ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 15, Issue, 05, pp. 68298-68305, May, 2025
https://doi.org/10.37118/ijdr.29303.05.2025

RESEARCH ARTICLE OPEN ACCESS

OPTIMIZATION AND HOLD QUEUE DEMAND MANAGEMENT IN THE PUBLIC SANITARY UNITS IN MAPUTO CITY - (MOZAMBIQUE)

***Noe Santos Macuacua**

Master's Degree, Laulana - Maputo, Bairro 3 De Fevereiro Rua Da, Quarentinha, Maputo- Bairro 03 De Fevereiro, Moçambique

ARTICLE INFO

Article History:
Received 20th February, 2025
Received in revised form
25th March, 2025
Accepted 16th April, 2025
Published online 25th May, 2025

Key Words:
Simulation, Queues, Analytical Model, Satisfaction Level.

****Correspondingauthor:*** **Noe Santos Macuacua**

ABSTRACT

This scientific article was used to conduct a study on the optimization and management of queue demand in the public health sector in Mozambique. It was based on queue theory and parallel simulation between the analytical model. The simulation allowed the reproduction of complex systems, making it possible to visualize how they behave. The study was carried out in a discrete, cross-sectional comparative manner between April 19 and May 15, 2024, covering 218 users of four Health Centers in the City of Maputo, chosen intentionally with the aim of determining user satisfaction in relation to the management of waiting queues and associated factors. Waiting time management was analyzed in relation to the following indicators: waiting time, treatment and cordiality, understanding of the explanation for the reason for the delay in service in outpatient consultations in the triage, and pharmacy and laboratory services of each center. Satisfaction regarding waiting time management was classified into two levels: positive (satisfied) and negative (dissatisfied). Of the respondents, 67.89% were female and 32.11% were male, aged between 21 and 30 years old, and 50.5% were students. Regarding the indicators evaluated, the overall satisfaction of users regarding waiting time management was 45.10%, with the highest level being at the Polana Cimento Health Center (54.76%) and the lowest at the José Macamo Health Center (40.26%). Parameters such as availability of medicine and waiting time management recorded the lowest levels of satisfaction. An association was found between gender and the level of satisfaction regarding time management and social demographic variables, gender and academic level.

Citation: Noe Santos Macuacua. 2025. "Optimization and Hold Queue Demand Management in the Public Sanitary Units in Maputo City - (Mozambique)". *International Journal of Development Research*, 15, (05), 68298-68305.

INTRODUCTION

Mozambique is a developing country located in southern Africa. The country has a population of 36 million inhabitants and it is developing rapidly thanks to its natural resources. Despite the significant importance of healthcare in Mozambique, the country lacks an effective system of queuing and patient management within the healthcare sector, resulting in inefficiencies and suboptimal decision-making capabilities for healthcare providers and the Ministry of Health. This situation negatively impacts the overall quality of healthcare services and the patient experience, hindering the country's efforts to improve health outcomes and deliver high-quality healthcare to its citizens. The illiteracy rate is very high in Mozambique which creates many difficulties in basic health care and patient management to seek medical. Doctors attend several patients, but because of high number of cases, it is very difficult to manage the queues. Many diseases such as malaria affecting the population of Mozambique require good organization of public hospitals. Analysis of the historical trajectory of Mozambique's healthcare system reveals that, despite gaining independence in 1975, the country's National Health Service has been plagued by a chronic shortage of healthcare personnel.

This shortage, which can be traced back to the country's post-independence economic challenges and the subsequent migration of healthcare workers, has had a significant impact on the accessibility, quality, and coverage of healthcare services, posing a significant challenge to the healthcare system's ability to meet the health needs of its population. The impact of this shortage extends beyond mere staffing gaps and has far-reaching implications for the healthcare system's effectiveness in delivering quality care and disease prevention interventions. The inability to attract and retain sufficient healthcare personnel has resulted in overcrowding in healthcare facilities; long wait times, compromised quality of patient care, reduced access to preventative health services, and a general decline in the overall performance of the healthcare system. The set of human resource problems of the Ministry of Health in Mozambique can be highlighted in the following aspects:

- Acute shortage of systematic health professionals, particularly in remote areas where the workers are less motivated;
- Limited Health Training Institutions;
- Inefficient allocation of resources and deep imbalances in the composition and distribution of the workforce;
- Work overload;

- Lack of humanization of patient care;
- Inefficiency in the HR Management, that is lack of HR manual;
- Low wages, and
- Lack of career advancement

The central point for developing this research was mainly based on the experience the researcher acquired over time interacting with the patients while conducting inspections and visits in the general hospitals of José Macamo, Mavalane and Chamanculo located in the sub-urban and peri-urban zones of Mozambique. During these inspections and visits, it was observed that there were possibilities of circumstances where users spent almost the whole day in hospital. For example, a user would arrive at 05:00 AM and would possibly leave at 02:00PM. Some other users would end up abandoning the line after they waited so long in the sanitary unit. Thus, they would express their dissatisfactions. Unfortunately, these dissatisfactions would translate into "complaints" reported to the media of the country and through informal conversations. This research seeks to find mechanism to resolve this problem of long queues in the health facilities of the country. Among the many quantitative models is a simulation modeling. It is a numerical method that involves the development of a real system model, and conducting experiments in this model to understand the behavior of the system. Romanelli, Anchieta, Mourão, Campos *et al.* (2013), propose a theoretical discussion that has as its backdrop the dichotomy Methods Analytical versus Numerical Methods. Oliveira, Savelsbergh, Veelenturf, Van Woensel (2017) show that research was conducted through simulation models to analyze the queues for checkout at supermarkets and identify the optimal number of cashier stations required to minimize wait times and enhance customer satisfaction.

