SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2022
ORGANIZATION AND MANAGEMENT SERIES NO. 167

A QUEUEING MODEL OF PATIENT SERVICE IN ACCORDANCE WITH THE S.T.A.R.T STANDARD

Wojciech M. KEMPA[1*], Arkadiusz BANASIK[2]

[1] Silesian University of Technology, Faculty of Applied Mathematics, Department of Mathematics Applications and Methods for Artificial Intelligence; wojciech.kempa@polsl.pl, ORCID: 0000-0001-9476-2070
[2] Silesian University of Technology, Faculty of Applied Mathematics, Department of Mathematics Applications and Methods for Artificial Intelligence; arkadiusz.banasik@polsl.pl, ORCID: 0000-0002-4267-2783
* Correspondence author

Purpose: Currently, in Poland, the process of patient registration at the hospital emergency department (SOR) is carried out in accordance with the S.T.A.R.T. (*Simple Triage And Rapid Treatment*) standard. According to this procedure, based on a preliminary interview, patients are selected and assigned to one of five risk categories, described by colors: red, orange, yellow, green, or blue. Since this standard is relatively new (introduced in 2019), its analysis using quantitative, not only qualitative, methods are highly desired.
Design/methodology/approach: The so-called mean-value approach for studying the proposed queueing model is used in the paper.
Findings: For each category of patients, the mean number of patients present in the system, the mean queue length, and the average values of waiting time for admission by a doctor and time spent in the emergency department are found. Two priorities are considered: non-preemptive (absolute) and preemptive-resume (relative). Numerical calculations illustrate theoretical results.
Originality/value: Explicit representations for key queueing characteristics are found analytically. Moreover, illustrating numerical results simulating the "behavior" of a real hospital emergency department are presented and discussed.

Keywords: hospital emergency department, priority service, quality of service (QoS); queueing model, S.T.A.R.T. standard.

Category of the paper: research paper.

1. Introduction

Hospital emergency departments (SOR acronym commonly used in Poland), by definition, should meet the highest quality standards for patient care. They should also be perfectly managed to specialist medical equipment, patient service time, and human resources (medical personnel). Currently, in Poland, the process of patient registration at the hospital emergency department is carried out in accordance with the S.T.A.R.T. (*simple triage and rapid treatment*) standard. The algorithm was initiated by the Newport Beach Fire and Marine Department and Hoag Hospital in Newport Beach (California, U.S.A.) in 1983. According to the S.T.A.R.T. procedure, based on a preliminary interview, patients are selected and assigned to one of five risk categories, described by colors: red, orange, yellow, green, or blue. Different modifications of the original approach were proposed later. For example, in 1996, the S.T.A.R.T.-S.A.V.E. (*Secondary Assessment of Victim Endpoint*) standard was worked out by Benson, Koenig, and Schultz (Benson et al., 1996), in which some additional factors determining "survivability" of a patient over time were implemented. In the literature, one can find many studies and comparisons regarding the original algorithm and its modifications to various aspects of patient service in hospital emergency departments, the quality of this service, and the possibility of its improvement (see, e.g., Baker, 2007; Kahn et al., 2009; Badiali et al., 2017; Fink et al., 2018; Ferrandini et al., 2018).

Queueing systems are commonly used in modeling different-type real-life phenomena. In particular, they are proposed for solving management and quality-of-service problems occurring in production engineering, computer and telecommunication networks, transport and logistics, medical sciences, and health care procedures (see, e.g., Bose, 2002; Ng, Soong, 2008; Chan, 2014; Shortle et al., 2018; Lakatos et al., 2019). By constructing an appropriate queueing model and determining its stochastic characteristics, it is possible to answer many questions regarding the quality of service and the level of use of the service station capabilities.

