Queue theory Applied to the care of Pregnancies related to Pregnancy, Childbirth and the Puerperium

João Paulo Zordan da Silva[1], Wellington Gonçalves[2], Rodrigo Randow de Freitas[3], William Gonçalves[4], Giglliara Segantini de Menezes[5]

[1]Department of Production Engineering at the Federal University of Espírito Santo, UFES – São Mateus/ES, BRA
Email: joaopzordan@gmail.com
[2]Department of Engineering and Technology, Operations Research Laboratory, Logistics and Transport (POLT) at the Federal University of Espírito Santo, UFES – São Mateus/ES, BRA
Email: wellington.goncalves@ufes.br; rodrigo.r.freitas@ufes.br
[3]Department of Engineering and Technology at the Federal University of Espírito Santo, UFES – São Mateus/ES, BRA
Email: william.goncalves@ufes.br
[4]Administrative Management at the Federal University of Espírito Santo, UFES – São Mateus/ES, BRA
Email: william.goncalves@ufes.br
[5]Federal Institute of Espírito Santo, IFES - São Mateus/ES, BRA
Email: giglliara.menezes@ifes.edu.br

Abstract*— *This work investigates the increased length of care related to pregnancy, childbirth and the puerperium. It also proposes a simple approach to illustrate the effect of poor care in a maternity hospital. Applying Queue Theory with the Kolmogorov-Smirnov Test. In addition, it provides evidence that non-emergency patients, and also because of the lack of useful information in the primary care process, contribute to prolonged delays for all patient classes. Proposes a priority queuing model to reduce average wait time.

***Keywords*— Length of stay; Capacity management; Occupancy rate; Emergency patient flow.**

I. INTRODUCTION

Queues can be found in many everyday situations and are responsible for wasted time, productivity and money. Depending on their nature, queues can lead to personal and collective disruption, serious financial, economic, organizational and even life-threatening problems. For Peres et al. (2019) Patients are affected by the lack of availability of convenient times on the provider's schedule, especially when their care is urgent, and providers are affected by the uncertainty of appointments each day and the variety of cases that may arise.

According to Fogliatti and Mattos (2007), a queuing system can be defined as a process in which users from a certain population arrive to receive a service for which they expect, if necessary, to leave the system as soon as the service is completed. This wait happens when demand is greater than the service capacity offered, in terms of flow.

Existence as well as waiting time in queues is related to the efficiency and the ability of the system to provide services. The study of processes and queues aims at adapting the system in question to a queue model that allows the minimization of waiting times and, consequently, increase the efficiency of the studied system. According to Mendonça (2014), the objective of the queues study is to estimate the parameters involved in the model and to calculate some measures of its performance, considering the particularities of each case and, thus, to search systems that efficiently meet the needs of those who seek the service without let the system be idle for a long time.

Operational Research, through Queue Theory, is used to study queuing processes to produce real process performance indicators and proposed scenarios that will assist in decision making to minimize queue size and the permanence of individuals in them.

Care for pregnant women, parturients, mothers and newborns is a central problem in health systems worldwide (WHO, 2005). According to data from the Department of Information and Informatics of the Brazilian Unified Health System (DATASUS), between 2010 and 2016, more than 50 unborn children died, in addition to a total of over 100 maternal mortality related to pregnancy, childbirth and puerperium in the São Mateus microregion (comprised of the municipalities of São Mateus, Pedro Canário, Jaguaré and Conceição da Barra). It is noteworthy that the need to reduce maternal mortality rates, humanize births and the health prospects of unborn children depend, among other factors, also fundamentally on the quality and speed of care in health services. According to the National Health Agency (ANS), waiting time for care plays an unquestionable role in the health of pregnant women, mothers and newborns.

According to Dong et al. (2018), Delay announcements, common in service systems, can be used to influence quality perceptions and customer sentiment toward the service provider. In addition, these announcements may affect customer choices regarding emergency service providers, with subsequent effects on actual system operations.

According to Shi et al. (2015), a key factor contributing to the overcrowding of the emergency department (ED) is the extended waiting time for admission to inpatient wards, also known as hospital boarding time. However, there are no systematic records of waiting times for hospitalizations related to pregnancy (pregnancy, childbirth and the puerperium) in the Unified Health System (SUS). To estimate these times, we used the models of queuing theories that have been widely applied in health. Silva et al. (2018) applies the concepts of Queue Theory to analyze referrals from SUS patients to the medical specialty of angiology. The literature indicates that such models are particularly useful in analyzing childbirth-related phenomena. Although natural, the phenomenon of childbirth is a kind of emergency, since the moment of occurrence cannot be perfectly predicted and whose process, once triggered, cannot be much delayed. Delay in care usually entails serious health risks for the parturient and the unborn child. On the other hand, births cannot be freely scheduled either. The moment of childbirth has a strongly random character, which gives the phenomenon a degree of exogeneity appropriate to the queuing theory models that will be used in the present work.

