Yugoslav Journal of Operations Research
xx (20xx), Number xx, xxx–xxx
DOI: https://doi.org/10.2298/YJOR240316015T

Research Article

QUEUING THEORY & DISCRETE SIMULATION AS A TOOL TO IMPROVE MEDICINE DELIVER CENTER SERVICE LEVELS

Maria Alexandra HINESTROZA TORRES
Faculty of Engineering, Corporación Universitaria Comfacauca, Colombia
mariahinestroza@unicomfacauca.edu.co, ORCID: 0009-0001-3681-7050

Camila RUIZ MORCILLO
Faculty of Engineering, Corporación Universitaria Comfacauca, Colombia
camilaruiz@unicomfacauca.edu.co, ORCID: 0009-0002-7397-3228

Yesid Ediver ANACONA MOPAN*
Faculty of Engineering, Corporación Universitaria Comfacauca, Colombia
yanacona@unicomfacauca.edu.co, ORCID: 0000-0001-7255-8503

Helmer PAZ OROZCO
Faculty of Engineering, Corporación Universitaria Comfacauca, Colombia
hpaz@unicomfacauca.edu.co, ORCID: 0000-0002-1270-7981

Mario CHONG
Faculty of Engineering, Universidad del Pacífico, Perú
m.chong@up.edu.pe, ORCID: 0000-0002-1231-0992

Ana LUNA
Faculty of Engineering, Universidad del Pacífico, Perú
ae.lunaa@up.edu.pe, ORCID: 0000-0002-1359-2469

Irineu DE BRITO JUNIOR
Faculty of Engineering, University State Paulista, Brasil
irineu.brito@unesp.br, ORCID: 0000-0003-2977-6905

Received: March 2024 / Accepted: January 2025

*Corresponding author

Abstract: The increasing demand for medical services and supplies as medicine delivery has been found in applied mathematical models, such as queuing theory, a tool to analyze and improve medical processes. However, these models often assume specific arrival intervals and demands without real variability and uncertainty demand. This research proposes a methodology in real systems, considering queuing theory and discrete simulation. The goal is to provide reliable scenarios and facilitate the decision-making process in medical centers. The data were collected in drug dispensing centers in Colombia as service times, service level and users'arrival. The data were processed in a simulation model by Flexsim with multiple scenarios. The results revealed a 50% reduction in average waiting time, identified staff constraints and critical processes, and improved resource allocation. This applied case demonstrated that the queue theory and simulation integration become a valuable tool for managing variability in this medical service delivery system.

Keywords: Queuing theory, discrete simulation, waiting time.

MSC: 68N15, 68M20, 68U20.

1. INTRODUCTION

Queuing theory, a fundamental mathematical model for analyzing congestion and delays in waiting systems [1], focuses on the interaction between clients and servers [2]. From its initial conceptualization by Erlang in the field of telecommunications to its application in contexts as diverse as inventory management, traffic planning or healthcare [3], this discipline offers a precise and powerful analytical framework to address mathematical studies of queuing and efficient resource management in systems where elements [4], whether people, products, or information, compete for access to a limited service [5].

In the Colombian context, this theory is especially relevant for those who face long queues for various services, such as medical care, administrative procedures and banking services. For example, according to data from the National Health Institute, the average waiting time for primary health care is approximately 2 hours [6]. Likewise,the Superintendence of Notaries and Registries estimates that 70% of Colombians experience delays in administrative procedures [7], while more than 50% of bank customers report waiting times of more than 30 minutes in branches [8]. These figures show the significant impact of queues on the population, generating widespread inconvenience and harm.

Nevertheless, considering that the population under study is part of the drug dispensing centers, reference is made to some of the limitations, such as an initial maturity medical system, inadequate health infrastructure, qualified personnel shortage, and a non-integrating local and regional network [9], [10]. According to the Colombian Health Secretary's data, the country has approximately 4,329 drug supply centers, attending 43,159,524 General Social Security System in Health affiliates (SGSSS), representing 96% of the population, with an average of 10,000 affiliates/supply center [11]. This scenario reflects insufficient resources to align the medical supply with the growing service demand [12]. As a result, delaying medical procedures, aggravating medical disease situations, and increasing queues in every process has a bullwhip effect in this socio-technical complex system [13]. Therefore, one of the strategies is to increase the number of dispensing points, but this implies additional costs for the Health Promoting Entities (EPS) [14]. The challenge is balancing service quality and operating costs [15].

Within the field of queuing theory, there are control algorithms that are essential to efficiently manage waiting systems [16]. These queuing algorithms determine the order in which processes access system resources, prioritizing those with longer waiting times, which directly influences operational efficiency and customer satisfaction [17]. Ranging from basic policies such as FIFO (first in, first out) to more sophisticated approaches such as SJF (shortest job first), these algorithms optimize the distribution of resources, thus contributing to maximizing the performance of the system as a whole [18]. In addition, queuing theory provides key performance indicators that enable informed decision making, which is crucial in business environments where efficiency and service level are priorities [19].

Several methodologies have been proposed concerning this challenge and informed decision-making [20]. The queuing theory model presents key performance indicators (KPI) [21] as the system customer arrival rate (λ), the service rate (μ), the server utilization rate (ρ) as $\rho = \lambda/\mu$ and a desired service level: $S = \lambda * W$, where W is the maximum queue waiting time [22], acceptable waiting times, arrival and service levels.

Previous research demonstrated queuing theory applications in different industries and business environments, such as logistics, telecommunications, customer service, finance, and medicine [23]. Logistics focused on two queue line models and found the optimal service rate and economic viability, favoring the M/M/1 model [24]. Telecommunications analyzed the utilization of servers using the M/M/c queue model to meet the demand for cloud services and reduce waiting times [25]. In the customer service area, explore work-in-process demand and cycle time behavior in a service system, employing the analytical model of M/M/c queuing theory and Arena to accurately service system [26].

However, many authors have emphasized the queuing theory's usefulness in healthcare management [27]. This theory has accurately addressed healthcare principles highlighted in Pakistan's healthcare system and challenges [28]. In addition, queuing theory and blockchain cryptocurrency modeling [29] improve the public healthcare system and patient waiting times [30] optimize personal allocation in a service system [31], queuing in emergency departments [32] queuing in pharmacy services [33].

Other authors consider combining the queuing theories model with other mathematical models [34], [35]. This modality has proven helpful in addressing the management of complex processes and services in various industries [36]. However, the theory model excludes several behaviors, as assumed service times and user arrivals are constant, even though there is variability in both inter-arrival intervals and time windows throughout the day [37], [38]. In healthcare management, hybrid approaches that combine discrete simulation with other methods have proven to be particularly effective. For example, the article entitled "A Novel Simulation-Based Two Stage-Optimization Approach for Nurse Planning", which aimed to develop a novel simulation-based two-stage optimization approach to determine the required number of nurses and schedule work shifts in the Covid-19 inpatient service of a Turkish public hospital [39]. Another relevant study is "A Comprehensive and Integrated Hospital Decision Support System for Efficient and Effective Healthcare Services Delivery Using Discrete Event Simulation", whose main objective was to develop a comprehensive hospital-level decision support system to assess and respond to the needs of local populations. These studies provide valuable methodologies and tools that can be adapted and applied in various healthcare institutions to improve resource planning and decision making [40].

This article aims to explore the theoretical foundations of queuing theory [15], as well as its application in the practical case of a medicine supply center that presented failures in queue management [41], generating dissatisfied users due to the waiting service and creating stress in the workers of the service modules. Due to this disparity between capacity and demand, queues are generated in the system, reflecting a shortage of resources to meet the growing demand for health services [42]. Therefore, this main contribution of this study is to propose a methodology that links queuing theory with simulation models and incorporating variability in natural systems. The variability is an intrinsic characteristic in many systems, and its inclusion can generate more accurate and reliable results, facilitating decision-making.

Finally, the rest of the paper is organized in the following manner: Section 2 presents the literature review, Section 3 materials and methods, Section 4 defines the results achieved, and Section 5 summarizes the conclusions and future work prospects.

2. LITERATURE REVIEW

Queuing theory is a crucial field in waiting systems and stochastic processes research. From its initial conception applications in various disciplines, queuing theory offers valuable tools and approaches to understand and improve the system's efficiency with waiting components. In this section we conducted a comprehensive literature review on queuing theory and found 69,756 documents in Scopus, 1,282 focused on medicine. This review covered the period from 1900 to 2023 and revealed studies on queue congestion problems, operations improvement, and service level reduction, especially in smaller systems [43].