In the article, a queueing system with many classes of customers and priority service is proposed for modeling the process of qualifying and serving patients. The stream of incoming patients is described using the Poisson process, while the time of processing of a single patient is assumed to be hyper-exponentially distributed. For each category of patients, the mean number of patients present in the system, the mean queue length, and the average values of waiting time for admission by a doctor and time spent in the emergency department are found. Two priorities are considered: non-preemptive (absolute) and preemptive-resume (relative). Numerical calculations illustrate theoretical results.

2. Description of a queueing model

Arrival and service processes

Let us assume that patients arrive at the emergency department according to the Poisson process with a certain constant intensity λ. The Palm-Khintchine theorem justifies this assumption (see, e.g., Heyman, Sobel, 1982), according to which the superposition of many "rare" and independent streams of events (namely, here, streams of patients coming from different "directions") can be well approximated by the Poisson event stream. After arriving patients are registered, a general interview is carried out, and on its basis, an initial assessment of their health condition and related risks is performed. The consequence of this procedure is the selection of patients and their division into five different classes marked with bands of different colors: red (R), orange (O), yellow (Y), green (G), or blue (B). The appropriate priority of patient service is associated with a specific class (color) determined by the patient's health condition. We will assume that patients marked in red have priority in handling (e.g., when performing laboratory or radiological tests) over all other patients, patients in orange "give way" to only "red" patients, etc. The principles of giving priority to handling specific patients are graphically presented in the following diagram:

RED > ORANGE > YELLOW > GREEN > BLUE

Figure 1. Priority rule in the considered queueing model.

Assume that q_j is the fraction of arriving patients qualified as j-type patients, where $j = R, O, Y, G, B$. Relating to splitting property of a Poisson process (Tijms, 2003), j-type patients arrive at the emergency department according to a Poisson process with intensity $\lambda_j \stackrel{\text{def}}{=} q_j \lambda$ $(j = R, O, Y, G, B)$. Within a given class, patients are served in accordance with the FIFO (*First In First Out*) discipline, i.e., in the order in which they appeared in the emergency department. The queue of waiting patients is unlimited in advance. The service time for a single patient is closely related to her/his health condition. It is not necessarily the longest in the case of the highest priority patients (such patients are often referred directly to a specific hospital ward, e.g., for urgent surgery). In general, therefore, we will assume that the service time for a single j-type patient has a hyper-exponential distribution with the following probability density function (PDF):

$$f_j(t) \stackrel{\text{def}}{=} \sum_{i=1}^{k_j} p_i^{(j)} exp\left(-\mu_i^{(j)} t\right), t > 0 \tag{1}$$

where $j = R, O, Y, G, B$. Values k_R, k_O, k_Y, k_G, and k_B denote numbers of possible complete diagnostic paths for a j-type patient. In practice, these numbers can be the same for all patients. However, it is not a rule. The value $\frac{1}{\mu_i^{(j)}}$ is the mean duration of the ith complete diagnostic path offered for the j-type patient ($i = 1, \dots, k_j$).

Let us note that the occupation rate ρ_j of the system relating to j-type patient (where $j = R, O, Y, G, B$) equals

$$\rho_j = \lambda_j \cdot E(B_j) = \lambda_j \cdot \sum_{i=1}^{k_j} \frac{p_i^{(j)}}{\mu_i^{(j)}} \tag{2}$$

where B_j stands for the service time of a j-type patient. To ensure the existence of the stationary state of the system, let us assume that

$$\rho \overset{\text{def}}{=} \rho_R + \rho_O + \rho_Y + \rho_G + \rho_B < 1 \quad (3)$$

Priority rules

In the studied queueing model of patient service at the hospital emergency department, we will consider the following two models of the priority rule (discipline):

- *absolute priority* (preemptive-resume priority) in which an arriving higher-priority patient interrupts the ongoing service of a lower priority patient,
- *relative priority* (non-preemptive priority) in which the ongoing service process of a lower priority patient cannot be interrupted by a higher priority patient arriving at the hospital emergency department.