The goal of the study was to optimize the checkout process, thereby increasing customer loyalty and reducing the likelihood of shoppers abandoning their purchases due to excessive wait times. Zhang, Wang, Feng, Zhang *et al.* (2020), present a study where an application of delay analysis in railway systems and proposes a novel optimization methodology to minimize delay times. The methodology involves the use of advanced modeling techniques and sophisticated optimization algorithms to identify and address critical factors contributing to delay, with the ultimate goal of enhancing operational efficiency and maximizing customer satisfaction. This study will demonstrate the importance of modeling and simulation for various purposes and in particular for checking the simulation line models |M|M|1 and |M|M|C. The study will then compare the models obtained from simulation with analytical models, demonstrating that when applied appropriately, models for simulation can reproduce satisfactorily real systems with enough reasonable results, corroborating these complex situations.

METHODOLOGY

According to Malhotra, Nunan and Birks (2020), research can be broadly classified as exploratory or conclusive. And as a subdivision of conclusive research, there is descriptive research. In this study, mixed research was used in order to draw conclusions about the population under study. To achieve the objectives of the study, a questionnaire was designed to serve as an instrument for data collection. Quota sampling was used to collect data. This type of sampling is not probabilistic or random. According to Sudman (1966), in this type of sampling, the samples are obtained by dividing the population into categories or strata and selecting a certain number (quota) of elements from each category in a non-random manner. For this study, the aim was to select a sample that represented all the strata into which the user population was divided. Here, the strata represented the different services provided by the health units in outpatient consultations, which are: triage consultation, laboratory, treatment room, pediatric triage and pharmacy. The results from this sample needed to be reliable at a level of 95%, with the maximum estimate of the absolute error being 0.046 (4.6 percentage points). This was supposed to apply in the single domain of analysis, that is, the set of four Health Centers in the city of Maputo without considering their strata. In order to meet the objectives of this study, a probabilistic sampling plan was adopted and the sample selected in a single stage. In other words, the sample for this study was:

- **Probabilistic or Random**: the probability of selecting each user is known and is different from zero which allows estimating the precision and inferring the sample results;
- **Single-stage:** after creating and identifying the strata, the selection was made in a single stage, which consisted of obtaining the users for the interviews;
- **Stratified:** the users were divided into groups according to the services required in the health units.

Table 1. Sample Allocation and Distribution

Health Unit	Users served per week	Sampling fraction	Expected events in the sample	Expected interviews
Cs Chamanculo Annex	2825	32.37	91448	91448
Cs attachment by Jose Macamo	2221	25.45	56524	56524
Cs June 1st	2640	30.25	79862	79862
Cs of Polana Cement	1041	11.93	12418	12418
Total	8727	100	872700	872700

Source: Waiting Lines at Health Units (2024)

Table 2. Sample Quotas by sex in each Health Unit

Unidades Sanitarias	Women	Men	% Women	% Men	N◦ Number of Patients Interviewed	Sampling fraction
Cs Chamanculo Annex	31	20	22.1	25.6	51	23.4
Cs attachment by Jose Macamo	43	34	30.7	43.6	77	35.3
Cs June 1st	39	9	27.9	11.5	48	22.0
Cs of Polana Cement	27	15	19.3	19.2	42	19.3
Total	140	78	100.00	100.00	218	100

Source: Waiting Lines at Health Units (2024)

In the case of this study, taking into account the variability of the information sought and considering that most of the estimates in this survey will be presented in the form of averages and totals, it can be considered that $n = \frac{z_\alpha^2 \, Var(\hat{\bar{x}})}{e^2}$ which will be a sample size sufficient (sometimes more than sufficient) for the entire domain to achieve the desired accuracy and confidence level. In this case, taking into account that the users who visit the health units constitute an infinite population, the variance of the average service time is estimated to be $\sigma^2 = Var(\hat{\bar{x}})$ which is equal to the maximum value that the variance that could be obtained could take if it were to estimate the proportion of users that visit health units.

Data Collection From users: Before collecting data from the field, the author of the study carried out a pre-test with a sample of 5 users. This was in order to assess the consistency of the questionnaire and, where possible, to conclude if it was consistent with using a Cronbach's Alpha of 0.81. To do this, the instrument for recording responses - the questionnaire - was first used in collaboration with the National Institute of Statistics of Mozambique. The questions in the questionnaire were addressed to each user of the health unit by service. The questionnaire was divided into three sections, namely:

waiting time/queue time in the different services, quality of services provided and level of user satisfaction in the different services. The questionnaire was applied to voluntary users in 4 Health Centers, intentionally chosen in the peri-urban, suburban and urban areas of Maputo City. The clients were approached when they were purchasing passwords at the triage and were asked if they were available to answer the questionnaire.

Method used to select Health Centers: The study was carried out in the City of Maputo, which has an area of 300 km² and an approximate population of 1,120,736 inhabitants (NIS, Census 2007). This city has 3 General Hospitals and 12 Health Centers of the National Health System (NHS). The study was carried out from April 19 to May 20, 2024, and was carried out in the 4 Health Centers of the City of Maputo intentionally chosen. It should be noted that regarding the intentional choice, the author worked for a certain time in the chosen health units, which provided easy collection of information within the expected time and at lower costs. Although the choice of centers for the study would have been random, in order to simplify the study, a choice was made according to the following aspects:

- Area with the largest number of users;
- The data presented in the 2008 reports at the level of the Health Department of the City of Maputo; and
- The area of inhabitants.

