital as an Erlang-Loss system. Information of performance measures for the queueing system are provided in Section 3. In Section 4, we evaluate main performance measures for the model, such as loss probability, optimal number of beds, mean loading of the unit. We show effects of the system parameters on its performance measures. In Section 5, data from Kyiv hospitals are described. In Section 6, simulation of queueing model for Kyiv hospitals is provided, based on data about correspondent patients flows and number of occupied beds. Analyzing system parameters based on the data and the simulation results, we can determine the probability of loss for the Oleksandrivsky hospital and for all together hospitals in Kyiv. Finally, conclusions are given in Section 7.

2. Queueing modelling. A critical care unit as an Erlang-loss system

Society encounters queueing systems every day. In many fields of production, household services, economy and finance, special systems that realize repeated execution of the same type of tasks play an important role. Such systems are called queueing systems. Examples of such systems are banks, various communication systems, loading and unloading complexes (ports, cargo stations), shops, ticket offices, hospitals, anti-aircraft, or anti-missile defense systems, etc.

The founder of queueing theory is the famous Danish scientist A.K. Erlang [12], an employee of the Copenhagen Telephone Company, who was the first to propose using Markov processes with a discrete set of states to describe the processes occurring in queueing systems. Nowadays there are many works devoted to various queueing models in various fields ([13-20], etc.)

To describe the characteristics of a queueing system, it is necessary to determine the probabilistic properties of the incoming flow of customers, service times and service discipline, in particular, the availability of waiting places. The arrival flow of the customers can be characterized by the distribution of the inter-arrival times, and usually the times are assumed to be independent and identically distributed random variables. Let the rate of the input flow is λ. The service times at each server of the system are supposed to be independent random variables, often exponentially distributed (with parameter μ). In the system can be a queue with a finite or infinite waiting places. As for the service discipline, FIFO (first in - first out) is used most often.

For the performance measures of a queueing system, the rate of traffic (traffic intensity) for a server is a crucial characteristic. It is defined as follows:

$$\rho = \textit{mean service time/mean inter-arrival time} = \lambda/\mu$$

One of the main goals of modeling is to determine the performance characteristics of the system which are the probabilistic properties of such random variables as queue length, waiting time, number of customers in the system, loading of capacities (utilization rate of the facilities), etc. For the healthcare unit, it means that we can evaluate the average number of occupied beds, the distribution of the number of occupied beds, the cost of the medical service, the probability of a patient being turned away in case if all beds are occupied, etc. Explicit formulas for stationary probabilities and other performance measures for the most basic types of queueing systems were obtained earlier and are well-known (see, for example, [13-15]). Having an appropriate model, correspondent performance measures for a healthcare unit can be evaluated.

So, we will consider an intensive care unit as a queueing system. It means that beds are servers, patients are customers. There are a finite number of servers. Let us their number be $n = 30$ in a block. Since the waiting time of patients in a queue should be eliminated when providing care of the type in intensive care units, a system without a queue is considered. The flow of patients is supposed to be a Poisson one, because the appropriate conditions of the Poisson process usually fulfilled for it. Since we are considering the flow of emergency patients, then the intensity of patient arrivals is assumed to be constant, because the change in the intensity of arrivals is little monitored depending on, for example, the day of the week or the time of day, in contrast to planned patients who are affected by these factors. Service, that is duration of stay in the intensive care unit, is supposed to be determined by an exponential distribution. The service discipline is FIFO.

So, considering the above assumptions, we get the $[\mathrm{M}|\mathrm{M}|\mathrm{n}|\mathrm{n}]$- queueing system. It is the oldest system described by Erlang, and it is called Erlang-loss system. Formulas for the performance measures for this type of system are well-known. We will use them to explain operating characteristics of a critical care unit in a hospital (loss probability, facilities utilization). Note, that if the real distribution of service times is a bit less or greater than exponential, the $[\mathrm{M}|\mathrm{M}|\mathrm{n}|\mathrm{n}]$- system will still good estimate loss probabilities. However, if it is substantially different, the $[\mathrm{M}|\mathrm{M}|\mathrm{n}|\mathrm{n}]$model may significantly underestimate or overestimate actual loss probabilities. It depends on variance of the distribution. If the variance is lower, the model will overestimate actual loss probabilities, while the converse is true if variance is greater. If the variance of service time is known, the loss probabilities can be calculated for a model with generally distributed service times.