2.1. Theoretical foundations

Since the 20th century, queuing theory has experienced exponential growth and diversification in its application [44]. Erlang applied the design of telephone exchanges, traffic engineering, and service management [45]. The M/M/s (arrival rate/service rate/number of servers) describes systems with simultaneously handle tasks [46] and introduces critical elements such as competition between servers and the ability to serve multiple clients concurrently [47]. This model simplifies relevant KPIs (key performance indicators), such as customers' average number in the system L_s, no customer probability in the queuing system $P_o)$ average time spent in the system W_s, average users in the queuing L_q, average time in the line W_q and service utilization rate (ρ) [48]. These variables are calculated from the following equations:

$$L_s = L_q + \frac{\lambda}{\mu}(L_s = \lambda W_s) \tag{1}$$

$$P_o = \frac{1}{\sum_{n=0}^{s-1} \frac{1}{n!}(\frac{\lambda}{\mu})^n + \frac{1}{s!}(\frac{\lambda}{\mu})^s(\frac{1}{1-\rho})} \tag{2}$$

$$W_s = W_q + \frac{1}{\mu} \tag{3}$$

$$L_q = \frac{(\frac{\lambda}{\mu})^s \lambda \mu}{(s-1)!(s\mu - \lambda)^s} \tag{4}$$

$$W_q = W - \frac{1}{\mu} = \frac{Lq}{\lambda} \tag{5}$$

$$\rho = \frac{\lambda}{n\mu} \tag{6}$$

Regarding performance, it is relevant to note that when the multiplication of the number of servers by the service rate (nμ) exceeds the arrival rate λ, the system is in a stable regime. However, when (nμ) is less than or equal to λ, the system can collapse, and the queue can increase without limits, as pointed out by various sources [49]. As for arrival λ and service μ rates, these can be modeled by continuous or discrete distributions, depending on the nature of the process being analyzed [50]. The exponential distribution is commonly employed for the arrival rate when customers arrive randomly and independently. On the other hand, the Poisson distribution is suitable for modeling customer arrivals in discrete intervals or integers [51]. The exponential distribution is also used for the service rate, especially when service times are independent and identically distributed [52]. However, in cases where service times follow a discrete distribution, geometric or Erlang distributions accurately represent the system [53].

2.2. Queuing theory advances

Queuing theory, a multidisciplinary field, has contributions from several researchers [54] concepts such as network analysis, little's rule [55], stochastic simulation [56], integrating queueing theory with Markov chain models in complex systems [57] and hybrid models with machine learning approaches [58]. In engineering and management [59] relationship between key variables, communication systems stability [60], stochastic simulation [61], queueing games, weak convergence, queueing redundancy scheduling, regenerative theory, and its applications, as well as stochastic coupling and ordering in a symmetric union, highlighting the shortest-tail model [62]. These researchers represent only a fraction of those who have influenced the continuing queueing theory evolution.

2.3. Simulation (or Discrete Simulation)

Use of simulation is heath systems has been invaluable in improving training and clinical decision making. It provides a safe environment for professionals to practice medical procedures, hone surgical skills and address complex situations without putting real patients at risk.

Discrete simulation to set up death scenarios during Covid-19 pandemics is emerging as an invaluable tool for analyzing various factors, such as the speed of virus spread, the effectiveness of containment measures and the capacity of the health system. In this way, it contributes to the design of more effective and adaptable strategies for dealing with health crises. A significant study supporting this assertion was conducted by the authors, who developed a discrete event simulation model as soon as the pandemic affected the city of São Paulo (Brazil). Using this model, several scenarios considering different levels of demand, from minimum to maximum (i.e., varying the number of expected deaths), were examined and evaluated [63].

Simulation methods:

These are techniques that make it possible to represent the behavior of real systems by means of mathematical and computational models. Different approaches have been developed over time to simulate complex phenomena, and several authors have been pioneers in this field. Some of the most relevant methods are mentioned below, together with the pioneering authors, key concepts, mathematical models and applications.

- **Monte Carlo Method:** Stanislaw Ulam and John von Neumann developed the Monte Carlo method in the 1940s while working on the Manhattan Project [64]. It is a probabilistic method that uses random samples to solve mathematical problems that can be deterministic or stochastic. It is based on the generation of random numbers to calculate integrals or solve statistical problems. For example, to calculate an integral, random points in the domain of the function are simulated and the proportions within the region of interest are estimated. Example:

$$\pi = \frac{4N_{inside}}{N_{total}} \tag{7}$$

  Where (N_{inside}) is the number of points inside the circle and (N_{total}) is the total number of points. Application: It is widely used in finance for option pricing, in physics to simulate particles, and in engineering to evaluate complex systems.

- **Finite Difference Method:** Lewis Fry Richardson is one of the pioneers in the application of this method in weather forecasting in the 1920s [65]. Finite difference methods are used to solve differential equations by approximating the derivatives by differences between discrete values. A typical example is the solution of the heat equation for the heat equation in one dimension:

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \tag{8}$$

  It can be discretion by finite differences:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \alpha \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} \tag{9}$$

  Where (u_i^n) represents the temperature at point (i) and in time (Δt). Application: This method is used to model phenomena such as heat diffusion, fluid dynamics and wave propagation.

- **Agent-based simulation (ABM):** John Holland is a reference in the development of the theory of complex adaptive systems and agent-based simulation [66]. In ABM, a system is modeled as a set of agents interacting with each other and with their environment, where agents are individual entities with specific behavioral rules. Agents are modeled by behavioral rules, which can be deterministic or stochastic. These rules can be influenced by the environment and interactions with other agents. Example: Agent Movement: If an agent moves on a grid, its position can be updated using:

$$X_{t+1} = X_t + V_x \Delta t \tag{10}$$

$$Y_{t+1} = Y_t + V_y \Delta t \tag{11}$$

Where (X_t) and (Y_t) are the current coordinates, $(V_x)$and (V_Y) are the velocities in the (x) and (y) directions, and $(\Delta t))$ is the time step.

Interaction between Agents: If the agents have an attraction/repulsion rule, a formula similar to the law of gravitation can be used:

$$F = G\frac{m_1 m_2}{r^2} \tag{12}$$

Where (F) is the interaction force, (G) is a constant, m_1 and m_2 are the masses of the agents, and (r) is the distance between them. Application: ABM is used in biology to model ecosystems, in economics to simulate markets and in sociology to study collective behaviors.

- **Finite Element Method (FEM):** Richard Courant and Alexander Hrennikoff pioneered the development of this method in the mid-20th century [67]. The finite element method divides a continuous domain (such as a physical structure) into small finite elements, over which partial differential equations are solved. The set of local solutions of these elements is combined to approximate the global solution. The FEM converts the partial differential equations (PDEs) into a system of algebraic equations that can be solved numerically. For example, in structural elasticity, it is used to solve the static equilibrium equations.

  $$K \cdot t\dot{u} = F \tag{13}$$

  Where (K) is the stiffness matrix, $(\dot{u})$ is the vector of displacements and (F) is the vector of applied forces.

- **Molecular dynamics simulation (MD):** B.J. Alder and T.E. Wainwright performed some of the first molecular dynamics studies in the 1950s [68]. Molecular dynamics simulation calculates the motion of particles (atoms and molecules) by solving Newton's laws for each particle. The motion of a particle in a force field can be described by Newton's second law:

  $$F = m \cdot a = m \cdot \frac{d^2 r}{dt^2} \tag{14}$$

  Where (F) is the force, (m) the mass, (a) the acceleration and (r) the position. Application: It is used in chemistry and biology to simulate molecular behaviors and study processes such as protein folding or material dynamics at the atomic scale.

- **Discrete Event Simulation (DES):** Geoffrey Gordon is one of the pioneers of discrete event simulation, known for developing the GPSS (General Purpose Simulation System) simulation language in the 1960s [69]. It simulates the behavior of a system at specific points in time, called discrete events. Unlike continuous methods, changes only occur when an event occurs that alters the state of the system. It uses a queuing system, in which events such as the arrival and departure of customers are modeled. A typical simulation scheme includes: Time between events: If events occur randomly, the time between events can be modeled using an exponential distribution:

$$T = -\frac{1}{\lambda}\ln(U) \tag{15}$$

Where (T) is the time between events, (λ) is the event occurrence rate, and (U) is a uniform random variable between 0 and 1. System state update: When an event occurs, the system state is updated. For example, in a queuing system, if a customer arrives, the number of customers in the queue (N) is increased:

$$N = N + 1 \tag{16}$$

Simulation clock advance: The simulation clock is advanced to the time of the next event in the list of future events:

$$t_{actual} - t_{proximoevento} \tag{17}$$

Applications: Computer network management, logistics systems, queuing in services (healthcare, airports). Each of these methods has applications in different fields of science and engineering, and have been fundamental to the development of simulation as a crucial tool in research and industry.