II. LITERATURE REVIEW

According to Haghighinejad et al. (2016), are elements of the queue: arrival process of customers; attendance process; queuing discipline and system capacity. In the opinion of Komashie et al. (2015), the Queue Theory study proves to be a tool that ensures optimization of service companies, a fact that can generate the satisfaction of system users.

According to Lima et al. (2016), queuing theory consists in obtaining adequate models of circumstances involving queues, thus predicting their behavior. This behavior is expressed by a number of performance measures, namely customer arrival rate and system service rate.

Fogliatti and Mattos (2007) point out that a characteristic queued system is determined by the arrival of customers who need a certain service, they wait in a queue that forms as the service capacity is less than the demand of the users, are served and leave the system then. Figure 1 illustrates an arbitrary queued system.

Brahma (2013) reports that in hospitals, queuing theory can be used to assess a multitude of factors such as prescription time, patient waiting time, patient counseling time, and staffing levels. Applying queuing theory can be particularly beneficial in hospitals with high-volume outpatient workloads and / or those providing multiple service points.

Cho et al. (2017) emphasize that due to the characteristics of medical services, it is very difficult to predict exactly when a patient will arrive and how long it will take for the service. Therefore, the ultimate goal of queuing theory is to achieve an economic balance between cost of service and patients' time wasted while waiting in line to be served.

For Hamacher (2017), the process of arrival of a user in the system can be deterministic when the amount of arrivals and the moments in which they occur are known, or stochastic, when there is a random behavior. Stochastic processes are characterized by probability distributions.

The arrival of customers in the system is also classified by Fogliatti and Mattos (2007) as individual or in groups. An example for the individual case is the arrival of a car in a parking lot of a mall. For group arrival, an example is data packets waiting to be processed.

Also called service stations, the number of service channels is classified by Hamacher (2017) as finite or infinite. The finite case is exemplified by boxes in a bank and infinity in self-service calls.

https://dx.doi.org/10.22161/ijaers.69.33 ISSN: 2349-6495(P) | 2456-1908(O)

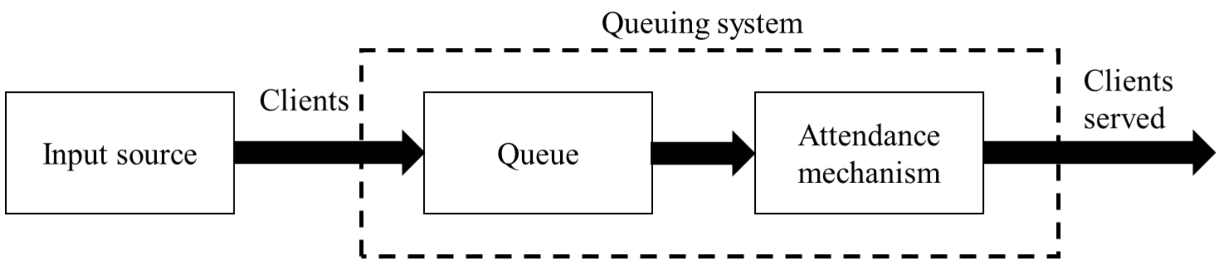

Fig. 1: Schematic representation of a queued system.

According to Brahma (2013), by better understanding queuing theory, service managers can make decisions that increase the satisfaction of all relevant groups - patients, staff and management.

Bhattacharjee and Ray (2018) define a system's capacity as the maximum number of users it can handle - including queue and service - and can be finite or infinite. When capacity is finite, customers arriving at the system after the maximum capacity is reached are not met. In the case of infinite capacity, one can cite the waiting for ships to attend a waterway environment for unloading in a port.

According to Fogliatti and Mattos (2007), the most commonly used service orders, or service literature, are: First In - First Out (FIFO): The service is performed according to the order of arrival; Last In - First Out (LIFO): The last user to arrive is the first to be served; Priority Service (PRI): Priority Service determined by system management is used for the occurrence of calls; e Service In Random Order (SIRO): Service is performed in random order. For the authors it is important to know and also understand what type of queue the system analyzed has, in its various service stations or units.