For each priority rule, we will analyze the four main stochastic characteristics (performance measures) affecting the quality of service (QoS) for $j = R, O, Y, G, B,$ namely:

- the mean number of j-type patients present in the system (we will denote it by $E(L_j)$),
- the mean number of j-type patients waiting in a queue (denoted by $E(Q_j)$),
- the mean sojourn time of a j-type patient (denoted by $E(S_j)$),
- the mean waiting time (for beginning the service) for a j-type patient (denoted by $E(W_j)$).

Moreover, in dependence on the type of a priority rule, we will use the superscript abs or rel in relation to the absolute or relative priority, respectively. So, for example, $E(S_R^{rel})$ stands for the mean sojourn time of a red-type patient in the case of the relative priority rule.

3. Analytic formulae for performance measures

Since the service time of a single patient has not got the memoryless property (is not exponential), we will introduce the so-called *residual service time*, which is the time needed to complete the ongoing (at the arrival moment of a patient) service. In general (see, e.g., Adan, Resing, 2015), it is possible to express the mean residual service time by using the first two moments of the service time. Indeed, we have

$$E(R_j) = \frac{E(B_j^2)}{2E(B_j)} \quad (4)$$

where $E(R_j)$ and $E(B_j^2)$ are the mean residual service time of the j-type patient and the second moment of its service time ($j = R, O, Y, G, B$), respectively. Adjusting the formula (4) to the hyper-exponential service times (see (1)), we have

$$E(R_j) = \frac{\sum_{i=1}^{k_j} \frac{p_i^{(j)}}{\left[\mu_i^{(j)}\right]^2}}{2\sum_{i=1}^{k_j} \frac{p_i^{(j)}}{\mu_i^{(j)}}} \tag{5}$$

The representations for mean values of the number of patients (of each type) present in the considered system, number of patients waiting in the queue, waiting, and sojourn times are essentially dependent on mean residual service times, occupation rates, and average service times and arrival rates. In the case of relative priority, the following formulae are accurate (see Adan, Resing, 2015) – we adjust the formulae to the notation introduced in the paper:

$$E(S_j^{rel}) = \frac{\sum_{i\in\{R,O,Y,G,B\}} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} + E(B_j) \tag{6}$$

and

$$E(W_j^{rel}) = \frac{\sum_{i\in\{R,O,Y,G,B\}} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} \tag{7}$$

where $j = R, O, Y, G, B$.

Now we will apply the well-known (see, e.g., Bose, 2002) Little's formulae $E(L) = \lambda E(S)$ and $E(Q) = \lambda E(W)$, where L, λ, S, Q, and W stand for the number of customers (here: patients) present in the system, the arrival rate, the sojourn time, the waiting time and the number of customers (patients) waiting in the queue, respectively. We get from (6) and (7) the following representations:

$$E(L_j^{rel}) = \frac{\lambda_j \sum_{i\in\{R,O,Y,G,B\}} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} + \rho_j \tag{8}$$

and

$$E(Q_j^{rel}) = \frac{\lambda_j \sum_{i\in\{R,O,Y,G,B\}} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} \tag{9}$$

In the formulae above, the notation $\sum_{i\geq j}\rho_i$ stands for the sum over all groups of patients with the priority that is higher than or equal to j. Similarly, in the case of $\sum_{i>j}\rho_i$, we take higher priorities than j.

In the case of the absolute priority we have (Adan, Resing, 2015)

$$E(S_j^{abs}) = \frac{\sum_{i\geq j} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} + \frac{E(B_j)}{1-\sum_{i>j}\rho_i} \tag{10}$$

and

$$E(W_j^{abs}) = \frac{\sum_{i\geq j} \rho_i E(R_i)}{(1-\sum_{i\geq j}\rho_i)(1-\sum_{i>j}\rho_i)} \tag{11}$$

where $j = R, O, Y, G, B$.