The General Hospital of Mavalane has the largest area and the largest number of inhabitants seeking basic health services, so two centers were chosen to carry out the study. In the other health facilities, only one center was selected because the facilities were manageable in terms of size and population using.

Data collection: In order to obtain accurate results regarding the adopted model, data collection was carried out at different times, that is, on different days and times. In this way, it was possible to analyze the performance of the system variables considering peak or non-peak times, weekdays, and weekends.

Three specific moments were observed during data collection:

a) The time each patient arrived;
b) The waiting time in the queue for the care office in the different services;
c) The time spent waiting inside the office in the different services.

The data collection stage took place in 4 periods, namely (1) from April 19 to 24, 2024 from 7:30 AM to 3:30 PM (weekdays) and on Saturday from 7:30 AM to 12:00 PM at the Chamanculo Annex Health Center (peri-urban area); (2) from April 26 to May 1, 2024 from 7:30 AM to 3:30 PM (weekdays) and on Saturday from 7:30 AM to 12:00 PM at the José Macamo Annex Health Center (peri-urban area); (3) from 3 to 8 May 2024 from 7:30 AM to 3:30 PM(weekdays) and on Saturday from 7:30 AM to 12:00 PM at the Polana Cimento Health Centre (urban area) and, (4) at the 1 de Junho Health Centre (suburban area) from 10 to 15 May 2024.

Theoretical: The problem around queue management which has devastated the international community has been mostly faced in public healthcare facilities characterized with a large influx of patients. This problem was studied for the first time by (ERLANG, 1913), associated with an area of knowledge called "Queuing Theory" to balance the costs of provision of good service with cost of having customers waiting (ENTRINGER, 2020).

Waiting line: Queues are common in our day-to-day. Callers to tele-care or go to banks or cinema are all subject to queuing to be served. The queuing theory uses a mathematical concept of queues. The formation of these is a common event that occurs whenever the current demand for a particular service exceeds the current capacity to meet this service (SEN; ZHOU; ZHANG; YOON *et al.*, 2002). To specify the queuing system, a pattern that represents all the features that can be specified in this system was created. This pattern is called Kendall notation and consists of six parameters. Their representation has the following form:

A / S / M / B / K / SD

Where each parameter can be specified as follows.

The parameter 'A' is the arrival process which specifies the distribution to be respected to know the behavior of the arrival processes. The most widely used arrival process is called Poisson arrival which means that the interval between arrivals is Independent and Identically Distributed (IID) subject to an exponential distribution. The service time is the time the system spends in responding to a request. It is common that the distribution used in this parameter is also IID. The distribution of service time is represented by the variable 'S'. The number of servers is represented by variable 'M', which tells the total server that the queue system has to meet customer requests. Variable 'B' represents the system's maximum capacity which is the highest number of requests the system can support. The size of the population, defined by the variable 'K', is the maximum number of clients that can potentially send requests to the system. In most real systems, the population is limited. However, to facilitate the calculations, the queuing theory assumes that in most cases, the population is infinite. The 'SD' service discipline is how requests are met. The most widely used discipline is the 'First Come First Served' (FCFS) where requests are attended to in the order of arrival. When the system has used its parameters 'B' and 'K' defined as infinite and the service discipline as FCFS, only the first three Kendall notation parameters are required. In this case, to specify a system with three servers and arrival processes and service discipline respecting an exponential distribution, the notation to be used is M / M / 3. The most common line types are the following:

- ✓ M / M / 1: System that has only one server and inter-arrival times and service following the exponential distribution.
- ✓ M / M / m: System that has servers and exponential distribution for times between arrival and service.
- ✓ M / M / M / B: Same as the previous system, with the difference that only a B number of requests may be present in the system at the same time.
- ✓ M / G / 1: This the system applied in this work. It has only one server, the arrival processes respect the exponential distribution and service time follows a general distribution.

Simulation: The simulation has been studied and disseminated to the study of complex events. In this work, the simulation was used in order to analyze a queuing system, but the application would be much more comprehensive. Since the end of the twentieth century, thanks to the development of computing resources, simulation has been increasingly used in various areas – from the simulation of complex events to the most popular events. Currently, there are several simulation software applications such as Crystall Ball, @RISK, Decision Ro, Xcell, SLAM, Witness, Arena and MAP / 1 (ERLANG, 1913).According to (SEN; ZHOU; ZHANG; YOON *et al.*, 2002), simulation is a widely used method and progressively popular for the study of complex systems. There is often a set of assumptions about how a system works. To verify simulation, these hypotheses are used:

Emulation with its advantages and disadvantages: According to the perspective of (MORéT-FERGUSON; LAW; PROSKUROWSKI; MURPHY *et al.*, 2010), the use of simulation systems implies advantages and disadvantages, and pros and cons of its use as shown below.