Bhattacharjee and Ray (2018) point out that Kendall-Lee Notation is a universally used way of describing systems and thus simplifying their analysis by using five characteristics described as follows.- A/B/C /D/E – Where: A: refers to the probability distribution type of the system arrival process; B: refers to the type of probability distribution of the care process; C: determines the number of service stations in parallel; D: indicates the physical capacity of the system; and E: makes mention of the discipline of the queue.

Parameters A and B can assume D, M, E_k e G, where D indicates a deterministic or degenerate distribution, M denotes an Exponential (Memoryless or Markovian) distribution, E_k represents an Erlang distribution of the type k and G refers to an unspecified distribution.

For example, the notation $D/D/1/\infty/FIFO$ represents a system in which the intervals between successive arrival of users to the system and the attendance times behave according to a deterministic distribution. There is only one service desk, the system has infinite capacity and the service takes place according to FIFO discipline.

The Markovian process, according to Dai and Shi (2017), is a stochastic (probabilistic) process that has the property called memoryless. According to this property, the probability distribution of an event is independent of the events that occurred before it, but depends on the current state. A Markov chain is a sequence $X_1, X_2, X_3, ..., X_n$ of random variables. The scope of these variables, that is, the set of values they can assume, is called the state space, where X_n denotes the state of the process in time n. If the conditional probability distribution of X_{n+1} in the past states is a function only of X_n, so:

$$P_r(X_{n+1} = x / X_0, X_1, X_2, ..., X_n) = P_r(X_{n+1} = X_n),$$

where X it is some state of the process. The above identity defines Markov property.

The Markov Chain is a Markov process and can be represented by a diagram showing the discrete states of a system through nodes, and the transitions between these states through arcs (GNEDENKO, 2018). State is defined as the number of users in a system. A process of birth and death is a Markov Chain where, given state n, the only possible changes are to state $n+1$ or $n-1$. Birth is the change to state $n+1$ and death to the state $n-1$. Transitions depend only on the state of the system, through and, which are the user arrival and user attendance rates, respectively.

For Chan et al. (2016), performance measures should be used to evaluate the efficiency of a system by analyzing its characteristics. According to Dai and Shi (2017), the most commonly used performance measures are: L: average number of clients in the system; L_q: average number of clients queued; W: average waiting time on

https://dx.doi.org/10.22161/ijaers.69.33 ISSN: 2349-6495(P) | 2456-1908(O)

any customer's system; W_q: average queue time for any customer; $P(N \le n)$: It is likely that there will be no more than n; $P(T_q > t)$: probability that a customer will have to wait more than one time t in line; and $P(N < c)$: probability that a system with c service stations has some idle server.

Probability distributions can be continuous or discrete. According to Chan et al. (2016), a probability distribution of a discrete variable is the set of values of a discrete Random Variable (RV) and their respective probabilities. The difference for continuous variables is that, according to Dai and Shi (2017), instead of variable values, there is a range of values.

Next, only the exponential distribution will be addressed, as this is the one that fits the case study approached by this paper. For Gnedenko (2018), Exponential distribution is a continuous distribution that is widely used in queuing theory to describe the time spent developing a task. This author defines the exponential cumulative distribution function as follows (Equation 1):

$$F(x) = 1 - e^{\frac{x}{\alpha}}, \text{ since } x, \alpha > 0 \quad (1)$$

Where: x is the value of the random variable; and α is the mean value of the random variable.

However, after identifying the distribution, according to Hamacher (2017), it is necessary to perform adherence tests. These authors emphasize that the adherence test is a way of verifying if a population P follows a specified distribution P_0, that is, to verify the null hypothesis $H_0 : P = P_0$.

There are basically two types of adherence test: the Chi-square test, preferably used in discrete distributions, and the Kolmogorov-Smirnov test, preferably used in continuous distributions. Next, only the Kolmogorov-Smirnov test is presented as it was used in this work, due to the continuous characteristic of the variable in question - the time. According to Kasjanov and Szafran (2015), be: a sample $x_1, ..., x_n$ referring to a random variable X of a population P, $N(x)$: the number of observations less than or equal to x_n; $F(x)$: RV Cumulative Distribution Function (CDF); and $F_e(x)$: Empirical Density Function (EDF).

The purpose of the test is to verify the null hypothesis $H_0 : F(x) = F_0(x)$, for every x, that is, test whether the cumulative distribution of $X - F(x)$ matches a specified probability distribution $F_0(x)$.