Hence, utilizing Little's laws, we obtain

$$E\left(L_j^{abs}\right) = \frac{\lambda_j \sum_{i \geq j} \rho_i E(R_i)}{\left(1-\sum_{i \geq j} \rho_i\right)\left(1-\sum_{i>j} \rho_i\right)} + \frac{\rho_j}{1-\sum_{i>j} \rho_i} \tag{12}$$

and

$$E\left(Q_j^{abs}\right) = \frac{\lambda_j \sum_{i \geq j} \rho_i E(R_i)}{\left(1-\sum_{i \geq j} \rho_i\right)\left(1-\sum_{i>j} \rho_i\right)} \tag{13}$$

4. Numerical examples

In this section, we present illustrative numerical examples for some chosen sets of system parameters. The different sets of system input parameters correspond to the real-life scenarios of traffic intensity and patient service in the hospital emergency department. These parameters can be successfully estimated statistically based on concrete "learning" observations in a time window of a fixed width. In all computations (visualized in figures), we investigate the behavior of a given characteristic (e.g., the waiting time) in dependence on the intensity of patient arrivals. Let us note that these examples are illustrative and describe all possible practical situations only to a limited extent.

Scenario no. 1

In the first scenario, we assume that the number of possible diagnostic pathways for the most severely ill patients (red-type) is the highest, equal to 5. Then it decreases by one for the following categories (types) of patients. For patients in the lightest condition (blue-type), there is only one diagnostic path (the service time for such a patient is then described using one exponential random variable with a fixed parameter value). Moreover, we assume that random variables describing the pathways are different for different-type patients. So, we take a set of values presented in Table 1 as the input set of system parameters.

Table 1.
Model parameters for Scenario no. 1

Patient type (j)	**Parameter**			
	Number of diagnostic paths (k_j)	**Frequency (q_j)**	**Parameters of different-path service times ($\mu_i^{(j)}, i = 1, \ldots, k_j$)**	**Frequencies of choosing a particular path ($p_i^{(j)}, i = 1, \ldots, k_j$)**
Red	5	0.10	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.4, 0.3, 0.1, 0.1, 0.1)
Orange	4	0.15	(0.25, 0.35, 0.40, 0.50)	(0.4, 0.3, 0.2, 0.1)
Yellow	3	0.20	(0.30, 0.40, 0.50)	(0.5, 0.4, 0.1)
Green	2	0.15	(0.40, 0.55)	(0.6, 0.4)
Blue	1	0.40	(0.50)	(1.0)

In Figure 1, the behavior of the mean sojourn time in dependence on the arrival rate of incoming patients is presented in the case of relative priority. Similar results for the absolute priority can be observed in Figure 2. Particular colors of lines in figures correspond to the patient type (red, orange, yellow, green, and blue). The values of the arrival intensity λ are taken in such a way as to satisfy the stationary condition of the system ($\rho < 1$). In Figures 3-4, similarly, the behavior of the mean number of patients (of each type) present in the system is shown in the case of relative and absolute priority, respectively.