Regarding Advantages

- The most complex systems in the real world with non-stochastic elements can be accurately described by a mathematical model with analytical evaluation;

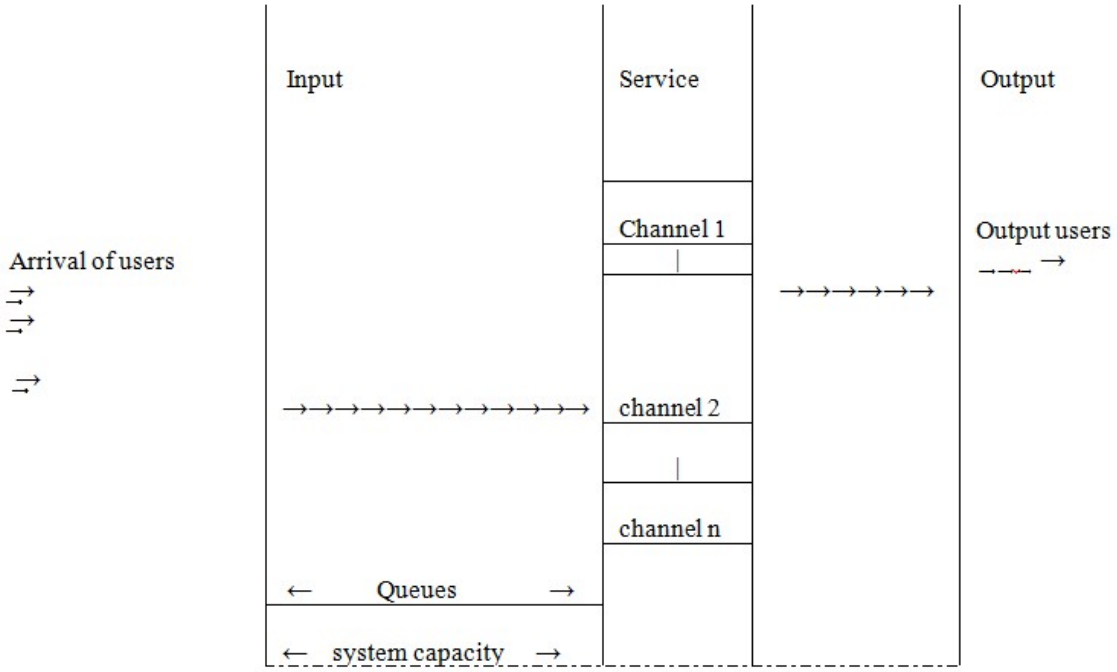
- The simulation often allows us to estimate the performance of existing systems designed under specific operating conditions;
- Other possible alternatives can be compared via simulation to see which best meets specific requirements;
- Simulation allows for a greater degree of experimental control, enabling the researcher to precisely manipulate conditions and parameters that would be challenging or impossible to control in the actual system. This level of control enables the researcher to investigate complex phenomena and obtain reliable, comprehensive insights into the system's behavior under various conditions.
- Simulation models, when properly designed and calibrated, can closely replicate the behavior of analytical models derived from theoretical principles. As such, they can serve as valuable tools for verifying and refining analytical models, as well as for exploring complex system behavior that may not be easily captured by analytical methods alone.

Regarding Disadvantages

- The execution of a stochastic simulation model produces estimates of the true features of a model for a particular set of input parameters. Therefore, several independent replications of the model can be obtained and will probably be required for each set of input parameters to be studied;
- Simulation models have always been very expensive and time consuming;
- If the simulation results, no matter how impressive look, provide little useful information about the real system, the model is not a "valid" representation of a system under study.

the system provides service to users, including processing time and probability of failure, and queuing discipline which is the rules or algorithms for determining which user gets served first and how users move through the queue. In queue system, the wait occurs when demand for a service is higher than that offered service capacity in terms of flow (SINAY, 2004). Therefore, if the service capacity of a queuing system is lower than the arrival rate of incoming requests, an accumulation of entities in the queue occurs, leading to delays or waiting times. This condition is known as an under-provisioned queuing system, and the server cannot process the incoming requests as fast as they arrive, thereby causing a backlog. Figure 1 is a schematic representation of a queue system M | M | C. through this representation, we observed the fundamental elements that make up the queue and how these relate. The queuing system is characterized by: users and population size, arrival process, service process, number of users, queue discipline, average queue length, maximum queue length, and average wait time in queue (PRADO; NATALE, 2004).

A) **Users and Population:** A User is part of a population. It is any individual that seeks service or access to a resource provided by a system. When the population of users is too large, the arrival of new users does not affect the arrival rate of subsequent users. Therefore, is can be concluded that arrivals are independent.
B) **Process of arrival:** According to (RAGSDALE, 2009), the time between two consecutive arrivals is called inter-arrival time. If the number of arrival follows a Poisson distribution with a mean λ, it can be demonstrated that the time between arrivals (TEC) follows an exponential distribution with mean $1 / \lambda$.



Source: Sinay (2004)

Figure 1. Schematic representation of a queuing system M | M | C

Queuing Theory: The formal study of queues, commonly known as queueing theory, was initiated in 1909 by Danish Mathematician, A. K. Erlang, who devised a theoretical framework for understanding and analyzing the behavior of queueing systems. The mathematical approach to queues started at the beginning of the 20th Century but it was only around 1950 after the Second World War that the queueing theory was applied (SEN; ZHOU; ZHANG; YOON et al., 2002). Queue system is any process where users come to receive any desired service. A queue system has three key components, namely; arrival process which is the pattern or rate at which users arrive at the system for the desired service, service process which is the manner in which

C) **Care process:** This is the elapsed time between the beginning of the service until the end and is called 'call time' (TA) (HILLIER; LIEBERMAN, 2013);
D) **Number of Servers:** Refers to the physical quantity and simultaneous supply of personnel and equipment available to attend to users.
E) **Discipline Queue:** Refers to the rules that govern how users are processed or attended to in a queueing system. One of these rules is one that defines the next user to be served commonly known as "first-come-first-served "(FIFO). There are several other disciplines and these different queue

disciplines can have a significant impact on the system's performance and the experiences of users in the queue.