It is necessary to compare the CDF of X under null hypothesis with the EDF from the realization of the difference between these functions. Asjanov and Szafran (2015) state that, given n sample points, the EDF is given by Equation 2.

$$F_e(x) = \frac{N(x)}{n} \quad (2)$$

The maximum absolute value D shall be obtained between the CDF of X, under zero hypothesis and the EDF (Equation 3).

$$D = \max |F(x_i) - F_e(x_i)|, \text{ for } 1 \le i \le n \quad (3)$$

The objective of the test is reached when the value found for D is compared to a critical value D_c, given by the Kolmogorov-Smirnov Distribution table - easily found in the literature - D by fixing the number n of samples and the significance level chosen. The significance level is the probability of rejecting the null hypothesis $H_0 : F(x) = F_0(x)$ if it is true, that is to say that the cumulative distribution of $X, F(x)$, does not match a specified probability distribution $F_0(x)$, when it actually matches. The significance level generally assumes values equal to 5%, 1% or 0.1%. Thus, if the value of D is less than the tabulated critical value, the null hypothesis H_0 is true, that is, the cumulative distribution matches a specified distribution. Otherwise, H_0 is rejected and therefore the parsed distribution does not match the specified distribution.

According to Asjanov and Szafran (2015), the critical value D_c for the Kolmogorov-Smirnov test, for a number of observations $n = 22$ and at 5% significance level, is 0.281.

In the literature there are several queue models, as can be seen in Fogliatti and Mattos (2007). However, due to the characteristics of the target problem of this work, the system of queues in SUS care in the pregnant and parturient sector of the hospital under study, we will present only the $M / M / 1 / \infty / FIFO$ model, as this is the model that represents the current situation in this sector of the hospital.

According to Bergsma (2018), in the $M / M / 1 / \infty / FIFO$ model, the intervals between successive arrivals and service times follow Markovian distributions, that is, these events occur according to exponential distributions, in a process of birth and death.

https://dx.doi.org/10.22161/ijaers.69.33 ISSN: 2349-6495(P) | 2456-1908(O)

This model has only one service desk and has infinite physical capacity for the queue. The service occurs according to the order of users' arrival at the system and has a constant rate, described by: $\lambda_n - \lambda, \quad \forall n \geq 0$

The system attendance rate is also constant, given by: $\mu_n - \mu, \quad \forall n \geq 1$

The parameter ρ, called system occupancy / utilization rate, is defined by Equation 4.

$$\rho = \frac{\lambda}{\mu} \tag{4}$$

Thus, in this framework two performance measures had their equations described for the $M / M / 1 / \infty / FIFO$ model. As presented above, there are other performance measures, however, for applicability purposes in this work, only two will be used (FOGLIATTI; MATTOS, 2007), namely:

a) Average queue time (W_q): Given by Equation 5.

$$W_q = \frac{\rho}{\mu - \lambda} \tag{5}$$

b) Average length of stay in the system (W): Given by Equation 6.

$$W = \frac{1}{\mu - \lambda} \tag{6}$$

III. METHODOLOGICAL APPROACH

The population of this study was surveyed using the sampling technique, which consists of choosing a representative part of the whole for the study, so that the results of the sample were sufficient to draw conclusions about the total population. Thus, the population of this research is formed by pregnant women and parturients users of SUS for hospitalization related to pregnancy, childbirth and the puerperium in the studied locality and region, and the sample is formed by data collected in the pregnant women care sector.

The surveyed sample consists of the people who arrive at the emergency room on a given day and time, but only the people who use the system, that is, those who arrive and receive care at the reception. Therefore, employees, visitors and service providers who pass through the pregnant women sector using it only as a route to enter the hospital are not accounted for in the sample.

The study in question is characterized by its exploratory and descriptive nature, since it is based on on-site observations with the purpose of investigating, describing and exposing the characteristics of the problem under study, as well as allowing the comparison of the existing scenario in the health sector pregnant women of this particular hospital with the Brazilian standards for pregnant women care, and thus it can be inferred whether the average waiting time in the pregnant women sector queue and the average length of stay in this sector contribute significantly to the total delay in care - ranging from the arrival of the user in the system to the doctor's care - including childbirth, prenatal examinations, etc. Thus, due to its characteristics and applied nature, it uses qualitative and quantitative approaches for diagnosis, description, discussion and comparison.