3. MATERIALS AND METHODS

This research has four phases. The methodological structure is presented in Figure 1.

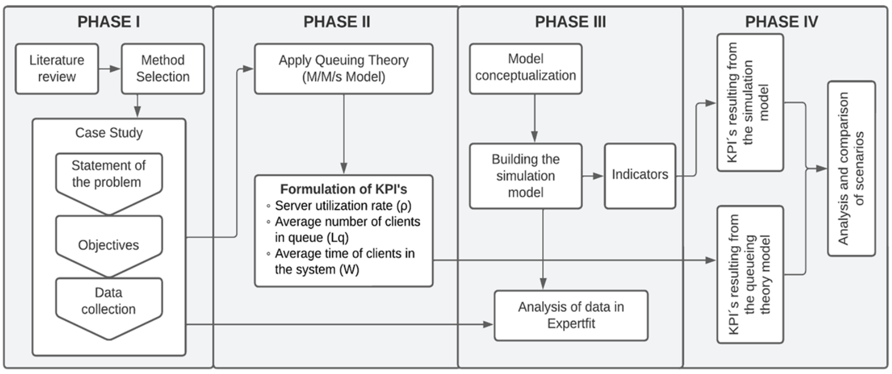

Figure 1: Methodological structure

In the first phase, the literature review, we identified methods to analyze waiting time problems. Subsequently, we took as a case study a drug dispensing center that served to validate the methodology proposed in this research. Sample data were taken for 29 days, recording the intervals between user arrivals and the attention times per user by means of stopwatches according to a previous calculation of the sample size. These data were analyzed using R, Python, and Expertfit [70]. The sample size ”n” was calculated from Eq. (18):

$$n = \frac{(N)(Z^2)(p)(q)}{e^2(N-1) + Z^2(p)(q)} \tag{18}$$

where ”n” represents the sample size in the month, ”N” corresponds to the sampling frame of the population, ”Z” refers to the confidence level, ”p” represents the probability that the event studied will occur, ”q” represents the probability that the event studied will not occur, and ”e” denotes the margin of estimation error. Replacing these values with those corresponding to the present case study, we obtain that n is approximately, n = 455.

In the second phase, applied queuing theory and key performance indicators (KPIs), such as queue user's number, use average time in the system, and server utilization. For this case study, selection criteria were defined based on several key factors, such as population served, technology implemented, operational capacity, location, and accessibility, selected to ensure the accuracy and reliability of the data collected in the research.

In the third phase, a simulation model was developed in FlexSim, adopting the methodological structure proposed in queuing theory [71]. In this phase, the simulated data was validated compared to the actual data, using a code in Python and Kolmogorov-Smirnov goodness-of-fit test validation. This statistical tool determines adequate simulation model accuracy and reliability.

Finally, in phase 4, a comparative results analysis was obtained between both models with FlexSim, facilitating bottlenecks identification and the improvement scenario proposal considering dispensing center efficiency and productivity.

4. RESULTS

4.1. Phase I. Data collection and analysis

In this first phase, we collected pharmacy patient flow data during the mornings and afternoons. These data included information on arrival-to-care times. The sample size was calculated using Equation 18. Expertfit statistical software examined multiple distributions and determined that the three presented in Table 1 and 2 have the highest relativity scores in the historical data set analyzed. This score evaluates the adequacy of the distribution models by comparing their effectiveness in representing historical data and brings transparency to the selection process. The results revealed that patient arrival and service times conformed to a Beta distribution. With a relativity score of 96.74, 94.57, 84.61 respectively, in this sense the Beta distribution is taken as a reference to represent the inter-arrival times of the clients. According to Cordeiro [72], the Beta distribution has three components: the lower endpoint, which in this case of customer inter-arrival times represents the minimum expected time between two consecutive arrivals; the upper endpoint, which represents the maximum expected time between two arrivals; alpha, which indicates the variability of the shorter times; and Beta, which indicates the variability of the longer times.

The resulting parameters for inter-arrival times are detailed in Table 1, while those related to service are presented in Table 2.

Table 1: Inter-arrival time distribution model's summary

Model	Relative Score(%)	Parameters	
		Lower endpoint(s)	9.22
Beta	96.74	Upper endpoint(s)	40.69
		Alpha	1.16
		Beta	1.06
		Lower endpoint(s)	8.53
Johnson SB	94.57	Upper endpoint(s)	41.52
		Alpha	-0.7
		Beta	0.69
		Location(s)	0
Weibull	82.61	Scale(s)	28.7
		Shape(s)	3.3

Table 2: Models for the distribution of service times summary

Model	Relative Score(%)	Parameters	
		Lower endpoint(s)	6.79
Beta	92	Upper endpoint(s)	56.26
		Alpha	5.35
		Beta	3.27
		Lower endpoint(s)	5.19
Johnson SB	92	Upper endpoint(s)	59.89
		Alpha	-0.63
		Beta	1.57
		Location(s)	0
Gamma	84	Scale(s)	1.84
		Shape(s)	20.38

The blue bars in the histograms in Figure 2 and Figure 3 represent the density of the collected sample data, and the red line represents the theoretical Beta distribution. Figure 2 shows a high frequency between time intervals of patient arrivals. Figure 3 shows the behavior of attention times in the module, in which high waiting peaks stand out. These figures show the relationship between the search for complex medications and high service times.

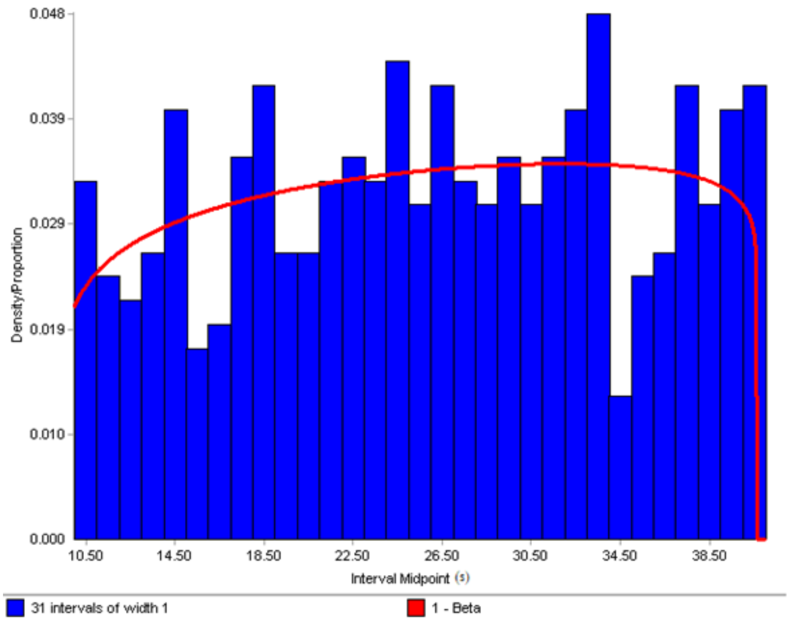

Figure 2: Input data histogram density (Source:The authors using Expertfit software)

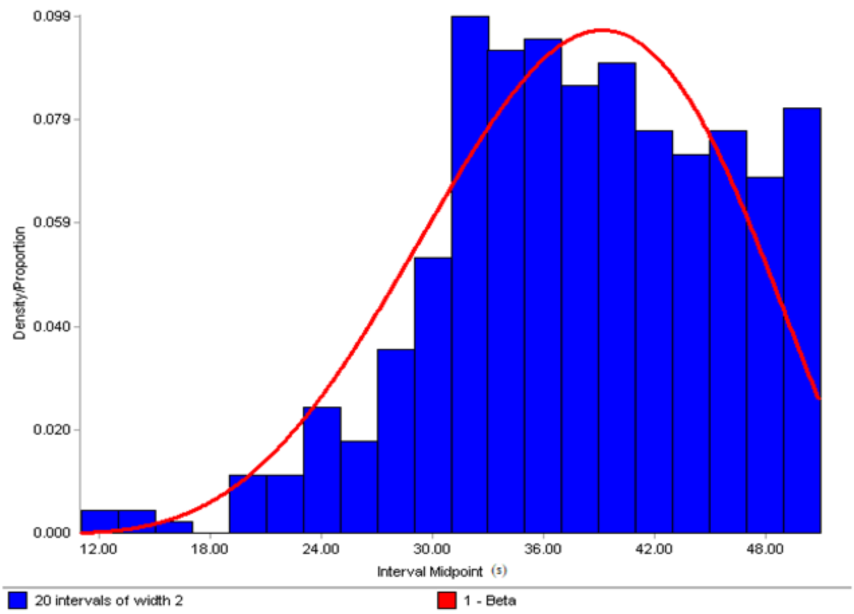

Figure 3: Input data histogram density (Source:The authors using Expertfit software)

4.2. Phase II. Queuing theory method application

The variables related to M/M/s queuing theory analysis from the collected data present two key parameters that stood out: the customer arrivals rate to the system (λ) and the service rate $(\mu$ [73]. This model applies in queuing systems involving multiple server(s), where customer arrivals follow a probabilistic Poisson distribution, and service times are exponential. There is an infinite population of customers and FIFO (first in, first out) for both morning and afternoon schedules. These approaches allow for a detailed queue behavior examination and its influence on the customer service process efficiency. Table 3 presents the results.