3. Performance measures of the system

3.1. Stationary distribution for the number of customers in the system

Let $N(t), t \geq 0$, be the number of customers in the $[M|M|n|n]$-system. It is a birth-death process with the following infinitesimal rates:

$$\lambda_k = \begin{cases} \lambda, & k < n; \\ 0, & k \geq n; \end{cases}$$

$$\mu_k = k\ \mu, \quad k = 1, 2, \ldots, n$$

Stationary distribution for the number of customers can be calculated by the following formula

$$P_k = P_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} = \frac{\left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}} = \frac{\frac{\rho^k}{k!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!}}, \quad k \leq n, \quad \rho = \frac{\lambda}{\mu}$$

The most important quantity out of obtained values P_k is P_n. This is the blocking probability of the system. The blocking probability is the probability that all *n* servers are busy, so it is the proportion of time that no new customers can enter the system, namely, they are blocked (or lost). It is therefore called time congestion.

$$P_b = P_n = \frac{\frac{\rho^n}{n!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!}} = B(n,\ \rho)$$

is known as Erlang's loss Formula, or Erlang *B*-formula, published first by A.K.Erlang ([12]). Due to the special properties of the Poisson process, in addition of being the proportion of time during which the calls are blocked, $B(n, \rho)$ also gives the proportion of calls blocked due to congestion; namely, it is the call congestion as well.

3.2. Heavy traffic regime

A system where the offered traffic load ρ is greater or equal to the system capacity is called a critically loaded system (a system operating in heavy traffic regime). Accordingly, in a critically loaded Erlang's system we have $n \leq \rho$. It is interesting that if we maintain $n = \rho$ and we increase them both, the blocking probability decreases, the utilization increases, and the product $B(n, \rho)\sqrt{\rho}$ approaches a constant $\tilde{C}$, that does not depend on ρ or n. This implies that in the limit, the blocking probability decays at the rate of $1/\sqrt{n}$. That is, for a critically loaded Erlang's system, we obtain

$$\lim_{n \to \infty} B(n, \rho) = \frac{\tilde{C}}{\sqrt{n}}$$

The low blocking probability in critically loaded large system can be explained. In such a case, the standard deviation of the traffic is very small relative to the mean, so the traffic behaves close to deterministic.

3.3.Performance measures for the system

Here some well-known useful formulas for calculating performance measures for the $[M|M|n|n]$ -system are given.

The average number of customers in the system (the average number of occupied beds)

$$\bar{N} = \rho(1 - P_n)$$

and utilization (utilization rate) of a server

$$U_s = \frac{\bar{N}}{n}$$

The mean idle period of a server

$$\bar{e} = \frac{n}{\lambda(1-P_n)} - \frac{1}{\mu}$$

Average working time of the system (busy period)

$$\mathrm{E}\delta_r = \frac{1-P_0}{P_0}$$

4. Effect of system characteristics on the operation of a critical care unit

The work of the intensive care unit can be affected by various factors, in particular, a change of the intensity of the arrival of patients, which can be associated with an increase or decrease in the population in the area where the hospital is located, with natural disasters, unpredictable outbreaks of diseases (for example, like COVID-19), by war, etc. The duration of stay of a patient in the unit is also affected, for example, if the rate of the input flow is very high, this period can be reduced to the minimum possible for individual patients, or if the workload of the unit is low, the duration of treatment can be extended in order to properly care for the patient in order to prevent relapses. And, of course, the functioning of the intensive care unit is affected by the number of beds that can be used by patients and the number of medical staff.

Here we show, how different values of system parameters affect the operation of the system.

4.1. The probability that *k* beds are occupied

First, let us consider how the different number of available beds will affect the workload of the unit, while the rate of input flow and rate of service remain the same.

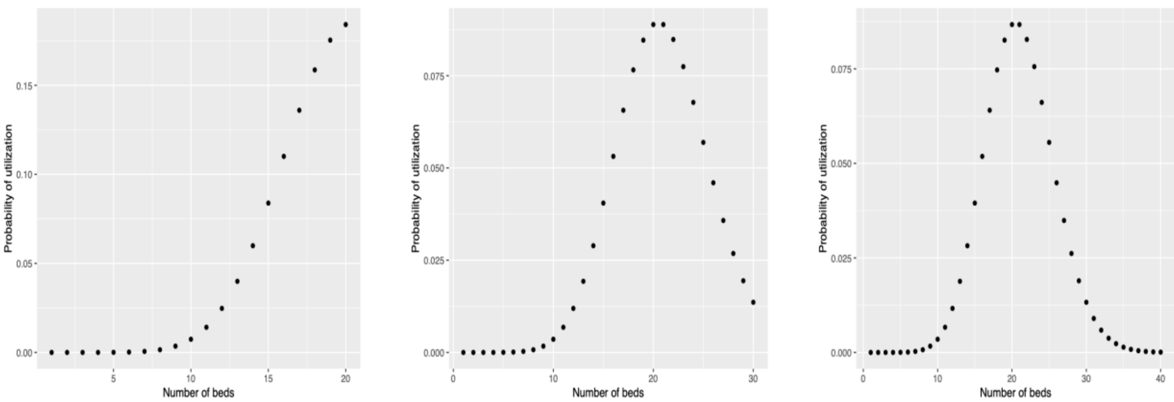

Figure 1: The probability that *k* beds are occupied for units with 20, 30 and 40 beds, $\lambda = 3$, $\mu = 1/7$.

Figure 1 shows the graph of the probability of occupied k beds, when their maximum number is 40, 30 and 20 beds. We can see that the smaller the number of beds, the more the graph is shifted to the right, which means that the probability of the maximum number of occupied beds increases, and therefore the load on the system increases.