F) **Average size of the queue:** The queue size, also known as the queue length, is the number of users waiting in a queue for service at any given point in time. It is an important metric for understanding the performance of a queueing system, as it can provide insight into the level of congestion and delays experienced by entities in the system. The queue size is not constant when the average rates of arrival and attendance are constant; the queue size oscillates around a mean value.

G) **Maximum size of the queue:** The systems are dimensioned for a certain maximum amount of waiting customers.

H) **Average waiting time in the queue:** The average waiting time depends on the processes of arrival and service. To adequately describe a queue system we use the notation of (KENDALL; HILL, 1953), which is represented as follows: A | B | C | D | E where A and B represent time between arrivals and service time (TA), C and D denote the number of service stations in parallel and the physical capacity of the system, and E is the discipline of service employed (LECOURT; HERAULT; PEARCE; SOLLOGOUB *et al.*, 2004)

Queuing model M | M | 1

In Model M | M | 1,is a single-server queueing model that assumes that the arrivals of entities (customers or data packets) follow a Poisson distribution (M) and that service times are exponentially distributed (M). The "1" indicates that there is only one server/attendant in the system(PRADO; NATALE, 2004). The most important probability distribution in the theory of queues is the exponential distribution. This is a type of continuous probability distribution, represented by a parameter, and it is one of the main properties being a memory less process where the time until the next arrival does not suffer any influence of the latest arrival. So assuming a random variable, T represents time between arrivals and service times. It is said that this variable has an exponential distribution with parameter α if its density function probability is: (KENDALL; HILL, 1953).

$$f(x,\lambda) = \begin{cases} \alpha e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad \text{……..(1)}$$

The cumulative probabilities are:

P (x) () = $1 - e^{-\lambda x}$ ….……..(2)

X ≥0,

P () = $e^{-\lambda x}$ ...…….(3)

And the expected value and variance of X are respectively:

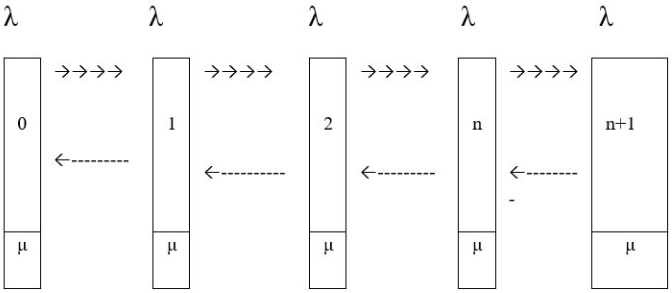
$$E(x) = \frac{1}{\alpha} \qquad \text{……..(4)}$$

$$\text{Var}(x) = \frac{1}{\alpha^2} \qquad \text{……...(5)}$$

The M | M | 1 | ∞ | queueing model is similar to the M | M | 1 model, but with the added assumption of infinite queue capacity. That is, the system can accept an infinite number of users into the queue without limit. In the model M | M | 1 | ∞ |, FIFO times between successive arrivals and service times follow exponential distributions. Arrivals and calls feature a process of birth and death; meaning that arrival and service rates are constant and there is a single service station and the order of service for users follows the order of arrivals (SINAY, 200), (KENDALL; HILL, 1953) state that the number of users arrive at the M line | M | 1 by minute is described by a parameter λ Poisson distribution, and service time for the μ parameter. Where λ is the average arrival rate of users per unit time and μ is the average attendance per unit time, and system load factor is represented by the following equation:

$$\rho = \frac{\lambda}{\mu} \qquad \text{…….. (6)}$$

The system in question can be considered stable. It is required that λ is less than μ or ρ <1. When we have ρ close to 1, the queue tends to increase infinitely (PRADO, 2004). Figure 2 below refers to a line. M | M | 1 | ∞ | FIFO is the flow and the correlation of time between arrivals and service time.



Fonte: Sinay (2004)

The study of queueing theory involves analyzing the dynamics of queues and their associated performance measures, which provide valuable insights into the efficiency and stability of queuing systems. By identifying the key performance measures, such as queue length, waiting time, and utilization, researchers can evaluate the impact of various system characteristics on the queuing system's ability to meet the demands of its users.

Table 3. Presents the formulas for calculary key performance measurements for a queuing system M | M | 1.