Table 3: Models for the distribution of service times summary

Indicators	Abbreviature	Data	Equation
Average number of customers in queuing	L_q	44	(Eq. 4)
The average time a customer spends in the system (s)	W_q	994	(Eq. 5)
Average service utilization (%)	ρ	0.8	(Eq. 6)

4.3. Phase III. Simulation model conceptualization

The simulation purpose in queuing theory is to provide a highly adaptable tool for modeling and analyzing queuing systems. This tool allows experimentation with configurations and strategies for improving efficiency and service quality in various applications and environments [74]. Thus, simulation becomes the option when real-world experiments are unfeasible or too costly due to lack of time, limited resources, or access to reliable historical data.

In this sense, we conduct simulation using FlexSim, which creates virtual models that represent processes or systems in the real world. In this specific case, we use that software to simulate and analyze the behavior of systems involving queue and customer service processes, allowing us to visualize and evaluate how different variables and scenarios affect the performance of these processes, thus providing valuable information for decision-making and optimization of customer service efficiency [75].

4.3.1. Model configuration

We initially structured the flowchart representing the simulation model in the Process Flow module of FlexSim. This tool establishes the logical sequence that governs the operation of a dispensing center. To begin with, we adjust the model to simulate the arrival of patients, ensuring that it places them in the queue according to their order of arrival. From this point, we implemented a process that allows the operator to assign shifts to patients based on their order of arrival. During the care process in the module, critical operations are simulated, such as the verification of medical orders and the collection of data necessary for the dispensing of medications. Finally, an exit procedure is established for users once they have obtained their medications, thus marking the closure of the process. This sequence of stages provides a complete and accurate representation of the operation of the dispensing center. Figure 4 presents the flowchart of the structured logic to simulate the case study.

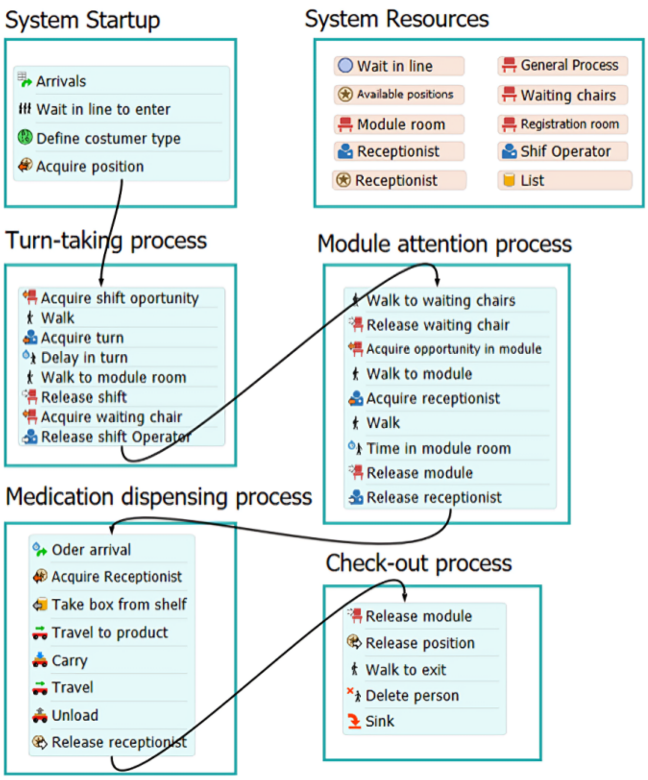

Figure 4: Dispensing Center Process Flow Configuration (Source:The authors using Flexsim)

4.3.2. Simulation Model Programming Algorithm

The diagram has allowed us to structure the logic of the simulation model, then the programming algorithm is built using the Python programming language. This code implements a simulation of a healthcare dispensing center using SimPy, a discrete event simulation library in Python. Here's an explanation of the structure and main functions used: Libraries used: The code uses the SimPy library for discrete event simulation. This library provides classes and methods to create and execute simulations. Functions and classes: DispensingCenter: This class represents the healthcare dispensing center. It has methods to simulate patient arrivals, module selection, and the check-out process. $patient_arrival$: This method simulates a patient's arrival at the dispensing center. The patient joins a queue, requests a healthcare provider and a care module if available. $select_module$: This method selects an available care module for the patient. If no modules are available, it returns a SimPy event indicating that the patient must wait. $check_out$: This method simulates the check-out process for a patient from the dispensing center. Patient generator: The $patient_generator$ function simulates continuous patient arrivals at the dispensing center. Each patient is processed using the $patient_arrival$ method. Simulation creation: A simulation environment (simpy.Environment) is created, and an instance of the DispensingCenter class is instantiated. Subsequently, the patient generator process is initiated in the simulation environment. Simulation execution: The simulation is executed using the env.run (until=455) method for a specific time duration (in this case, 455-time units). The following figure (show Figure 5) shows the programming code.

```
import simpy
import simpy
class DispensingCenter:
    def __init__(self, env, capacity, num_modules):
        self.env = env
        self.checkout_time = 3
        self.shift_duration = 4
        self.shift_provider = simpy.Resource(env, capacity=2)
        self.patient_queue = simpy.Store(env, capacity=80)
        self.modules = [simpy.Resource(env, capacity=1) for _ in range(num_modules)]
   def patient_arrival(self, patient):
        print(f"Patient {patient} arrives at the dispensing center at {self.env.now}")
        yield self.patient_queue.put(patient)
        with self.shift_provider.request() as request:
            yield request
            patient = yield self.patient_queue.get()
            print(f"Patient {patient} is assigned to a healthcare provider at {self.env.now}")
            module = yield self.select_module()
            if module is None:
                print(f"No module available for patient {patient} at {self.env.now}")
                return
            with module as module_request:
                yield module_request
                print(f"Patient {patient} is receiving care at module
                {self.modules.index(module)} at {self.env.now}")
                yield self.env.timeout(3)
    def select_module(self):
        for module in self.modules:
            if module.count == 0:
                return module.request()
        return None
    def check_out(self, patient):
        print(f"Patient {patient} checks out at {self.env.now}")
        yield self.env.timeout(self.checkout_time)
def patient_generator(env, dispensing_center):
    patient = 0
    while True:
        patient += 1
        env.process(dispensing_center.patient_arrival(patient))
        yield env.timeout(2)
env = simpy.Environment()
capacity = 28
num_modules = 5
dispensing_center = DispensingCenter(env, capacity, num_modules)
env.process(patient_generator(env, dispensing_center))
env.run(until=455)
```

Figure 5: System programming code (Source:The authors using Python)

4.3.3. Building the 3D model based on its infrastructure

The construction of the 3D model in FlexSim is based on the understanding of the challenges faced by the original drug distribution center. This center was characterized by having a waiting room, an internal storage warehouse, and five service modules, with manual assignment of shifts by an employee, resulting in queue delays. Since the waiting room capacity was limited to 48 people, the focus was improving operating efficiency. We implemented several key strategies to address these constraints. First, the number of service modules was increased to 6, allowing for a more even workload distribution and reduced waiting times. Second, an automated shift assignment system was incorporated into current technologies to eliminate delays associated with manual shift assignments. These proposals are based on optimizing care and adapting to the current high-demand environment, ensuring a more efficient and satisfactory patient experience. Figure 6 depicts the simulated 3D model of the medication dispensing center.