4.2. Effect of the number of beds on the loss probability

The number of beds in the intensive care unit is closely related to the probability of patient rejection, if all beds are occupied. Let us analyze this dependence by constructing the plot shown in Figure 2, where the number of beds varies from 20 to 40, and the rate of arrivals and rate of service are the same as in the initial version of the unit simulation. It is obvious that increasing the number of beds will reduce the probability of failure, but it should be remembered that the utilization of the ward will also decrease to too low a value, so the task of the head of the unit is to find a compromise that will be beneficial for their situation.

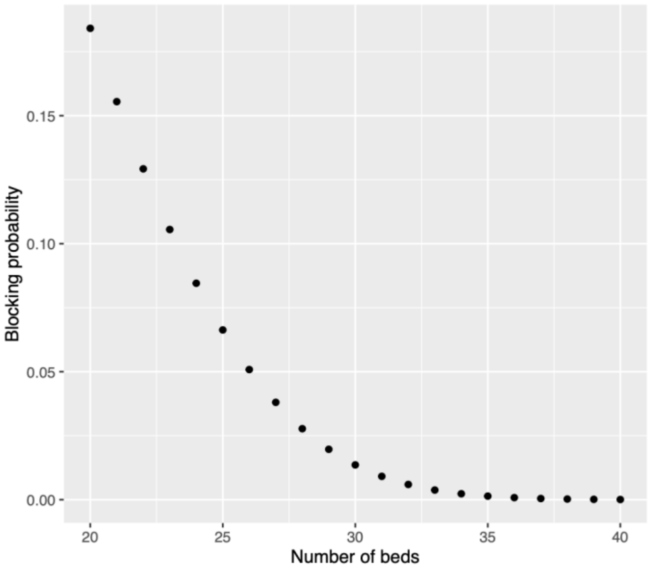

Figure 2: Effect of the number of beds on the loss probability, $\lambda = 3$, $\mu = 1/7$.

So, with 30 beds in the ward, the probability of failure is equal to 0.014, Figure 2 shows that in order to this probability to be less than 0.001, the number of beds should be equal to 35, that is, it is necessary to expand the modelled unit by 5 beds. With the available 38 places the loss probability is almost zero, which means accepting for treatment all patients who arrive in the intensive care unit. If for some reason the hospital is forced to reduce the number of beds in the intensive care unit. This will lead to an increase in the probability of failure, which will have a bad effect on the quality of service. For example, if the number of beds is reduced to 24, the rate of failure will be about 0.09, which is unacceptable, because a large percentage of patients will not be able to receive the immediate necessary care, which will lead to a deterioration in their health or even death.

4.3. Effect of the rate of arrivals and rate of service on the unit occupancy

We see a similar situation as when changing the number of beds: the greater the rate of arrivals, the more the dynamics of the probability that k patients are in the unit is shifted to the right. When one person per day is admitted to the intensive care unit, its workload will be much lower than when the intensity of admission is equal to 5.

At different values of λ and μ, the occupancy of the unit changes. To analyze this situation, let us build a 3D plot (Figure 5), and write the digital data into Table 1. We can observe a natural dynamic: when the rate of arrivals increases, the workload of the unit increases, similarly, when the length of stay in the unit of patients increases. It should be noted that the increase in the load factor with an increase in the arrival rate occurs faster than with an increase in the length of stay.

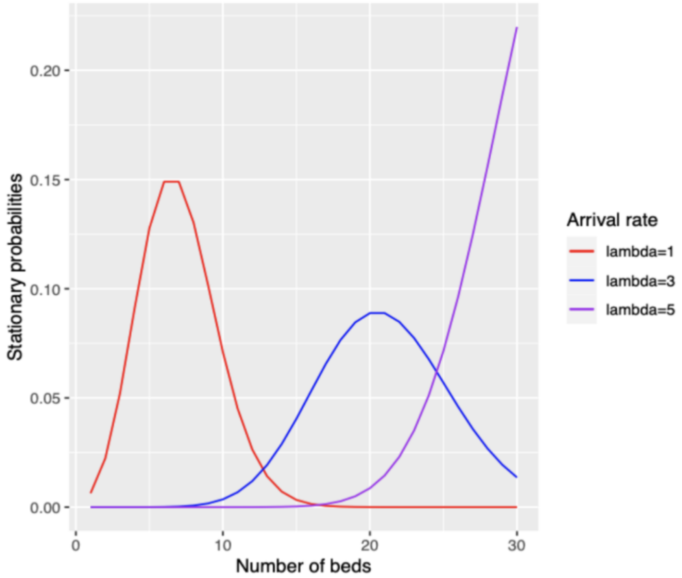

Figure 3: Effect of the patient flow rate on the probability of k patients in the unit (*n*=30, $\lambda = 1; 3; 5$, $\mu = 1/7$, $\rho = 7;\ 21;\ 35$)