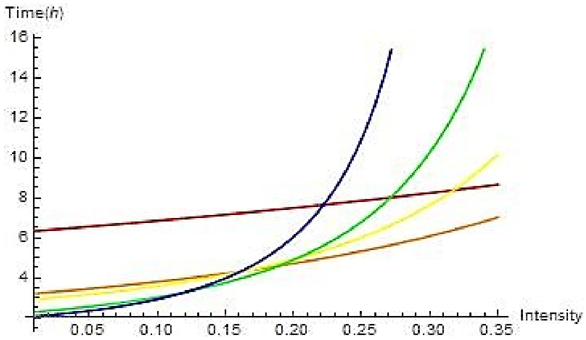

Figure 1. Mean sojourn time for Scenario no. 1 and relative priority

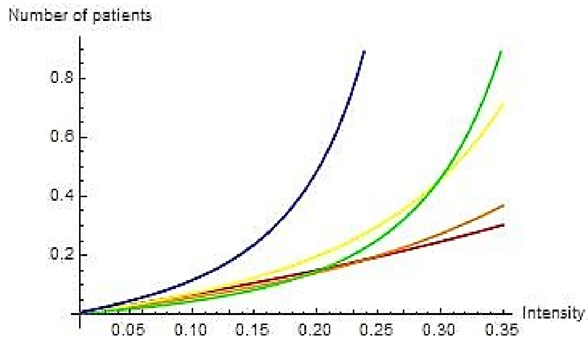

Figure 2. Mean number of patients for Scenario no. 1 and relative priority

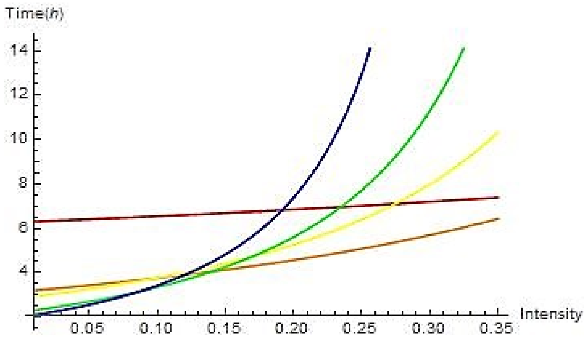

Figure 3. Mean sojourn time for Scenario no. 1 and absolute priority

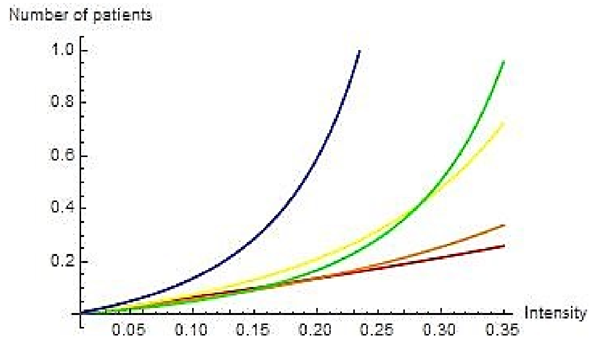

Figure 4. Mean number of patients for Scenario no. 1 and absolute priority

Let us note a significant difference between the behavior of individual characteristics for the relative and absolute priorities. Indeed, in the absolute priority for red-type patients, other-type patients "do not exist". The result is that for the highest possible arrival intensities, e.g., the mean sojourn time for a red-type patient is visibly above 8 hours in the case of relative priority and about 7 hours in the case of absolute priority.

Scenario no. 2

In Scenario no. 2, we leave the number of possible diagnostic paths, the parameters of their distributions, and the frequencies of occurrence of successive types of patients unchanged. In this model, however, we will assume that the longest-lasting diagnostic paths (i.e., those with the lowest values of parameters $\mu_i^{(j)}$) appear the least frequently in practice, and most often those in which the diagnostic and therapeutic process lasts the shortest. The parameters of the model under consideration are presented in Table 2.

Table 2.
Model parameters for Scenario no. 2

Patient type (j)	Parameter			
	Number of diagnostic paths (k_j)	Frequency (q_j)	Parameters of different-path service times ($\mu_i^{(j)}, i = 1, \ldots, k_j$)	Frequencies of choosing a particular path ($p_i^{(j)}, i = 1, \ldots, k_j$)
Red	5	0.10	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.1, 0.1, 0.1, 0.3, 0.4)
Orange	4	0.15	(0.25, 0.35, 0.40, 0.50)	(0.1, 0.2, 0.3, 0.4)
Yellow	3	0.20	(0.30, 0.40, 0.50)	(0.1, 0.4, 0.5)
Green	2	0.15	(0.40, 0.55)	(0.4, 0.6)
Blue	1	0.40	(0.50)	(1.0)

In Figures 5-6, the mean waiting times and the mean numbers of patients waiting in the line are visualized, respectively, for the case of the relative priority, in dependence on the accumulated arrival intensity λ. Corresponding results for the absolute priority are presented in Figures 7-8.