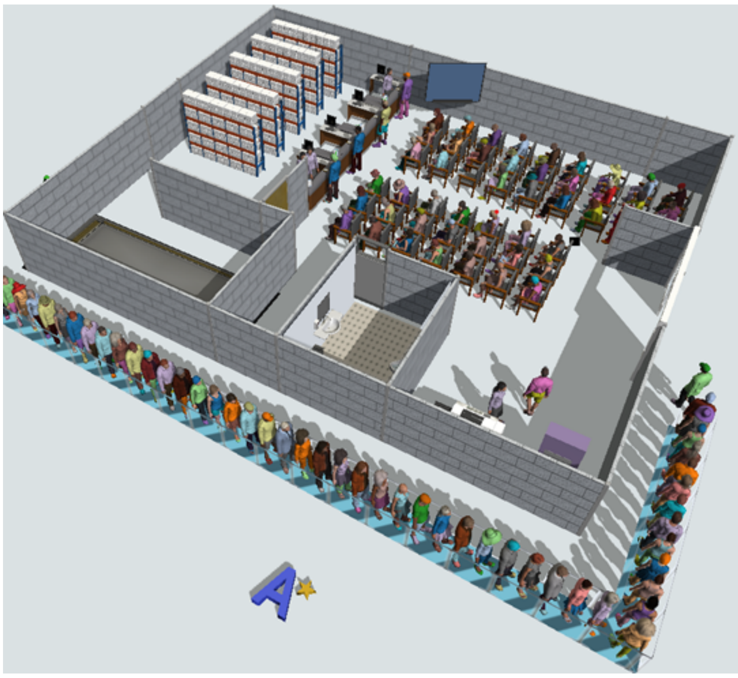

Figure 6: Medication dispensing center (Source:The authors using Flexsim)

4.3.4. Simulation results

The FlexSim provided the simulation model results, allowing the collection of relevant information. These results were generated by configuring and visualizing real-time data during the simulation run. In this section, we will present two simulation scenarios. In the first scenario, we will analyze the current queuing system to understand its structure and operation. The second scenario will explore a new module into the system and dispensing shifts automation. This exploration will allow us to evaluate how this addition might affect the overall operation and whether it has the potential to improve both efficiency and user satisfaction.

Current model

The current scenario has addressed multiple tasks simultaneously with five service modules, a single waiting queue for orderly customer management, a designated shift flow coordinator, and a 48-person maximum waiting room capacity, ensuring efficient flow control and customer comfort while waiting. However, due to the increased number of users entering the system, this dispensing center faces capacity issues to meet the demand, which may cause inconvenience to users. This analysis highlights the simulation's valuable utility in evaluating the moments of maximum flow fluctuation, previously considered constant time intervals according to queuing theory. Figure 7 presents a plot illustrating the variation in the number of people in the system during the peak congestion time of the day. An increase is observed around 9-10 hours when the number of people in the system exceeds 70, outstripping the facility's capacity under study and causing long queues. This finding highlights the need to adapt management and resources to these times of high demand, which can translate into a better user experience and greater operational efficiency of the facility.

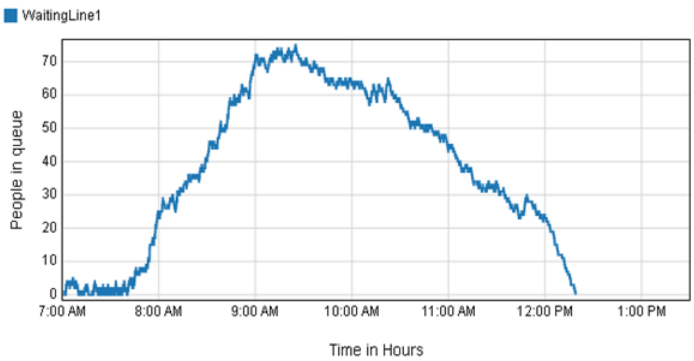

Figure 7: Number of people in queue vs time (hours) of the current model (Source:The authors using software Flexsim)

This service model has five logging modules in the room. Figure 8 reveals a high utilization rate of over 85% for each module, leading to problems such as congestion, long waiting times, operational inefficiency, and risk of system collapse. In this situation, it is essential to increase server capacity or redistribute the workload more efficiently. For optimal and satisfactory system performance, it is essential to keep utilization below 80%, following the Workload Profile method [37], which evaluates mental workload and notes that utilization above 80% implies a risk of occupational stress for the operator [76].

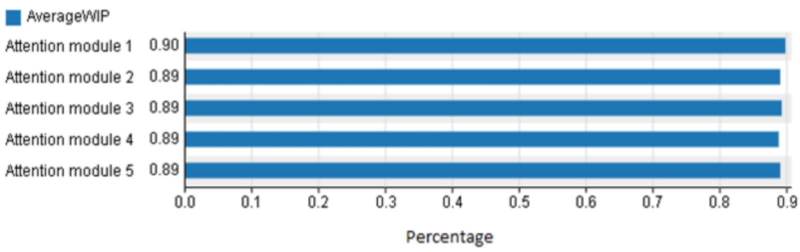

Figure 8: Current model use (%) (Source:The authors using software Flexsim)

Table 4 depicts the average waiting time, a critical indicator in queuing management systems. The average of 1639.59 seconds (approximately 27 minutes) observed in this figure suggests that customers waiting in queues experience significant waiting time before receiving attention. This extended waiting time can negatively impact customer satisfaction and the system's operational efficiency, which may require queuing management adjustments, such as allocating more resources, changing service policies, or reducing waiting time and improving customer experience strategies.

Table 4: Average queue time in the current scenario

Object	Average standby time	Minimum standby time	Maximum standby time
Waiting lines(s)	1639.59	19.17	3448.5

The new scenarios and two concrete alternatives suggest addressing these challenges. They are adding a new server during the hours when the simulation results show the highest fluctuation, i.e., between 9 am and 10 am. Figure 9 presents user waiting behavior at the drug distribution center during peak congestion hours, which provides crucial information on the efficiency and capacity of the facility. These improvement implementations reduce peak waiting times, especially between 9 am and 10 am. During this period, the people number in the system decreased considerably, reducing by 40 individuals. This decrease is a positive indication of a smoother flow of care and better management of resources by the drug distribution center. As a result, reducing long queues can translate into a more satisfactory user experience and improved center operational efficiency, which has a positive impact on customer satisfaction and resource management efficiency.

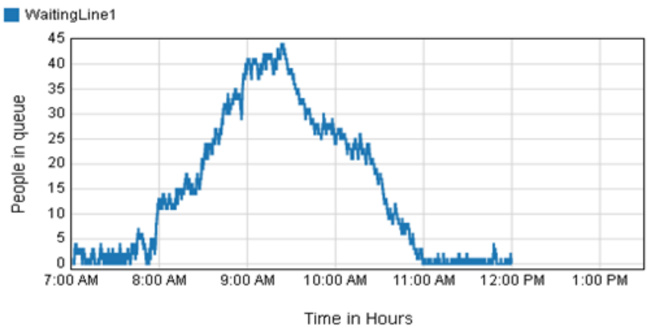

Figure 9: Number of people in queue vs time (hours). Proposed improved model (Source:The authors using software Flexsim)

As illustrated in Figure 10, the utilization rate decreased substantially in the improved scenario, ensuring that each module did not exceed the critical 80% utilization limit. The reduction was due to implementing measures, such as adding a new server, which allowed for a more efficient workload redistribution. This strategy manifested itself as improving system performance, reducing congestion risks, and long waiting times, and supporting this tool's importance in improving operability in healthcare.

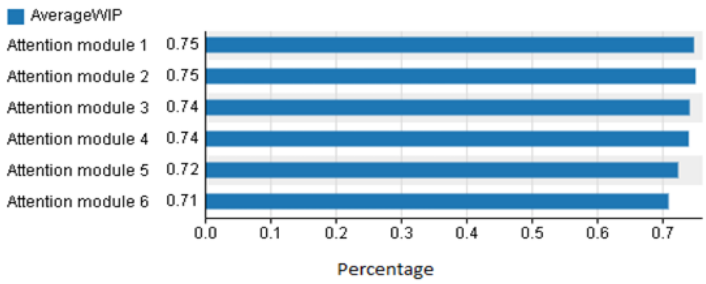

Figure 10: Percentage of use of the proposed model (Source:The authors using software Flexsim)

Table 5 shows a significant reduction in the average waiting time, which decreases from a considerably high initial value of 1,639.59 seconds to a significantly lower value of 591.41 seconds (equivalent to 9.857 minutes). Therefore, users or customers reduced their waiting time by approximately 96%, translating into a more satisfactory perception and a better experience.

Table 5: Average waiting time of the proposed model

Object	Average standby time	Minimum standby time	Maximum standby time
Waiting lines(s)	591.41	19.17	1824.9

We perform a goodness-of-fit test using the Kolmogorov-Smirnov test for the variable inter-arrival times and attention times in seconds to validate the simulation data with the actual data. In this test, we set two hypotheses:

- The null hypothesis (H0) would be that the data generated by the simulator follows the same distribution as the actual data.
- The alternative hypothesis (H1) would be that the data generated by the simulator does not follow the same distribution as the actual data.