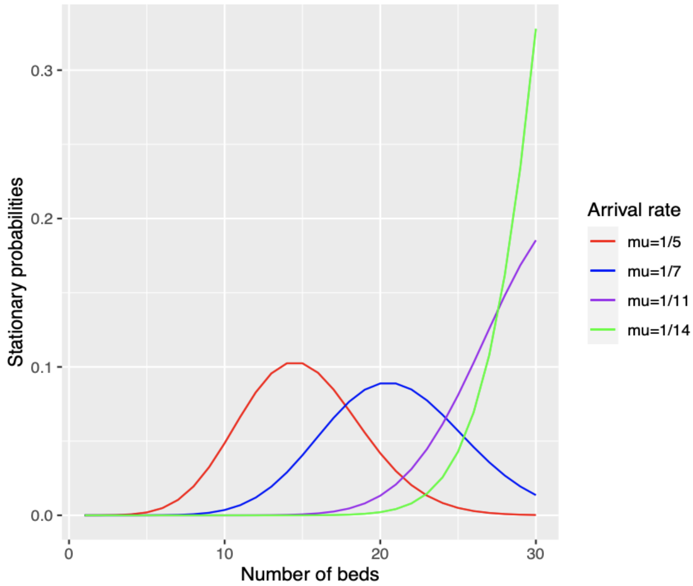

Figure 4: Effect of the service rate on the probability of k patients in the unit (*n*=30, $\lambda = 3$, $\mu = 1/5; 1/7;\ 1/11;\ 1/14$, $\rho = 15;\ 21;\ 33;\ 42$)

Table 1 shows the occupancy rate of the unit in the initial version of the simulation, i.e. with the arrival rate equal to 3 people per day and an average length of stay of 7 days. We can see that if the arrival rate increases by one person, and the average length of stay in the unit increases by one day, then the workload of the ward will increase by about 18%. And if $\lambda = 5$, $\mu = 1/9$, (the average length of stay in the ward is 9 days), then the occupancy rate is about 95%, which means the constant work of the intensive care unit almost at the peak of its capabilities, and there is a high probability of patient rejection. We can also see that when $\lambda = 1$ and $\mu = 1/5$, the occupancy is 16.6% which is a bad value, because most of the beds in the unit are not used at all, and therefore the funds allocated for their maintenance are wasted.

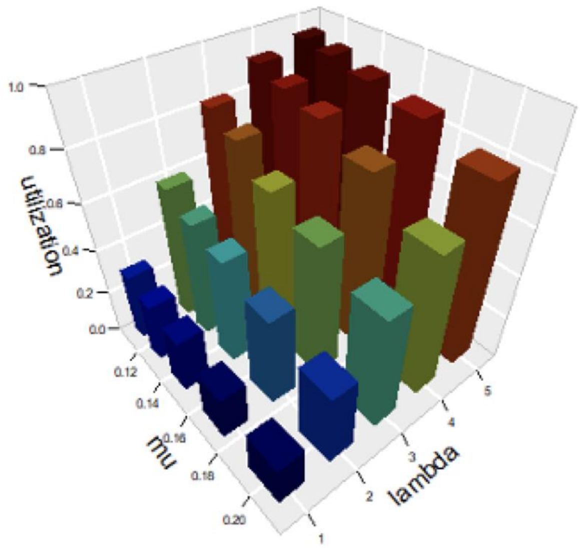

Figure 5: Effect of the rate of arrivals and service on unit occupancy

Table 1
Effect of the rate of arrivals and service on a bed utilization

μ | λ	1	2	3	4	5
1/5	0.167	0.333	0.5	0.661	0.789
1/6	0.201	0.402	0.601	0.77	0.869
1/7	0.233	0.466	**0.69**	0.841	0.91
1/8	0.267	0.533	0.768	0.888	0.934
1/9	0.3	0.599	0.827	0.916	0.949
1/10	0.333	0.661	0.868	0.934	0.959
1/11	0.37	0.724	0.899	0.948	0.966

4.4. Effect of the rate of arrivals and rate of service on loss probability

Similar dynamics we observe (Fig. 6, Table 2) when watch dependence of loss probability on the rate of arrivals and rate of service.

Table 2
Effect of the rate of arrivals and service on a bed utilization

μ | λ	1	2	3	4	5
1/5	0	0	0	0.00846	0.0526
1/6	0	0	0.00275	0.0413	0.135
1/7	0	0	**0.0135**	0.0977	0.219
1/8	0	0	0.0401	0.168	0.299
1/9	0	0.00265	0.0823	0.237	0.368
1/10	0	0.00846	0.132	0.299	0.425
1/11	0	0.0223	0.191	0.36	0.478