Parameters	Symbols	Formula	
Expected number	L	$\frac{\lambda}{\mu - \lambda}$	(7)
Expected number in queue	Lq	$\frac{\lambda^2}{\mu(\mu-\lambda)}$	(8)
Expected waiting time (includes service time)	w	$\frac{1}{\mu - \lambda}$	(9)
Expected time in queue	Wq	$\frac{\lambda}{\mu(\mu - \lambda)}$	(10)
Probability that the system is empty	Po	1- $\frac{\lambda}{\mu}$	(11)

Source: Barbosa (2009) apud adapted from Moore & Wheatherford (2005)

Line model M | M | C: This model, also known as the "multi-server model," is similar to the M | M | 1 model, but with more than one server (or multiple servers that process the users in parallel). This means that model M | M | C is the model in which there is a single row and multiple servers. Both the arrivals and the service are Marcovianos, and the utilization rate in the system is represented by the equation (PRADO, 2004):

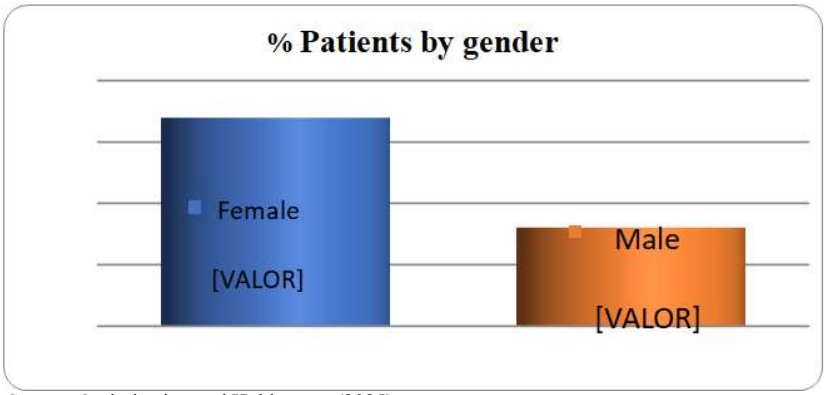
$$\rho = \frac{\lambda}{s\mu} \qquad \text{……..……(12)}$$

Where 'λ' is the rate of arrival, 's' the number of servers and the rate of 'μ' attendance. As the traffic intensity parameter ρ approaches 1 in the M | M | 1 queueing model, the system becomes saturated, resulting in an infinite queue length. This phenomenon is known as "queue explosion" or "instability" and leads to unacceptably long waiting times for users (PRADO, 2004). In this model, basically, we have the same performance measures of the model M | M | 1 but with different calculation formulas, as shown in Table 4.

Table 4. Model parameters M / M / C

Parameters	Symbols	Formula	
Probability that the system is empty	Po	$\frac{1}{\sum_{s=0}^{c=1} \frac{\rho^s}{f!} + \frac{\rho^2}{(s-1)!(s-\rho)}}$	(13)
Probability that all channels are busy	Pbusy	$\frac{\rho^s}{(s-1)!(s-\rho)}\rho 0$	(14)
Expected number on queue	Lq	$\frac{p_o(\frac{\lambda}{\mu})^s}{s!(1-p)^2}$	(15)
Expected number in systems	L	Lq + ρ	(16)
Estimated time in line	Wq	$\frac{Lq}{\lambda}$	(17)
Expected waiting time (including service)	W	Wq + $\frac{1}{\mu}$	(18)

Source: Barbosa (2009) apud adapted from Moore & Weatherford (2005)



Source: Optimization and Hold queue (2025)

Figure 3. Patients by Gender

SPSS Software and Arena Software for analyzing results: SPSS software is a powerful computer tool that allows users to perform complex statistical calculations and view their results in a matter of seconds. However, two obstacles stand between the user's good intentions and their goal: knowing which statistical test to use to answer their questions; and correctly interpreting the results of the statistical calculations performed (AZEN; WALKER, 2021). Arena Software is one of the most widely used simulation software programs due to its wide range of scenario configuration possibilities, and ease of use and interpretation of results. It can be understood as a simulation language that offers a work environment suitable for testing, with several analysis tools and advanced animation resources. Its use allows managers to simulate different scenarios involving the interaction of various elements of the process under study, such as people, equipment, inputs, raw materials and rules of behavior. With the software, managers can test new ideas and projects, in addition to predicting the results, thus assisting in decision-making. Another advantage of the software is that it can be used in different activities and processes, allowing simulation from the "factory floor" to the front office of organizations, and can therefore be used in different environments. The Arena Software has a graphical interface based on MS Office standards, with commands, buttons and menus that provide functions similar to those available in other Windows software. Another advantage, therefore, is its ease and operability in use, where modeling is done visually with simulation-oriented objects and only with the help of the mouse, without requiring any programming logic commands (AZEN; WALKER, 2021).

DISCUSSION OF RESULTS

The questionnaire survey was administered to 218 users of health centers across Maputo City, specifically chosen to represent a cross-section of urban, peri-urban, and suburban areas, in order to gather a detect some lines of sociographic and professional distinction between them. It was clear that, in terms of gender, in a set of 218 valid cases, 67.89% were female and 32.11% were male. Before proceeding with the study on time management, we observed whether the time between arrivals generally followed an exponential distribution. As for normality (Kolmogorov-Smirnov (K-S)) by(FILION, 2015), independence is not satisfied because there is clear evidence that leads us to conclude that the appropriate model to use in the study will be M/G1. Average time it takes for an employee to attend to 1 (one) patient upon acceptance.

$\sigma^2 s = V(s) = 0.36$, variance $\lambda = 11\ patiente/h$, users arrive at an average of 11 per hour in acceptance

μ = 1(Patient)/ 3,3878 (1hour/60min) = 18 Patients/h

$\rho = \frac{\lambda}{\mu}$, 11patients/ 18patients/h = 0.61 the occupancy rate is not explosive ρ<1.

Total time in the system for users seeking services from acceptance, pediatric triage, laboratory and pharmacy:

120 + 33 + 38 + 54= 245 min = 4 h of time that the patient takes until leaving with the services provided.