The empirical cumulative distribution function (ECDF) is calculated from a code in the Python programming language to test these hypotheses. The test results yielded a test statistic (D) of 0.012 for inter-arrival times and 0.024 for service times, with p-values of 1 and 1.2, respectively. Since the p-value exceeds the commonly established significance level of 0.05 in both cases, the null hypothesis cannot be rejected, indicating that the simulated data adequately fit the distribution of the accurate data.

4.4. Phase IV. Comparison of results of both methods and improvement scenario

Throughout this phase, we compared and evaluated the data obtained during the simulation in Phase III and the actual data collected in Phase II. During this process, we carefully analyzed the discrepancies between the theoretical predictions and the practical results, which allowed us to identify possible deviations and areas for improvement within the system. This phase played an essential role in assessing the M/m/s model's effectiveness in accurately representing the natural system. It provided valuable information that guided decision-making and necessary improvements to the system.

As shown in Table 6, we have used metrics to calculate the performance parameters in different scenarios. There is a noticeable difference when comparing the queuing theory results with the actual simulation results. This simulation considers both infrastructure and queue speed, providing us with a scenario much closer to reality and, therefore, is the most reliable option to support decision-making.

In addition, when analyzing the key indicators (L_q, W_q, and ρ) of the simulation model for both scenarios, it is observed that the indicators of the proposed simulation model are lower since the system resources are better optimized.

Table 6: Comparison of the three scenarios

Indicators	Abbreviature	theorical scenario	Simulated scenario	Proposed scenario
Average number of customers in queuing	L_q	43.9	39.51	11.5
Average time in the system(s)	L_q	994.49	16.59	591.41
Average service utilization rate	ρ	0.8	0.9	0.74

These results brought essential advantages to the system, as they reduced the theoretical queue indicators and clarified the availability of space in the infrastructure for future adaptations in the case study scenario. It is relevant to highlight the notable decrease in waiting customers (L_q) in the proposed scenario, with a reduction of close to 50% compared to the current simulation scenario. Furthermore, with the improvements implemented, greater efficiency has been achieved by reducing the average time customers spend in the system (W_q). That is particularly significant because we added a server to the proposed system, from 5 to 6 service modules, and shift automation via a dispenser was introduced.

5. DISCUSSION

The present research adds to a growing body of studies on queuing systems and service management in healthcare settings. Reviewing the international literature, we found results that vary in congruence with this study. Some of these studies are consistent with our findings, while others differ significantly. For example, in a study conducted in a specialized hospital in 2012, a patient intake rate was observed that far exceeded the service delivery rate, which is consistent with our results. However, other studies have reported divergent results [77]. This variability highlights the importance of considering each healthcare facility's context and particularities when addressing queuing management issues.

This study has conclusively demonstrated that changes in the number of servers and reorganization of resources have a significant impact on waiting times and system performance. These results support the idea that strategic interventions can substantially improve healthcare efficiency. This finding aligns with previous research highlighting the importance of effective resource management [78]. Incorporating multitasking staff is an effective strategy to change the queuing model and improve patient experience.

A prominent feature is applying simulation modeling and queuing theory to address management problems in the healthcare environment. This methodological choice relies on the growing evidence of the effectiveness of these tools in optimizing drug supply systems. The results support this trend, showing that simulation could be a valuable tool for analyzing and improving processes in a drug dispensing center. Therefore, this research adds to the growing literature highlighting simulation tools' potential in strategic healthcare decision-making. In this sense, the most notable result is the significant reduction in patient waiting times. This improvement was achieved by reorganizing the mix of service providers and implementing a more efficient queuing system. These results align with previous research that shows that system changes can lead to tangible improvements in waiting times [79]. This study provides additional evidence in this area and highlights the importance of considering similar strategies to address congestion problems in healthcare settings.

6. CONCLUSION

This paper uses queuing theory and simulation to address variability in real systems, using a medicine dispensing center as an example. It contrasts the results of queuing theory with simulation. It finds that simulation is better suited to reality by using the beta distribution to model arrivals and attendances, considering the variability of time windows. These findings support the usefulness of this combination of approaches for forecasting scenarios to improve system performance.

The determination of these scenarios in the simulation method followed a systematic approach, including literature review, data collection, and scenario analysis. Consultation with experts ensured that the scenarios were relevant in real healthcare settings. This comprehensive approach made the scenarios not only theoretically sound, but also practically applicable, providing a solid basis for evaluating and improving healthcare systems.

Through this study, it was possible to simulate the impact of implementing an additional server and an automatic shift distribution system in the analyzed network. The results are highly encouraging, as they reveal a 50% decrease in waiting times in the system, which implies a substantial improvement in user service efficiency. Also, the adding of a server contributed to a better balance in the workload between the existing servers, which further optimized the system's overall performance.

This contribution to the knowledge framework is valuable, enabling industry practitioners and academics to develop evidence-based best practices to address queuing management issues in the pharmaceutical and other healthcare environments. The evidence from the research provides a sound basis for informing management decisions and strategies, which, in turn, can be reflected in significant improvements in the quality of care and patient satisfaction in these vital healthcare settings.

In this study, several limitations are recognized, such as the realization in a single drug dispensing center, which limits the generalizability of the results to other settings; the collection of data in specific periods, which does not capture all the variability; and the use of simulation models based on assumptions that do not fully reflect the real complexities.

In future research, incorporating virtual reality technology can be considered to understand the systems better. In addition, implementing more accurate forecasting systems to optimize server scheduling and management based on specific needs and schedules is suggested. Finally, the methodological framework used in this study can be applied in other service contexts, becoming a practical approach to developing health management policies and improving care for users with congenital and hereditary diseases that require constant medication.

Acknowledgements: The drug distribution center, the value chain research group, related subject matter experts, and the advisor thanked.

Funding: This research did not receive external funding.

REFERENCES

[1] S. Ghosh and R. Hassin, "Inefficiency in stochastic queueing systems with strategic customers," *European Journal of Operational Research*, vol. 295, no. 1, pp. 1–11, Nov. 2021. doi: 10.1016/j.ejor.2021.03.065.

[2] N. G. Micheal Mathavavisakan and K. Indhira, "A literature review on retrial queueing system with Bernoulli vacation," *Yugoslav Journal of Operations Research*, vol. 34, no. 1, pp. 109–134, 2024. doi: 10.2298/YJOR230415020M.

[3] M. A. Cohen et al., "Bespoke Supply Chain Resilience: The Gap between Theory and Practice," *Journal of Operations Management*, vol. 68, no. 5, pp. 515-531, 2022. doi: 10.1002/joom.1184.

[4] N. T. Mai and D. M. Cuong, "The Application of Queueing Theory in the Parking Lot: a Literature Review," in *Proceedings of the International Conference on Emerging Challenges: Business Transformation and Circular Economy (ICECH 2021)*, Hanoi, Vietnam, 2021, pp. 190–203. doi: 10.2991/aebmr.k.211119.020.

[5] Dr. A. Adeniran, M. Sani Burodo, and Dr. S. Suleiman, "Application of Queuing Theory and Management of Waiting Time Using Multiple Server Model: Empirical Evidence From Ahmadu Bello University Teaching Hospital, Zaria, Kaduna State, Nigeria," *International Journal of Scientific and Management Research*, vol. 5, no. 4, pp. 159–174, 2022, doi: 10.37502/ijsmr.2022.5412.

[6] G. A. Girón-Restrepo, S. Tena-Monferrer, and J. C. Fandos-Roig, "A review of the Colombian healthcare system: challenges and opportunities," in *Modern Healthcare Marketing in the Digital Era*, K. Djakeli, Ed. Hershey, PA: IGI Global, 2024, pp. 150–163. doi: 10.4018/979-8-3693-0679-6.ch009

[7] Equipo Directivo de Superintendencia de notariado y registro, "Informe general rendición de cuentas Superintendencia Notariado y Registro," 2019, [Online]. Available: https://servicios.supernotariado.gov.co/files/portal/portal-informe_rdc20192020.pdf.

[8] M. C. Ramírez, "El tiempo de espera es el talón de Aquiles de las entidades bancarias," *Revista el Tiempo*, 2016, [Online]. Available: https://www.larepublica.co/consumo/el-tiempo-de-espera-es-el-talon-de-aquiles-de-las-entidades-bancarias-2354506.