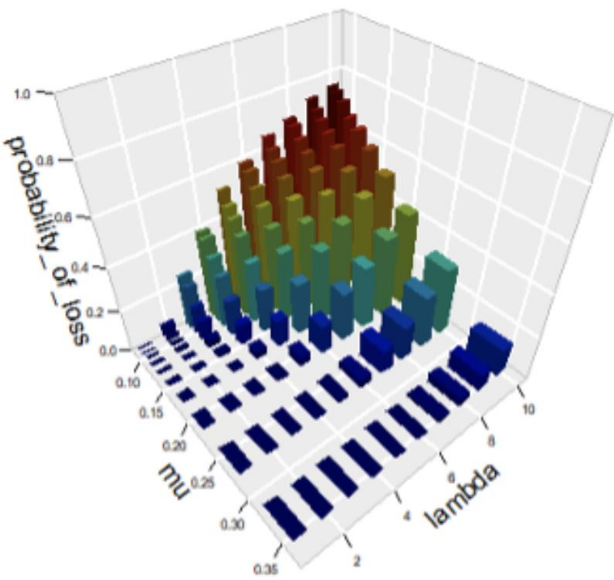

Figure 6: Effect of the rate of arrivals and service on loss probability

5. Data from hospitals in Kyiv

The comparison of the efficiency of intensive care units is performed on the basis of data about COVID emergency units taken from the report of Kyiv City Information analytical center of medical statistics ([21]) and from the Official portal of Kyiv (Kyiv City State Administration), Hospitals and medicine ([22]). The specified data provide characteristics of the work of Kyiv hospitals in 2021. Namely, in average for Kyiv hospitals, we have the following parameter values: the average number of patients per day $\lambda = 548$ and the average number of days a patient stays in the intensive care unit is $1/\mu = 11$. For the Oleksandrivsky Hospital in Kyiv, the characteristics are as follows: the average number of patients per day is $\lambda = 90$, and the average number of days a patient stays in the intensive care unit is $1/\mu = 7$. In Table 3 we include an additional row with approximate rate of arrivals in case if the units are not connected. Such a situation can be observed if the general input flow is divided, for instance, in many small cities. For Kyiv this situation, obviously, is not appropriate. But we show this rate for demonstration of real input flow that can be observed in a city with one intensive care unit.

Table 3
Data from hospitals in Kyiv

	All Hospitals of Kyiv	Oleksandrivsky Hospital
Average number of days a patient stays in the intensive care unit	11	7
Average number of patients per day	548	90
Number of beds in the intensive care units	5721	650
System utilization (according to the model)	0.997	0.955
Loss Probability (according to the model)	0.0628	0.014
Approx. rate for a block with 30 beds (NOT for a big unit)	2.87	4.15

6. Simulation results

The R simmer package is used to simulate the operation of the intensive care unit, which allows to create trajectories and simulate the operation of the system during a certain period.

After performing the simulation for 365 days, the obtained results are presented in Table 4. The data from the table can be used for further analysis and comparison of the performance of the Oleksandrivsky Hospital with the average Kyiv indicators. This allows to provide detailed analysis and comparison of the effectiveness of intensive care units in different medical institutions, which is important for improving medical care, ensuring patient satisfaction, and optimizing the use of resources.

In Table 4, the probability of patient failure and system load are shown. The probability of refusal for Kyiv hospitals is four times greater than the value of Oleksandrivsky hospital. Given that patients entering the intensive care unit require immediate and vital care, this figure is quite high. In further research, this indicator will be brought closer to the corresponding value in the Oleksandrivsky Hospital.

Table 4
Data from the simulation for 365 days

	All Hospitals of Kyiv	Oleksandrivsky Hospital
General number of patients	184147	31847
Number of patients served	174071	31429
Loss probability (estimated)	0.0518	0.0131
Average number of occupied beds	5563	606
Average duration of stay in the unit	10.9	6.7
System utilization	0.9492	0.9351

6.1. Resource usage

We build the evolution of the average number of patients in a unit during one year. Fig. 7 shows the system load during this time.

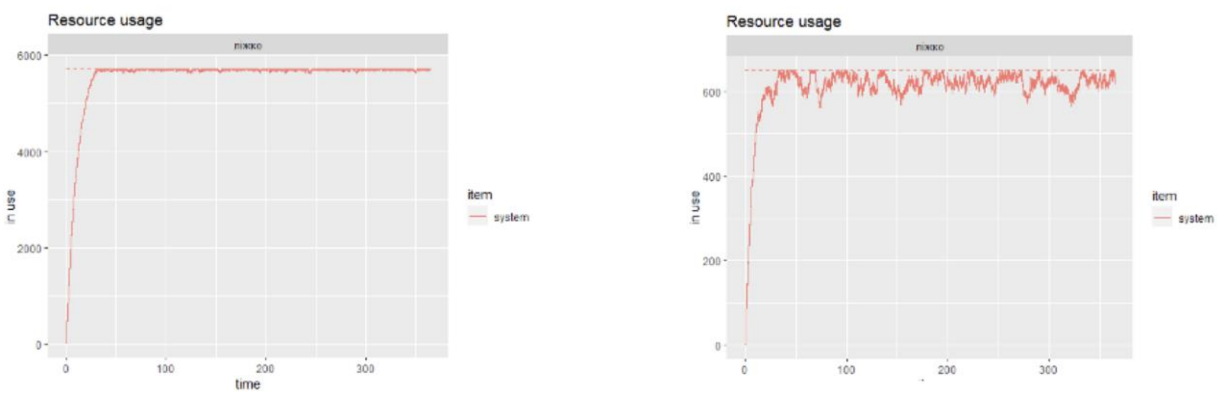

Figure 7: Resource usage in all Kyiv hospitals (on the left) and in Oleksandrivsky Hospital (on the right)