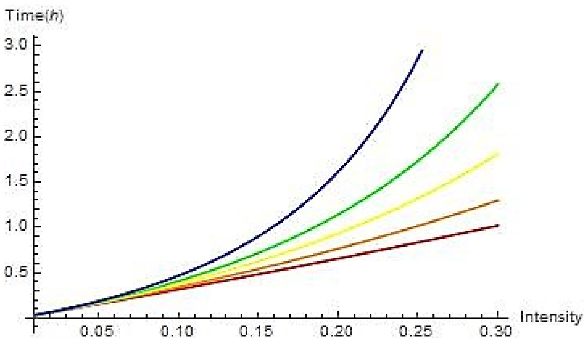

Figure 5. Mean waiting time for Scenario no. 2 and relative priority

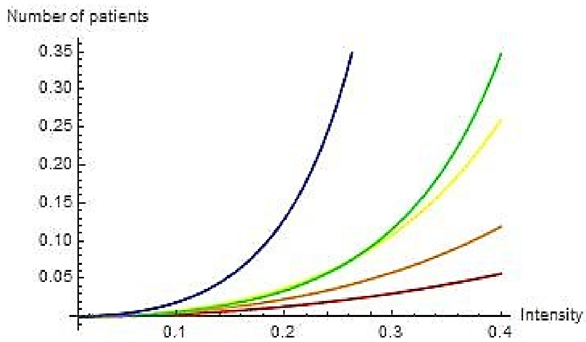

Figure 6. Mean number of patients waiting in the queue for Scenario no. 2 and relative priority

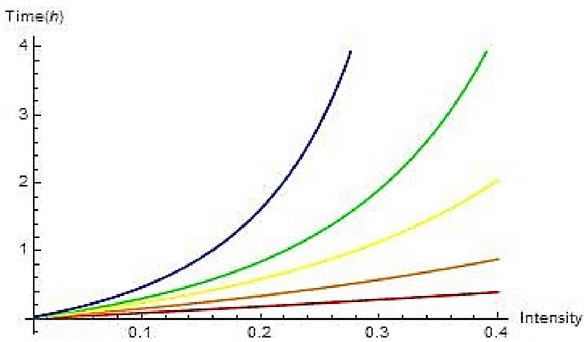

Figure 7. Mean waiting time for Scenario no. 2 and absolute priority

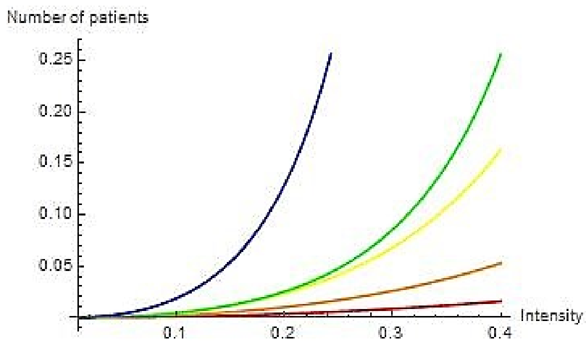

Figure 8. Mean number of patients waiting in the queue for Scenario no. 2 and absolute priority

Let us observe that in the case of Scenario no. 2, the mean waiting time and the mean number of patients waiting in the line (queue) for any lower-priority patient are always less than the corresponding values for a higher-priority one.

Scenario no. 3

In the last scenario, let us consider the situation in which the frequencies of occurrence of each type of patient are the same and equal 0.20. We have five different diagnostic paths for any patient, described by exponential distributions with the same parameters. The essential difference is that the frequency of choosing the longest-lasting one is the greatest for the red-type patient and the smallest for the blue-type patient. All values of parameters of the considered model are presented in Table 3.