[9] D. Pabón Guerrero, Y. I. Muñoz Silva, J. P. Muñoz Valencia, and S. A. Noguera Muñoz, "Barreras de acceso en la prestación de los servicios de salud para los usuarios de la IPS Totoguampa del Municipio de Silvia en el Departamento del Cauca, 2022," Universidad Católica de Manizales, 2022. [Online]. Available: https://repositorio.ucm.edu.co/handle/10839/4072.

[10] D. A. Collo Burbano et al., "Retos de la farmacovigilancia con la venta, comercialización y dispensación adecuada de medicamentos homeopáticos y medicamentos tradicionales a base de plantas en un establecimiento farmacéutico," Universidad Nacional abierta y a distancia, May 2023. [Online]. Available: http://repository.unad.edu.co/handle/10596/57613

[11] Ministerio de salud y Protección social, "Informe global del proceso de consulta pública del documento completo de Análisis de Impacto Normativo (AIN) en la temática de alimentos para propósitos médicos especiales (APME) en Colombia.," 2023. [Online]. Available: https://www.minsalud.gov.co/Normativa/Paginas/analisis-de-impacto-normativo.aspx.

[12] Ministerio de salud y Protección social, "Perfil farmacéutico de la República de Colombia," 2012.[Online]. Available: file:///C:/Users/Unicomfacauca/Downloads/pscp_colombia_1.pdf.

[13] X. Wang and S. M. Disney, "The bullwhip effect: Progress, trends and directions," *European Journal of Operational Research*, vol. 250, no. 3, pp. 691–701, May 2016, doi: 10.1016/J.EJOR.2015.07.022.

[14] D. A. Figueredo De Pérez and I. Vargas-Chaves, "El acceso a medicamentos en Colombia y los contornos de un derecho y una política farmacéutica a medio camino," *Justicia*, vol. 25, no. 37, pp. 125–150, Jun. 2020, doi: 10.17081/just.25.37.3528.

[15] D. Worthington, M. Utley, and D. Suen, "Infinite-server queueing models of demand in healthcare: A review of applications and ideas for further work," *Journal of the Operational Research Society*, vol. 71, no. 8, pp. 1145–1160, Aug. 2020, doi: 10.1080/01605682.2019.1609878.

[16] W. F. Angeles Bachet, "Análisis de la demora y las salidas ineficaces del proceso desde un enfoque de sistemas y de Teoría de colas. Una revisión sistemática,"*Revista de Climatología*, vol. 23, pp. 1104–1113, Jul. 2023, doi: 10.59427/RCLI/2023/V23CS.1104-1113.

[17] M. M. Hamdi, H. F. Mahdi, M. S. Abood, R. Q. Mohammed, A. D. Abbas, and A. H. Mohammed, "A review on Queue Management Algorithms in Large Networks," *IOP Conference Series: Materials Science and Engineering*, vol. 1076, no. 1, p. 012034, Feb. 2021, doi: 10.1088/1757-899X/1076/1/012034.

[18] H. Attar, M. R. Khosravi, S. S. Igorovich, K. N. Georgievan, and M. Alhihi, "Review and performance evaluation of FIFO, PQ, CQ, FQ, and WFQ algorithms in multimedia wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 6, p. 155014772091323, Jun. 2020, doi: 10.1177/1550147720913233.

[19] F. F. Galimulina and N. V. Barsegyan, "Application of Mass Service Theory to Economic Systems Optimization Problems—A Review," *Mathematics*, vol. 12, no. 3, p. 403, Jan. 2024, doi: 10.3390/MATH12030403.

[20] R. Renteria, M. Chong, I. de Brito Junior, A. Luna, and R. Quiliche, "An entropy-based approach for disaster risk assessment and humanitarian logistics operations planning in Colombia," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 11, no. 3, pp. 428–456, Jul. 2021, doi: 10.1108/JHLSCM-03-2020-0018.

[21] R. Delgado de la Torre, "Recordando a Erlang: Un breve paseo (sin esperas) por la Teoría de Colas," *Materials matemàtics*, pp. 0001–0033, 2009. https://ddd.uab.cat/record/97914

[22] J. Linares-Cos, J. A. Vilalta Alonso, and R. Garza Ríos, "Concepciones teóricas sobre la calidad en la gestión de los servicios y su efectividad desde la teoría de colas," *Revista Cubana de Administración Pública y Empresarial*, vol. 4, no. 1, pp. 128–143, 2020.

[23] E. López and L. G. Joa, "Queueing theory applied to the study of the service system of a pharmacy," *Revista Cubana de Informática Médica*, vol. 10, no. 1, pp. 3–15, 2018.

[24] J. M. Izar Landeta, C. B. Ynzunza Cortés, and J. Garnica González, "Análisis y optimización de dos sistemas de líneas de espera de empresas de logística y transporte de los Estados de Querétaro y Colima," *Investigación administrativa*, vol. 47, no. 121, pp. 1–20, 2018,

[25] S. Adhikari et al., "Analysis of average waiting time and server utilization factor using queue-

ing theory in cloud computing environment," International *Journal of Nonlinear Analysis and Applications*, vol. 12, no. Special Issue, pp. 1259–1267, Dec. 2021, doi: 10.22075/ijnaa.2021.5636.

[26] N. S. Alejandro Pérez, S. Hernández González, J. A. Jiménez García, and V. Figueroa Fernández, "Análisis de un sistema de servicio con Teoría de colas y simulación," *Pistas Educativas*, vol. 41, no. 135, 2020.

[27] G. P. Loor Alcívar, S. M. Rodríguez Merchán, O. B. Santos Vásquez, and B. J. Loor Alcívar, "Teoría de colas y optimización de proceso de atención al usuario," *AlfaPublicaciones*, vol. 4, no. 3, pp. 22–38, Jul. 2022, doi: 10.33262/ap.v4i3.221.

[28] M. A. Kalwar, H. Bux Marri, M. A. Khan, and S. A. Khaskheli, "Applications of Queuing Theory and Discrete Event Simulation in Health Care Units of Pakistan," *International Journal of Science and Engineering Investigations*, vol. 10, no 109, p. 6-18, 2021

[29] R. A. Memon, J. P. Li, and J. Ahmed, "Simulation Model for Blockchain Systems Using Queuing Theory," *Electronics* , vol. 8, no. 2, p. 234, Feb. 2019, doi: 10.3390/electronics8020234.

[30] P. Amalia and N. Cahyati, "Queue analysis of public healthcare system to reduce waiting time using flexsim 6.0 software," *International Journal of Industrial Optimization*,vol. 1, no. 2, pp. 101–110, 2020.

[31] W. M. Salazar Gallardo, "Modelo de teoría de colas y simulación para la asignación del recurso humano en un sistema operativo de servicios.," Escuela Politécnico Nacional, Quito, 2022. [Online]. Available: http://bibdigital.epn.edu.ec/handle/15000/22795

[32] A. Elalouf and G. Wachtel, "Queueing Problems in Emergency Departments: A Review of Practical Approaches and Research Methodologies," *Operations Research Forum*, vol. 3, no. 1, pp. 1–46, Mar. 2022, doi: 10.1007/S43069-021-00114-8.

[33] J. E. Távara Pacheco, "Aplicación de la teoría de colas para proponer mejoras en la atención del paciente en el servicio de farmacia del Hospital III José Cayetano Heredia-Piura," Universidad Nacional de Piura, 2019. [Online]. Available: https://repositorio.unp.edu.pe/handle/UNP/1873

[34] C. Chen and L. K. Tiong, "Using queuing theory and simulated annealing to design the facility layout in an AGV-based modular manufacturing system," *International Journal of Production Research*, vol. 57, no. 17, pp. 5538–5555, 2019, doi: 10.1080/00207543.2018.1533654.

[35] P. O. Peter and R. Sivasamy, "Queueing theory techniques and its real applications to health care systems – Outpatient visits," *International Journal of Healthcare Management*, vol. 14, no. 1, pp. 114–122, 2021, doi: 10.1080/20479700.2019.1616890.

[36] A. Zhang, X. Zhu, Q. Lu, and R. Zhang, "Impact of Prioritization on the Outpatient Queuing System in the Emergency Department with Limited Medical Resources," *Symmetry*, vol. 11, no. 6, p. 796, Jun. 2019, doi: 10.3390/SYM11060796.

[37] L. A. Velasquez Silva, "Análisis del impacto de la aplicación de la teoría de colas en la reducción de tiempos y la satisfacción de clientes en el rubro de alimentos, en los últimos 5 años: una revisión de la literatura científica," Universidad Privada del Norte, Lima, 2020. [Online]. Available: https://hdl.handle.net/11537/24858

[38] J. Deng, "Resource Management in FlexSim Modelling: Addressing Drawbacks and Improving Accuracy," *Applied Sciences*, vol. 13, no. 9, p. 5760, May 2023, doi: 10.3390/app13095760.