Initially, there were no occupied beds in the ICU model. If we take into account the first 20-40 days of work, then this period does not give a reliable idea of the functioning of the system as a whole. However, after this period, the work stabilizes, and it can be observed that the number of patients in the unit approaches the average value. This means that the system starts working in a stable (stationary) regime.

Let us consider how changing the basic parameters will affect the system load and the probability of failure, bringing the value of the probability of failure in Kyiv hospitals closer to the corresponding value in Oleksandrivsky hospital.

6.2. Effect of changing the number of beds

One of the factors affecting the change in the number of beds may be technological or medical changes. The introduction of new diagnostic or treatment methods may require specialized beds or equipment, which may change the total number of beds in the unit. In addition, changing medical standards or protocols may also lead to a revision of the number of beds in order to take into account new requirements and recommendations.

The relationship between the number of beds in the intensive care unit and the probability of patient failure can be analyzed using the graph shown in Figure 8. The range of the number of beds in the graph is from 5721 to 6000.

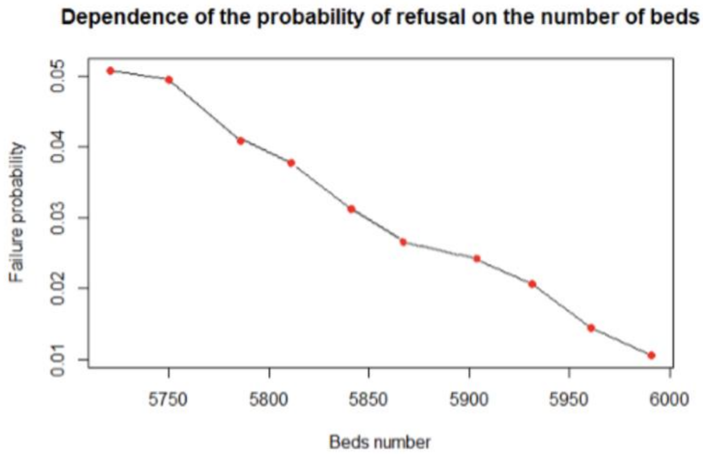

Figure 8: Effects of changing the number of beds

Naturally, with the increase in the number of beds, the probability of patient refusal will decrease. However, this also reduces the occupancy of the ward to a very low level, which leads to idle beds, which increases the hospital's costs for their maintenance. Maybe, the number of beds 5954 is close to the optimal level. It is 233 more than the initial value. In this case, the loss probability is 0.013. Only 1.3% of patients admitted to the intensive care unit are denied the necessary medical care. The low level of the probability of refusal indicates the efficient functioning of the unit and the ability to meet the needs of patients at the desired level of Oleksandrivsky Hospital. The number of patients who received adequate care increased by 6,294 patients, and the average number of occupied beds is 96% of the total number. The corresponding results and their comparison are shown in Table 5.

Table 5

Effects of changing the number of beds

Number of beds	Number of patients	Number of treated patients	Loss probability	Average number of occupied beds	Unit utilization
5721	184147	174071	0.0518	5563	0.9492
5954	190942	190609	0.0135	5747	0.9086

System load decreased to 90% from 94%. The solution to the optimal situation is to find a compromise between workload and the probability of patient rejection. Ward occupancy that is too low can lead to significant bed idleness, leading to suboptimal use of resources and excessive costs. On the other hand, too high a load leads to an increase in the probability of patient refusal, which has negative consequences.

6.3. Changes of the patient arrivals rate

The rate of patients input flow can vary due to various factors that affect the need for emergency medical care. One of the factors that can affect the intensity of the arrival of patients is the epidemiological situation or a war. In the event of an outbreak of an infectious disease epidemic, for example, there may be a significant increase in patients requiring emergency medical care. This may create a temporary peak in the patient admissions to the ICU, requiring adequate response and resources to meet the increased demand.

To reduce the probability of patient rejection, it is important to reduce the number of patients arriving the intensive care unit during the day. However, in some situations, this may not be a realistic option due to the presence of epidemiological outbreaks, military conflicts, or other crisis situations, when the number of patients increases dramatically. The dependence of the probability of refusal on the number of patients per day is shown in Figure 9, where the number of patients varies from 498 to 548 with a step of 10.