Probability of exactly two patients being in the system, one being treated and the other waiting in line:

$Pn = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n$, $Pn = \left(1 - \frac{11}{18}\right)\left(\frac{11}{18}\right)$ 2= 0.144969

The capacity of a system is directly related to the increase in time, that is, the closer capacity utilization gets to one hundred percent, the more the waiting time increases rapidly.

Table 5. One-Sample Kolmogorov-Smirnov Test

N		214
Exponential parameter.	Mean	3.74
Most Extreme Differences	Absolute	0.23
	Positive	0.13
	Negative	-0.23
Kolmogorov-Smirnov Z		3.43
Asymp. Sig. (2-tailed)		0.00

Source: Optimization and Hold queue (2025)

Table 6. Study of waiting time management in different services at each interview location (Health Units)

	N	Minimum	Maximum	Mean	Std. Deviation
Aceitação	49	1	8	3.3878	0.36

Source: Optimization and Hold queue (2025)

Systems Modeling: In order to improve system performance, system modeling allows some participating variables to be evaluated. System modeling consists of two techniques widely used to analyze processes in which queues are formed: queueing theory and simulation. As a result of the application of system modeling, it is expected that the system will present an optimized and efficient operation, that is, that costs will be adequate and users will be satisfied with the services provided in the Health Units. Below, three scenarios are presented that were analyzed with different variables and simulation time. For this first simulation, only one queue with one employee was considered. The results of scenario 1 are presented in the table above.

Table 7. Results of Scenario 1

Simulation time	720 Minutes (12 hours)
Average queue time	0.16
Time the employee was busy	0.55(55%)
Average number of users in the queue	5
Users served	218
Average time between the user's arrival in the queue and departure with the services provided	1.55 Min
Minimum time between the user's arrival in the queue and leaving with the service provided	0.8 Min
Maximum time between the user's arrival in the queue and leaving with the service provided	3.14 Min

Source: Optimization and Hold queue (2025)

Scenario 2: For the second scenario, the same time values were used and also only one queue, but with 2 attendants. After the simulation considering the descriptive configuration, the following results were obtained.

Table 8. Results of Scenario 2

Simulation time	720 Min (12h)
Average queue time	0.00
Time the employee was busy	0.55(55%)
Average number of users in the queue	0
Utentes atendidos	218
Average time between the user's arrival in the queue and departure with the services provided	1.49 Min
Minimum time between the user's arrival in the queue and leaving with the service provided	0.8 Min
Maximum time between the user's arrival in the queue and leaving with the service provided	2.18 Min

Source: Optimization and Hold queue (2025)

Scenario 2: This scenario considered the configuration of the model with 2 independent queues, with each queue containing 1 service. The service rate remained the same as in the previous scenarios (0.75 minutes, 1.5 minutes and 2.25 minutes), but since each queue will receive half the number of users compared to the previous scenario, the arrival rate was divided by 2, receiving the values 0.3 minutes, 1.25 minutes and 2.5 minutes.

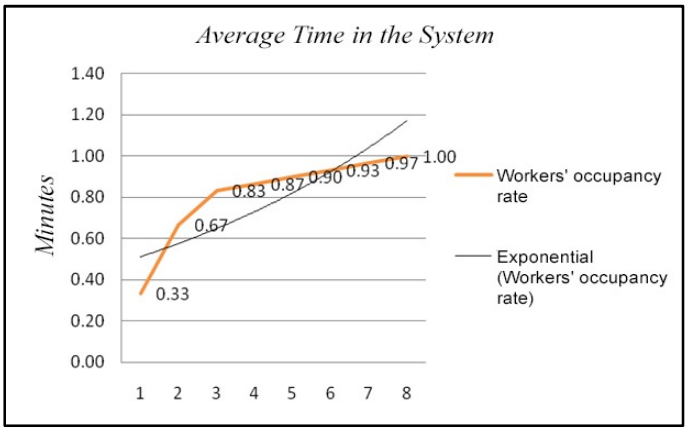
Scenario 3:

Table 9. Results of Scenario 3

Simulation time	720 Min (12h)
Average queue time 1	39.46 Mi
Average queue time 2	37.74 Min
Average queue time	0.16
Time the employee was 2 was busy	1.00 (100%)
Average number of users in queue 1	28.13
Average number of users in queue 2	30.10
Users served	961
Average time between the user's arrival in the queue and departure with the services provided	40.48 Min
Minimum time between the user's arrival in the queue and leaving with the service provided	1.38 Min
Maximum time between the user's arrival in the queue and leaving with the service provided	86.15 Min

Source: Optimization and Hold queue (2025)

Observations on the results and suggestions: Regarding the observation of the results, it can be obesrved that there was little change in the results of scenarios 1 and 2. The main change concerns the average time between the arrival of the user in the queue and the departure with the services provided. However, since the change was only a 6-second improvement in relation to 1 and 2 employees as attendants, it can be concluded that it is not worthwhile to bear the costs by keeping 2 employees, since the user's time in the queue will decrease by only 6 seconds on average. In relation to scenario 3, with 2 employees serving in 2 independent queues, an average queue time of 38 minutes was observed. On top of that, 961 users were served in the 2 queues, compared to only 218 users with 2 employees in a single queue. Therefore, it can be concluded that during peak times, scenario 3 is more appropriate, since it allows for greater service. If the manager concludes that during peak hours, the average queue time (38 minutes) is too high or that the queue is too long (an average of 28 users in queue 1 and an average of 30 in queue 2) and that users will therefore have a negative perception of the service, new experiments can be carried out with the adoption of an additional employee in the system, or new alternatives for configuring different queues can be evaluated.