Table 3.
Model parameters for Scenario no. 3

Patient type (j)	Parameter			
	Number of diagnostic paths (k_j)	Frequency (q_j)	Parameters of different-path service times ($\mu_i^{(j)}, i = 1, \dots, k_j$)	Frequencies of choosing a particular path ($p_i^{(j)}, i = 1, \dots, k_j$)
Red	5	0.20	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.35, 0.25, 0.20, 0.15, 0.05)
Orange	5	0.20	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.30, 0.25, 0.20, 0.15, 0.10)
Yellow	5	0.20	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.10, 0.20, 0.40, 0.20, 0.10)
Green	5	0.20	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.10, 0.15, 0.20, 0.25, 0.30)
Blue	5	0.20	(0.10, 0.20, 0.30, 0.40, 0.50)	(0.05, 0.15, 0.20, 0.25, 0.35)

In Figures 9-10 the results for the mean sojourn time and the mean number of patients present in the system are shown, respectively, for the relative priority. The corresponding results for the absolute priority are presented in Figures 11-12. Let us note that in the considered model, an "approximate" equilibrium occurs in the case of the mean sojourn time, i.e., for the certain intensity of patient arrival, the mean waiting time is approximately the same for all-type patients.

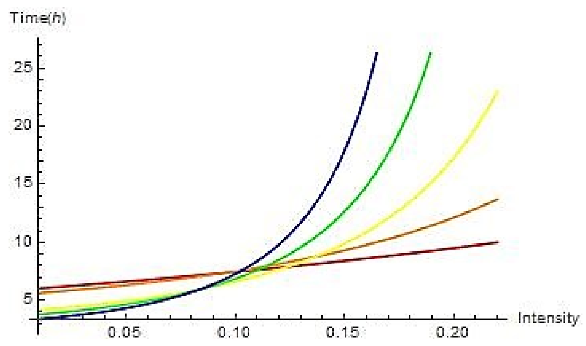

Figure 9. Mean sojourn time for Scenario no. 3 and relative priority

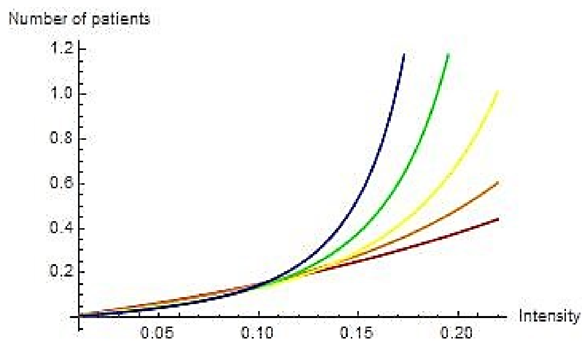

Figure 10. Mean number of patients for Scenario no. 3 and relative priority

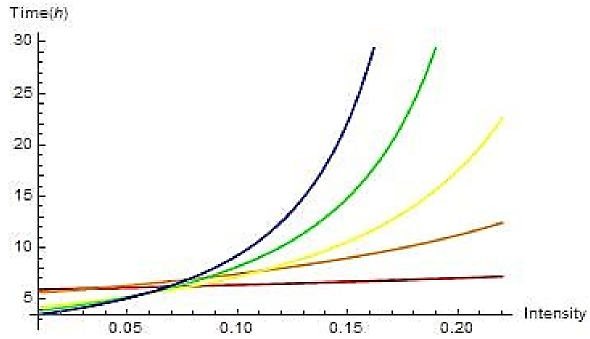

Figure 11. Mean sojourn time for Scenario no. 3 and absolute priority

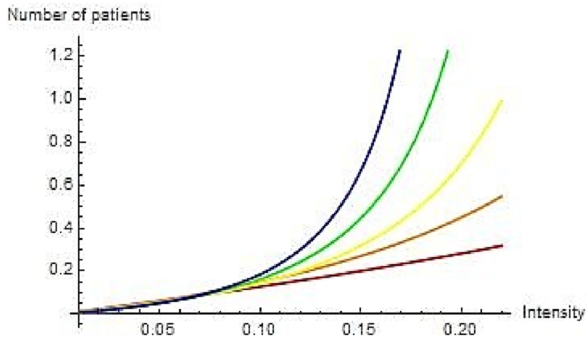

Figure 12. Mean number of patients for Scenario no. 3 and absolute priority

5. Conclusions

The paper proposed a queueing model of patient service in a hospital emergency department, compliant with the currently used S.T.A.R.T standard. The model assumed two types of priority patient service depending on their health condition: service with absolute priority and relative priority. The results for the average sojourn time of the each-type patient, the average waiting time for the start of the diagnostic process, as well as for the mean number of patients present in the system and the mean number of patients waiting in the queue were obtained. Analytical results were illustrated by numerical calculations considering three different scenarios. The proposed model creates vast application possibilities for a more precise assessment and optimization of the work of the hospital emergency department in various conditions that are planned in the future.