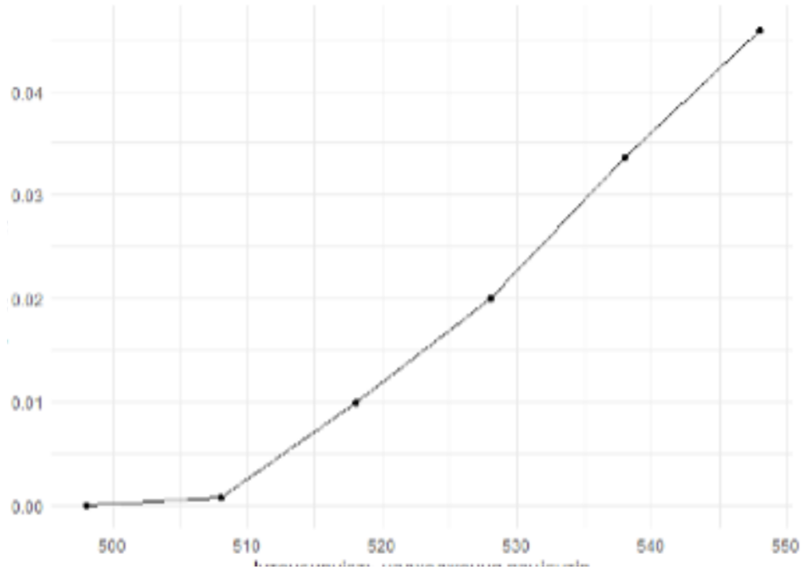

Figure 9: Effect of changes of the patient arrivals rate on the loss probability

The good number of patients is 518 people per day, then the loss probability is 0.008. The number of patients served increased by 7,752 patients.

Table 6
Effect of changes of the patient arrivals rate

Number of patients per day	General number of patients	Number of treated patients	Loss probability	Average number of occupied beds	Unit utilization
548	184147	174071	0.0518	5563	0.9492
510	183277	181823	0.0080	5487	0.9077

As we saw before, the duration of treatment has a smaller effect on the probability of failure compared to the intensity of patient arrival. It may be necessary to consider measures to reduce the number of patients, for example by improving the processes of transfer or treatment of patients.

6.4. Changes of the service time of one patient

Table 7 shows that after changing the service time of one patient from 11 days to 10, the loss probability became 0.0023.

Table 7
Effect of changes of the patient arrivals rate

Duration of bed occupancy	General number of patients	Number of treated patients	Loss probability	Average number of occupied beds	Unit utilization
11	184147	174071	0.0518	5563	0.9492
10	194179	193734	0.0023	5299	0.8477

6.5. Dependence of loss probability on λ and μ

In the Table 8 we can see joint effect of λ and μ on the loss probability.

Table 8
Dependence of loss probability on λ and μ

μ | λ	508	518	528	538	548
1/11	0.0006	0.0079	0.0197	0.0365	**0.0518**
1/12	0.0633	0.0771	0.0952	0.1087	0.1306
1/13	0.1279	0.1424	0.1622	0.1778	0.1889
1/14	0.1934	0.2065	0.2212	0.2384	0.2537
1/15	0.2470	0.2569	0.2688	0.2873	0.2973

7. Conclusions

In this work, we show the advantages of the tools of queueing theory for real-life systems modeling, which allows us to evaluate model parameters and to see how changes of some characteristics will affect others. This mathematical method is important for evaluating of intensive care unit parameters, because in real life any mistake can rise additional risks for patients in hospital. Therefore, the main task to be solved by the heads of hospitals and units is the distribution of resources in such a way as to find a balance between the average workload and the probability of failure.

In order to understand the capabilities of the intensive care unit, we assumed different situations and options for system parameters and looked at how they would affect the whole system. One of the most important characteristics is the probability of patient refusal. We investigate how it can be reduced, or under what circumstances this probability will increase. The most obvious way to reduce the probability of failure is to increase the number of beds. It is also obvious that the probability of refusal is influenced by the intensity of patients' arrival and their average length of stay in the intensive care unit. But it should be noted that the loss probability grows faster with an increase in the intensity of arrival than with an increase in the average length of stay. It works similarly with a decrease in these indicators.

An equally important characteristic of the considered system is the average occupancy of the unit. We saw that the change in the intensity of admission and the average length of stay in the unit have almost the same effect on the system load as on the loss probability. Increasing the number of beds can lead to both positive and negative consequences, because under high load, a small number of additional beds can relieve the system. But if the average load is not at a high enough level, and the number of beds is increased, for example, to reduce the probability of failure, the load is very small, resulting in most beds being idle, which is not cost-effective.

The obtained results make it possible to formulate and solve a number of optimization problems of minimizing the costs of ensuring the current service of the wards, including personnel, and minimizing the risks associated with the refusal of service to patients, possibly due to a certain logistical structure of transfers of urgent patients to other structural units.

We have demonstrated that the intensity of patient arrivals and the length of their stay in the unit are important factors that affect system load. With the increase in the intensity of the arrival of patients and the duration of their stay in the unit, the load on the system increases. This means that more patients arrive and stay in the unit at the same time. A particularly noticeable increase in workload is observed with an increase in the intensity of patient arrivals. The increase in the load factor occurs faster with an increase in the arrival rate than with an increase in the length of their stay.