For the development of the present work, visits and data collection were made in the pregnant women care sector, with the purpose of diagnosing the current conjuncture of expectation, arrival and care of pregnant women. From this initial diagnosis, a study of Queue Theory was made in order to find a queue model that, depending on its parameters and its applicability, best represents the case in question. To find this queue model, from the data collected above, the average arrival time of customers in the queue and the average attendance rate were calculated. From these data, an adhesion test was applied - in this case the Kolmogorov-Sminorv test - as a function of the continuous characteristics of VR to verify the data adherence with the exponential distribution and, from this, to be able to identify the row model. in question, calculate their parameters - W_q and W – and then compare them with the average standards in this type of system in Brazil.

IV. APPLICATION AND ANALYSIS OF RESULTS

The queuing system studied in this paper is associated with the reception of the pregnant women service sector, as mentioned earlier. This system is delimited by the door through which users enter the reception and the door that gives access to the interior of the hospital, comprising the service window and some seats for users and companions (Figure 2).

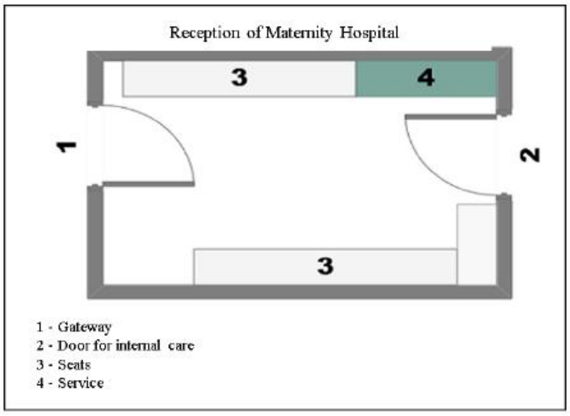

Fig 2: Physical area that served as research unit.

https://dx.doi.org/10.22161/ijaers.69.33 ISSN: 2349-6495(P) | 2456-1908(O)

Pregnant women arrive at the system through the reception's front door and wait for the service in line. At the counter, the service consists of checking personal documents and filling out a form with patient data. Queue discipline follows the FIFO pattern, so that service is handled in order of arrival to queue.

This research began with seeking authorization from the hospital administration in question for data collection. It was also asked the administration which would be the busiest days and times in this sector, in order to diagnose the formation of the queue. Therefore, as indicated by the administrator, the analyzed period was a Friday (4/19/2019) from 07h40min to 12h30min, encompassing the peak service - which takes place between 08h30min and 11h00min. These collection periods that demonstrate a pre- and post-peak service clearance are called pre- and post-warming, respectively, and are necessary because, according to Bergsma (2018), there is no precise method or way to establish a starting and ending point for the equilibrium state of a queued system.

Data processing began with the application of the Kolmogorov-Smirnov adherence test, to verify that the times between arrivals and service times behaved according to an Exponential distribution.

From this finding, the equations referring to Queue Theory were applied, according to section 2 of this paper. Thus, the results generated informed the current situation of the Emergency Service studied in relation to two performance measures: The average waiting time in the queue and the average user time in the system.

V. ANALYSIS OF CUSTOMER ARRIVAL PROCESS

Table 1 shows the data collected regarding the users' arrivals to the system.

Table. 1: Values involved in the Kolmogorov-Smirnov test for the arrival interval.

Patient	Arrivals Time	Interval (s)	Orderly	$F(x)$	$F_e(x)$	$\lvert F(x)-F_e(x)\rvert$
1	08:11:20					
2	08:20:29	549	31	0.0450	0.0476	0.0027
3	08:31:27	658	40	0.0576	0.0952	0.0376
4	08:39:10	463	65	0.0920	0.1429	0.0509
5	08:40:50	100	100	0.1379	0.1905	0.0526
6	08:49:13	503	132	0.1779	0.2381	0.0602
7	08:54:50	337	140	0.1876	0.2857	0.0981
8	08:57:02	132	160	0.2114	0.3333	0.1220
9	09:09:24	742	271	0.3311	0.3810	0.0498
11	09:22:35	520	337	0.3936	0.4762	0.0826
12	09:30:00	445	445	0.4834	0.5238	0.0404
13	09:30:31	31	463	0.4970	0.5714	0.0744
14	10:23:55	3204	503	0.5260	0.6190	0.0931
15	10:28:35	280	520	0.5378	0.6667	0.1289
16	10:30:55	140	549	0.5573	0.7143	**0.1570**
17	10:31:35	40	658	0.6234	0.7619	0.1385
18	11:11:05	2370	742	0.6675	0.8905	0.1420
19	11:36:05	1500	1500	0.8921	0.8571	0.0349
20	11:37:10	65	1640	0.9123	0.9048	0.0075
21	12:04:30	1640	2370	0.9703	0.9524	0.0179
22	12:07:10	160	3204	0.9914	1.0000	0.0086

Table 1 indicates that 22 clients arrived in the system during the surveyed period. Generating a rate of arrival of $\lambda = 0.0933$ clients per minute. The average interval between successive arrivals is $\alpha_1 = 673.81$ seconds.