Source: Optimization and Hold queue (2025)

Figure 4. Average time in the System

To solve the problem of queue formation,(PETERSEN; SCHMENNER, 1999) understands that there are multiple ways to manage it: increasing capacity, investing in statistical quality control or adopting any other measures that guide the reduction of service variances. (FITZSIMONS, 2000) state that the phenomenon of queue formation occurs when demand exceeds service capacity. This occurs when the arrival time of a new customer is greater than the service and service delivery time. (CORRÊA; GIANESI, 1994) emphasize that the service provider must manage queue formation so that the

customer does not wait too long to be served, otherwise it may have a negative impact on service quality. Therefore, the services provided must be offered with the minimum quality assurance so that the customer has the most positive perception possible. Queue management deserves to be studied by service managers due to its influence on the formation of users' perception of service quality, directly interfering with user satisfaction in health units. Inadequate queue management may result in users waiting for a long time in the queue, making them unhappy, with the feeling of being "left aside" and that their time is less important than that of the service provider. Therefore, service managers must understand the need for effective queue management, adopting mechanisms that result in reducing the feeling of waiting time for the customer in the queue, preventing them from developing a negative perception of the quality of the services provided. Through the use of computer simulation programs, managers can simulate different scenarios to study the behavior of variables during the service process. In this sense, system simulation can allow the identification of factors responsible for the formation of queues so that the manager can seek solutions aimed at optimizing processes.

REFERENCES

AZEN, R.; WALKER, C. M. Categorical data analysis for the behavioral and social sciences. Routledge, 2021. 0429330308.

CORRÊA, L. H.; GIANESI, I. G. Qualidade e Melhoria dos Sistemas de Serviços. _____. Administração Estratégica de Serviço, São Paulo: Atlas, p. 195-207, 1994.

ENTRINGER, T. C. Simulation and analysis of queues in banks: a case study of an agency in the southern State of Rio de Janeiro. Independent Journal of Management & Production, 11, n. 3, p. 892-907, 2020.

ERLANG, A. New alternating-current compensation apparatus for telephonic measurements. *Journal of the Institution of Electrical Engineers,* 51, n. 222, p. 794-799, 1913.

FILION, G. J. The signed Kolmogorov-Smirnov test: why it should not be used. Gigascience, 4, n. 1, p. s13742-13015-10048-13747, 2015.

FITZSIMONS, G. J. Consumer response to stockouts. *Journal of consumer research,* 27, n. 2, p. 249-266, 2000.

HILLIER, F. S.; LIEBERMAN, G. J. Introdução à pesquisa operacional. McGraw Hill Brasil, 2013. 8580551196.

KENDALL, M. G.; HILL, A. B. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General),* 116, n. 1, p. 11-34, 1953.

LECOURT, T.; HERAULT, A.; PEARCE, A. J.; SOLLOGOUB, M. *et al.* Triisobutylaluminium and diisobutylaluminium hydride as molecular scalpels: the regioselective stripping of perbenzylated sugars and cyclodextrins. *Chemistry–A European Journal,* 10, n. 12, p. 2960-2971, 2004.

MALHOTRA, N. K.; NUNAN, D.; BIRKS, D. F. Marketing research. Pearson UK, 2020. 1292308761.

MORÉT-FERGUSON, S.; LAW, K. L.; PROSKUROWSKI, G.; MURPHY, E. K. *et al.* The size, mass, and composition of plastic debris in the western North Atlantic Ocean. Marine pollution bulletin, 60, n. 10, p. 1873-1878, 2010.

OLIVEIRA, A. S.; SAVELSBERGH, M. W.; VEELENTURF, L.; VAN WOENSEL, T. Crowd-based city logistics. 2017.

PETERSEN, C. G.; SCHMENNER, R. W. An evaluation of routing and volume-based storage policies in an order picking operation. Decision Sciences, 30, n. 2, p. 481-501, 1999.

PRADO, R. D. M.; NATALE, W. Calagem na nutrição de cálcio e no desenvolvimento do sistema radicular da goiabeira. Pesquisa agropecuária brasileira, 39, p. 1007-1012, 2004.

RAGSDALE, S. W. Nickel-based enzyme systems. *Journal of Biological Chemistry,* 284, n. 28, p. 18571-18575, 2009.

ROMANELLI, R.; ANCHIETA, L. M.; MOURÃO, M. V. A.; CAMPOS, F. A. *et al.* Risk factors and lethality of laboratory-confirmed bloodstream infection caused by non-skin contaminant pathogens in neonates. Jornal de pediatria, 89, p. 189-196, 2013.

SEN, S.; ZHOU, H.; ZHANG, R.-D.; YOON, D. S. *et al.* Amplification/overexpression of a mitotic kinase gene in human bladder cancer. Journal of the National Cancer Institute, 94, n. 17, p. 1320-1329, 2002.

SUDMAN, S. Probability sampling with quotas. Journal of the American Statistical Association, 61, n. 315, p. 749-771, 1966.

ZHANG, M.; WANG, L.; FENG, H.; ZHANG, L. *et al.* Modeling method for cost and carbon emission of sheep transportation based on path optimization. Sustainability, 12, n. 3, p. 835, 2020.

****