Data from Kyiv hospitals are used to demonstrate possibilities of managing an intensive care unit. Simulation results provide us evaluation of crucial operational characteristics of such a unit for different values of the system parameters.

So, when designing and managing an intensive care unit, it is necessary to take into account the intensity of patient arrivals and the length of their stay in order to ensure optimal use of resources and provide adequate medical care to patients. It is important to find a balance between these two factors to ensure efficient operation of the unit and minimize the probability of patient rejection. Queueing modelling and simulation are very efficient tools for this.

8. Acknowledgements

Authors would like to thank the anonymous referees for their valuable suggestions.

9. References

[1] N. T. J. Bailey, Queueing for medical care, Journal of the Royal Statistical Society 3 (1954) 137–45. doi:10.2307/2985372.

[2] F. Downton, On limiting distribution arising in bulk service queues, Journal of the Royal Statistical Society 18 (1956) 265–274.

[3] M. J. Jacob, A. Krishnamoorthy, T. P. Madhusoodanan, Transient solution of a finite-capacity $m|ga,\ b|1$ queueing system, Naval Research Logistics 35 (1988) 437–441. doi: 10.1002/1520-6750(198806)35:33.0.CO;2-W.

[4] C. J. Rosenquist, Queueing analisys: a useful planning and management technique for radiology, Journal of medical systems 11 (1987) 413–419. doi: 10.1007/BF00993008.

[5] L. Green, Patient Flow: Reducing Delay in Healthcare Delivery, International Series in Operations Research Management Science, Springer, Boston, MA., 2006, pp. 281–307. doi: 10.1007/978-0-387-33636-7_10.

[6] I. W. Kabak, Blocking and delays in m(n)|m|c bulk queueing systems, Operations research 16 (1968) 830–840.
[7] S. Kim, I. Horowitz, K. K. Young, T. A. Buckley, Flexible bed allocation and performance in the intensive care unit, Journal of operations management 18 (2000) 427–443. doi: 10.1016/S0272-6963(00)00027-9
[8] I. Komenda, Modelling critical care unit activities through queueing theory, 2013. URL: https://api.semanticscholar.org/CorpusID:70621129
[9] G. B. Yom-Tov, Queues in hospitals: stochastic networks with ReEntering customers in the QED Regime, Ph.D. thesis, Israel Institute of Technology, Tamuz 5770, Haifa, 2010.
[10] R. K. Palvannan, K. L. Teow, Queueing for healthcare, Journal of Medical Systems 36 (2012) 541–547. doi: 10.1007/s10916-010-9499-7.
[11] S. Creemers, M. Lambrecht, Modeling a healthcare system as a queueing network: the case of a Belgian hospital, Research Report 0710, Department of decision sciences and information management, Research center for operations management, KU Leuven, 2007.
[12] A. Erlang, Solution of some problems of the theory of probabilities of significance in automatic telephone exchanges, The post office electrical engineers' journal (1918) 189–197.
[13] L. Kleinrock, Queueing Systems, volume 1, John Wiley Sons, Brisbane, Toronto, 1975.
[14] M. Zukerman, Introduction to queueing theory and stochastic teletraffic models, 2023. arXiv:1307.2968.
[15] J. Sztrik, Basic Queueing Theory, volume 1, University of Debrecen, 2021.
[16] H. Livinska, E. Lebedev, On transient and stationary regimes for multi-channel networks with periodic inputs, Applied Statistics and Computation (2018) 13–23. doi:10.1016/j.amc.2017.01.017.
[17] E. Lebedev, H. Livinska, Queueing Theory. Advanced Trends, volume 1 of Mathematics and Statistics, Sciences, ISTE J. Wiley, London, 2020, pp. 177–212.
[18] E. A. Lebedev, V. D. Ponomarov, O. V. Pryshchepa, On multiserver state-dependent retrial queues operating in stationary regime, in: A. Dudin, A. Nazarov, A. Moiseev (Eds.), Information Technologies and Mathematical Modelling. Queueing Theory and Applications, Springer International Publishing, Cham, 2021, pp. 95–107.
[19] I. Makushenko, I. Usar, H. Livinska, M. Sharapov, Optimal threshold strategies for retrial systems with queue, Journal of Computational and Applied Mathematics 427 (2023) 115136. doi:10.1016/j.cam.2023.115136.
[20] M. Katsman, V. Matsyuk, V. Myronenko, Modeling the reliability of transport under extreme conditions of operationas a queuing system with priorities, Control, Navigation and Communication Systems (2023) 10–17. In Ukrainian.
[21] Kyiv City Information analytical center of medical statistics, URL: https://medstat.kiev.ua/statistichni-zviti/
[22] Official portal of Kyiv (Kyiv City State Administration), Hospitals and medicine, URL: https://kyivcity.gov.ua/likarni_ta_medytsyna/zdatni_zupiniti_400608/zapovnenist_opornikh_likaren_persho_khvili_khvorimi_na_Covid-19/