In order to perform this test the null hypothesis is that the intervals between consecutive arrivals at the reception of the pregnant women sector correspond to an Exponential probability distribution. and with parameter $\alpha_1 = 673.81$ seconds. Equation 7 was obtained.

$$F(x) = 1 - e^{\frac{x}{673.81}}, x > 0 \quad (7)$$

Initially the values related to the intervals between

https://dx.doi.org/10.22161/ijaers.69.33 ISSN: 2349-6495(P) | 2456-1908(O)

user arrivals were ordered. From Equation 7 the values of the cumulative distribution function of $X, F(x)$ under H_0 were calculated and by Equation 2 the values for the empirical density function were calculated. With the values of the cumulative distribution function and the empirical distribution function $F_e(x)$ the absolute value of the difference between these functions was obtained. The results are shown in Table 1.

It can be seen from Table 1 that the maximum absolute value of the differences between $F(x)$ and $F_e(x)$ has the value of $D = 0.1570$ (highlighted in bold). According to values found in the Kolmogorov-Smirnov table. the critical value for the test is $D_c = 0.2810$ with 22 intervals observed. Thus. since D is less than D_c. it can be seen that the intervals between users' arrival at the system fit an Exponential distribution with parameter $\alpha_1 = 673.81$ seconds at the 5% significance level. The value of 5% was chosen for the significance level based on literature indications (FOGLIATTI; MATTOS. 2007).

Figure 3 illustrates this grip. The cumulative relative frequencies of the arrival intervals and the exponential parameter function $\alpha_1 = 673.81$ seconds are displayed.

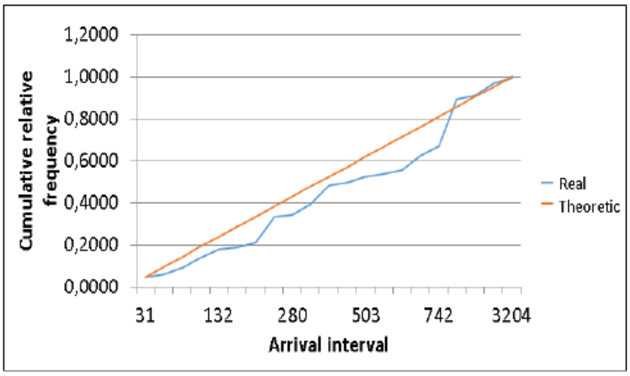

Fig. 3: Cumulative distribution of arrival interval function.

Table 2 shows the data collected regarding the users' arrivals to the system.

Table. 2: Values involved in the Kolmogorov-Smirnov test for the care of each user in the system.

Patient	Duration (s)	Orderly	$F(x)$	$F_e(x)$	$\lvert F(x) - F_e(x) \rvert$
1	156	35	0.1849	0.4545	0.1395
2	40	40	0.2084	0.0909	0.1175
3	94	60	0.2967	0.1364	0.1593
4	35	65	0.3159	0.1818	0.1341
5	190	80	0.3733	0.2273	0.1461
6	185	94	0.4225	0.2727	0.1498
7	130	100	0.4424	0.3182	0.1243
8	217	130	0.5321	0.3636	**0.1684**
9	315	133	0.5402	0.4091	0.1311
10	360	135	0.5455	0.4545	0.0910
11	264	137	0.5508	0.5000	0.0508
12	137	138	0.5534	0.5455	0.0080
13	210	140	0.5586	0.5909	0.0323
14	60	156	0.5980	0.6364	0.384
15	80	185	0.6606	0.6818	0.0212
16	100	190	0.6704	0.7273	0.0569
17	582	210	0.7068	0.7727	0.0660
18	138	217	0.7185	0.8182	0.0997
19	135	264	0.7861	0.8636	0.0775
20	140	315	0.8412	0.9091	0.0679
21	65	360	0.8779	0.9545	0.0766
22	133	582	0.9666	1.0000	0.0334

Table 2 indicates that 22 clients were served during the surveyed period. indicating a rate of attendance of $\mu = 0.3505$ clients per minute. Average answer time is

$\alpha_2 = 171.18$ seconds.