References

1. Adan, I., Resing, J. (2015). *Queueing Systems*. Eindhoven. Available online: https://www.win.tue.nl/~iadan/queueing.pdf.
2. Badiali, S., Giugni, A., Marcis, L. (2017). Testing the START Triage Protocol: Can It Improve the Ability of Nonmedical Personnel to Better Triage Patients During Disasters and Mass Casualties Incidents? *Disaster Medicine and Public Health Preparedness, Vol. 11, No. 3*, pp. 305-309.
3. Baker, M.S. (2007). Creating order from chaos: part I: triage, initial care, and tactical considerations in mass casualty and disaster response. *Military Medicine, Vol. 172, No. 3*, pp. 232-236.

4. Benson, M., Koenig, K.L., Schultz, C.H. (1996). Disaster triage: START, then SAVE – a new method of dynamic triage for victims of a catastrophic earthquake. *Prehospital and Disaster Medicine*, *Vol. 11, No. 2*, pp. 117-124.
5. Bose, S.K. (2002). *An Introduction to Queueing Systems*. New York: Springer Science+Business Media, LLC.
6. Chan, W.Ch. (2014). *An Elementary Introduction to Queueing Systems*. Singapore: World Scientific Publishing Co. Pte. Ltd.
7. Ferrandini, P.M., Arcos, G.P., Pardo, R.M., Nieto, F.-P.A., Cuartas, Á.T., Castro, D.R. (2018). Comparison of the Simple Triage and Rapid Treatment system versus the Prehospital Advanced Triage Model in multiple-casualty events. *Emergencias, Vol. 30, No. 4*, pp. 224-230.
8. Fink, B.N., Rega, P.P., Sexton, M.E. (2018). START versus SALT Triage: Which is Preferred by the 21st Century Health Care Student? *Disaster Medicine, Vol. 33, No. 4*, pp. 381-386.
9. Heyman, D.P., Sobel, M.J. (1982). *Stochastic Models in Operations Research. Vol. I: Stochastic Processes and Operating Characteristics*. Mineola, New York: Dover Publications, Inc.
10. Kahn, C.A., Schultz, C.H., Miller, K.T., Anderson, C.L. (2009). Does START triage work? An outcomes assessment after a disaster. *Annals of Emergency Medicine, Vol. 53, No. 3*, pp. 424-430.
11. Lakatos, L., Szeidl, L., Telek, M. (2019). Introduction to Queueing Systems. In: *Introduction to Queueing Systems with Telecommunication Applications*. Cham: Springer.
12. Ng, Ch.-H., Soong, B.-H. (2008). *Queueing Modelling Fundamentals with Applications in Communications Networks*. Chichester: John Wiley & Sons, Inc.
13. Shortle, J.F., Thompson, J.M., Gross, D., Harris, C.M. (2018). *Fundamentals of Queueing Theory*. Hoboken: John Wiley & Sons, Inc.
14. Tijms, H.C. (2003). *A First Course in Stochastic Models*. Chichester: John Wiley & Sons, Inc.