For the performance of this test the null hypothesis is that the service times at the reception of the PS match an Exponential probability distribution according to Equation 1. and with parameter $\alpha_2 = 171.18$ seconds.

$$F(x) = 1 - e^{\frac{x}{171.18}}, x > 0 \quad (8)$$

The test was performed analogously to the interval between arrivals. The values involved are shown in Table 2.

It can be seen from Table 2 that the maximum absolute value of the differences between $F(x)$ and $F_e(x)$ has the value of $D = 0.1684$ (highlighted in bold). According to values found in the Kolmogorov-Smirnov table. the critical value for the test is $D_c = 0.2810$ with 22 observed intervals. Thus. since D is less than D_c. it can be seen that the intervals between users' arrival at the system fit an Exponential distribution with parameter $\alpha_2 = 171.18$ seconds at the 5% significance level.

Figure 4 illustrates this grip. The cumulative relative frequencies of the arrival intervals and the exponential parameter function $\alpha_2 = 171.18$ seconds are displayed.

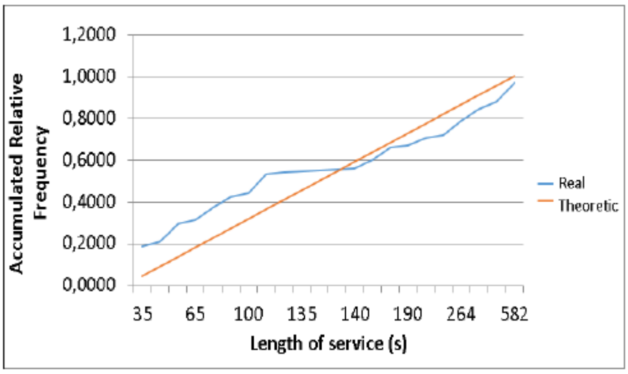

Fig. 4: Cumulative distribution function of service times.

It is important to note that it was found that the intervals between user arrivals and user service times for pregnant women are adjusted to exponential distributions. According to what was informed by the hospital management and observed in the data collection there is only one service station where customer service is received at reception and this service occurs in the order of arrival of users to the queue. In the data collection it was also noted that the queue starts at the reception desk and follows the extension of the reception room of this sector and that although the system is delimited from the entrance door to the reception desk this queue may extend beyond the door and continue on the street indicating that there is no limitation of the physical capacity of the system. Thus. it is concluded that the studied system is characterized as a $M / M / 1 / \infty / FIFO$ queue model.

The value found for the expected number of arrivals per unit of time $\lambda = 0.0933$ clients per minute - and the value found for the expected number of calls per unit of time $\mu = 0.3505$ clients per minute - were used in Equation 4 to calculate the system load factor – ρ. Thus. the system load / utilization rate is $\rho \cong 0.27 < 1$ indicating that 27% of the time the server is busy performing service.

VI. FINAL CONSIDERATIONS

The average waiting time in the queue (W_q) is 1.03 minutes and the average time spent in the system (W) is 3.89 minutes for the reception of the pregnant sector of the hospital under study. Therefore, the average waiting time in line and the average time spent at the reception of the pregnant women sector of this hospital are not significant when considering the entire process - from patient arrival to medical care. Therefore, in order to reduce this total time - time at reception plus internal time including hospitalization and medical care - management efforts and efforts should focus on reducing this internal time.

This study had a limitation in which different scales and compositions of medical teams were not considered in the scenario evaluation. In conclusion the service delivery system studied can be improved by reallocating available staff and also revising and modifying patient flow at no additional cost using queuing theory and simulation techniques. In other words, waiting times and patient waiting times can be reduced by using multitasking people and relocating them to the lengthy stage of filling in initial information without making any noticeable change in queuing characteristics.

ACKNOWLEDGEMENTS

Federal University of Espírito Santo (UFES)/ University Center North of the Espírito Santo (CEUNES). The Espírito Santo Research Support and Innovation Foundation - Announcements: FAPES-CNPq nº 13/2018 - PICJr; FAPES No. 21/2018 - Universal and FAPES No. 18/2018 - Researcher Capixaba Scholarship.

REFERENCES

[1] Fogliatti. M. C. and Mattos. N. M. C. *Teoria de Filas.* Rio de Janeiro: Interciência. 2007.

[2] Gnedenko. B. V. (2018). *Theory of probability*. Routledge.

[3] Kasjanov. V. and Szafran. K. (2015). Some hybrid models of subjective analysis In the theory of active systems. *Prace Instytutu Lotnictwa*. *3*(240). 27-31.

[4] Bergsma. C. (2018). *A cure for the queue: Scenario based optimization at ZGT's radiology department*. 74 f. Dissertation (Master in Industrial Engineering and Management) - Faculty of Behavioural Management and Social Sciences. Department of Industrial Engineering and Business Information Systems. Centre for Healthcare Operations Improvement and Research. University of Twente. Enschede. Netherlands.

[5] Bhattacharjee. P. and Ray. P. K. (2018). Scheduling appointments for multiple classes of patients in presence of unscheduled arrivals: Case study of a CT department. *IISE Transactions on Healthcare Systems Engineering*. *8*(3). 181-195.

[6] Chan. C. W., Dong. J. and Green. L. V. (2016). Queues with time-varying arrivals and inspections with applications to hospital discharge policies. *Operations Research*. *65*(2). 469-495.

[7] Dai. J. G. and Shi. P. (2017). A two-time-scale approach to time-varying queues in hospital inpatient flow management. *Operations Research*. *65*(2). 514-536.

[8] Departamento de Informática do SUS (DATASUS). (2019). *Informações de Saúde (TABNET) - Indicadores de Saúde e Pactuações*. Retrieved on September 10. 2019. from http://www2.datasus.gov.br/DATASUS/index.php?area=0201&id=6903.

[9] Dong. J., Yom-Tov. E. and Yom-Tov. G. B. (2018). The impact of delay announcements on hospital network coordination and waiting times. *Management Science*. *65*(5). 1969-1994.

[10] Hamacher. S. (2017). *Simulação de políticas de agendamento em serviços ambulatoriais*. 130 f. Dissertation (Master in Production Engineering) - Engineering Industrial department. Pontifical Catholic University – Rio de Janeiro. Rio de Janeiro. Brazil.

[11] Peres. I. T., Hamacher. S., Oliveira. F. L. C., Barbosa. S. D. J. and Viegas. F. (2019). Simulation of Appointment Scheduling Policies: a Study in a Bariatric Clinic. *Obesity surgery*. *29*(9). 2824–2830.

[12] Shi. P., Chou. M. C., Dai. J. G., Ding. D. and Sim. J. (2015). Models and insights for hospital inpatient operations: Time-dependent ED boarding time. *Management Science*. *62*(1). 1-28.

[13] Haghighinejad. H. A., Kharazmi. E., Hatam. N., Yousefi. S., Hesami. S. A., Danaei. M. and Askarian. M. (2016). Using queuing theory and simulation modelling to reduce waiting times in an Iranian emergency department. *International journal of community based nursing and midwifery*. *4*(1). 11-26.

[14] Mendonça. E. B. (2014). Teoria de filas Markovianas e Aplicações. 64 f. Dissertation bachelor of statistics - Science and Technology Center – Campina Grande. Paraíba. Brazil.

[15] Brahma. P. K. (2013). Queuing theory and customer satisfaction: a review of terminology. trends. and applications to hospital practice. *Asia Pacific Journal of Marketing & Management Review*. *2*(6). 83-89.

[16] Silva. R. K., Moraes. E. M., Mota. N. V. V. P. and Mourad. R. G. (2018). Simulação e teoria das filas aplicadas na análise dos encaminhamentos de pacientes SUS para a especialidade médica de angiologia. *Revista de Administração em Saúde*. *18*(72).

[17] Lima. B. P., Novais. A. da S. S., Carvalho. C. C. and Negrão. L. M. (2016). A teoria de filas como ferramenta de apoio na análise de um serviço de atendimento. *Anais*... Simpósio de Excelência em Gestão e Tecnologia (SEGET).

[18] Cho. K. W., Kim. S. M., Chae. Y. M. and Song. Y. U. (2017). Application of Queueing Theory to the Analysis of Changes in Outpatients' Waiting Times in Hospitals Introducing EMR. *Healthcare informatics research*. *23*(1). 35-42.

[19] Komashie. A., Mousavi. A., Clarkson. P. J. and Young. T. (2015). An integrated model of patient and staff satisfaction using queuing theory. *IEEE journal of translational engineering in health and medicine*. *3*. 1-10.