[39] K. Akin and H. & Ordu, "A novel Simulation-Based Two-Stage optimization approach for nurse planning," *International Journal of Simulation Modelling*, vol. 20, no. 4, pp. 591–602, Dec. 2022, doi: 10.2507/ijsimm21-4-618.

[40] M. Ordu, E. Demir, C. Tofallis, and M. M. Gunal, "A comprehensive and integrated hospital decision support system for efficient and effective healthcare services delivery using discrete event simulation," *Healthcare Analytics*, vol. 4, p. 100248, Dec. 2023, doi:

10.1016/j.health.2023.100248.

[41] S. W. Akuamoah, D. Lu, and D. Yaro, "Application of Queueing Theory in Dispensation of Resources and Optimization in Teleconsultation," *International Journal of Applied and Computational Mathematics*, vol. 6, no. 4, pp. 1–17, Aug. 2020, doi: 10.1007/S40819-020-00851-0.

[42] Y. A.-W. Nawusu, A. A. A. Danaa, and S. Ismail, "Queue Management during Health Pandemics: A Queuing Theory Perspective," *Asian Journal of Research in Computer Science*, vol. 6, n. 4, pp. 18–26, Dec. 2020, doi: 10.9734/ajrcos/2020/v6I430164.

[43] F. Lenin, V. Satama, M. L. Bernal, and D. I. Montenegro Gálvez, "Teoría de colas y líneas de espera, un reto empresarial en el mejoramiento continuo de los servicios," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 5, no. 5, pp. 8418–8440, Oct. 2021, doi: 10.37811/cl_rcm.v5i5.933.

[44] W. A. Massey and J. Pender, "Dynamic rate Erlang-A queues," *Queueing System*, vol. 89, no. 1–2, pp. 127–164, Jun. 2018, doi: 10.1007/s11134-018-9581-2.

[45] M. Demircioglu, H. Bruneel, and S. Wittevrongel, "Analysis of a Discrete-Time Queueing Model with Disasters," *Mathematics*, vol. 9, no. 24, p. 3283, Dec. 2021, doi: 10.3390/math9243283.

[46] R. Rathore, "A Review on Study of application of queueing models in Hospital sector," *International Journal for Global Academic & Scientific Research*, vol. 1, no. 2, pp. 01–05, Jun. 2022, doi: 10.55938/ijgasr.v1i2.11.

[47] W. Qu, T. Tao, B. Xie, and Y. Qi, "A State-Dependent Approximation Method for Estimating Truck Queue Length at Marine Terminals," *Sustainability*, vol. 13, no. 5, p. 2917, Mar. 2021, doi: 10.3390/su13052917.

[48] G. Ascione, N. Leonenko, and E. Pirozzi, "Fractional Erlang queues," *Stochastic Processes and their Applications*, vol. 130, no. 6, pp. 3249–3276, Jun. 2020, doi: 10.1016/j.spa.2019.09.012.

[49] J. W. S. de Lima, G. Callou, and E. Andrade, "Queueing Theory and Stochastic Petri Net: A tutorial," *Research, Society and Development*, vol. 10, no. 3, pp. e2810312826, Mar. 2021, doi: 10.33448/rsd-v10i3.12826.

[50] C. Blondia, "Evaluation of the Waiting Time in a Finite Capacity Queue with Bursty Input and a Generalized Push-Out Strategy," *Mathematics*, vol. 10, no. 24, p. 4771, Dec. 2022, doi: 10.3390/math10244771.

[51] I. Cruz Domínguez and E. García Jiménez, "Recolección y Análisis de Datos del Sistema de línea de espera del Instituto Soledad Acevedo De Los Reyes," *Anuario de Investigación UM*, vol. 1, no. 1, pp. 35–48, Aug. 2020.

[52] W. Q. Duan, Z. Khan, M. Gulistan, and A. Khurshid, "Neutrosophic Exponential Distribution: Modeling and Applications for Complex Data Analysis," *Complexity*, vol. 2021, no. 1, p. 5970613, Jan. 2021, doi: 10.1155/2021/5970613.

[53] D. L. Rizzini, F. Galasso, and S. Caselli, "Geometric Relation Distribution for Place Recognition," *IEEE Robot Autom Lett*, vol. 4, no. 2, pp. 523–529, Apr. 2019, doi: 10.1109/lra.2019.2891432.

[54] S. A. Afolalu, O. M. Ikumapayi, A. Abdulkareem, M. E. Emetere, and O. Adejumo, "A short review on queuing theory as a deterministic tool in sustainable telecommunication system," *Materials Today: Proceedings*, vol. 44, pp. 2884–2888, Jan. 2021, doi: 10.1016/j.matpr.2021.01.092.

[55] Y. Perwej, S. Qamar Abbas, J. P. Dixit, D. N. Akhtar, and A. K. Jaiswal, "A Systematic Literature Review on the Cyber Security," *International Journal of Scientific Research and Management*, vol. 9, no. 12, pp. 669–710, Dec. 2021, doi: 10.18535/ijsrm/v9i12.ec04.

[56] E. Jafarnejad Ghomi, A. M. Rahmani, and N. N. Qader, "Applying queue theory for modeling

of cloud computing: A systematic review," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 17, p. e5186, Sep. 2019, doi: 10.1002/cpe.5186.

[57] A. Dudin, O. Dudina, S. Dudin, and K. Samouylov, "Analysis of Multi-Server Queue with Self-Sustained Servers," *Mathematics*, vol. 9, no. 17, p. 2134, Sep. 2021, doi: 10.3390/math9172134.

[58] T. M. V. Anuruddhika, S. Prasanth, and R. M. K. T. Rathnayaka, "The Approaches Utilized in Queuing Modeling: A Systematic Literature Review," *Asian Journal For Convergence In Technology*, vol. 8, no. 2, pp. 24–30, Aug. 2022, doi: 10.33130/ajct.2022v08i02.006.

[59] L. Fratta, M. Gerla, and L. Kleinrock, "Flow Deviation: 40 years of incremental flows for packets, waves, cars and tunnels," *Computer Networks*, vol. 66, pp. 18–31, Jun. 2014, doi: 10.1016/j.comnet.2014.04.001.

[60] M. J. Blesa and A. Fernández Anta, "Maria Serna's contributions to adversarial queuing theory," *Computer Science Review*, vol. 39, p. 100348, Feb. 2021, doi: 10.1016/j.cosrev.2020.100348.

[61] P. W. Glynn and H. Lam, "Constructing simulation output intervals under input uncertainty via data sectioning," *Simulation Winter Conference*, vol. 2018-December, pp. 1551–1562, Jul. 2018, doi: 10.1109/wsc.2018.8632450.

[62] I. J. B. F. Adan, B. D'Auria, and O. Kella, "Special volume on 'Recent Developments in Queueing Theory' of the third ECQT conference," *Queueing Systems*, vol. 93, no. 1, pp. 1–2, Sep. 2019, doi: 10.1007/s11134-019-09630-1.

[63] I. de Brito, M. H. C. Cunha, L. A. Tozi, L. A. Franzese, L. da S. Frazão Márcia, and A. Bressane, "Managing funerary systems in the pandemic: lessons learned and an application of a scenario simulation in São Paulo City, Brazil," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 11, no. 3, pp. 481–492, 2021, doi: 10.1108/jhlscm-09-2020-0078.

[64] S. M. Ulam and S. Higuera De Frutos, "Aventuras de un matemático. Memorias de Stanislaw M. Ulam," *Pensamiento Matemático*, vol. 7, no. 1, pp. 199–202, 2017.

[65] P. A. C. Raats and J. H. Knight, "The contributions of Lewis Fry Richardson to drainage theory, soil physics, and the soil-plant-atmosphere continuum," *Frontiers in Environmental Science*, vol. 6, no. APR, p. 346275, Apr. 2018, doi: 10.3389/fenvs.2018.00013.

[66] H. John, "Holland, Adaptation in natural and artificial systems," Ann Arbor MI University of Michigan Press, vol. Ann Arbor, pp. 1-200?, 1992.

[67] U. Farooq and M A. Ahmad, "Finite Element Approach To Reinforced Concrete Using Ansys." *National conference on sustainable infrastructure development*, At Chandigarh, India. 2014.

[68] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, Aug. 1959, doi: 10.1063/1.1730376.

[69] System simulation, 2nd. ed. New Jersey, 1987. https://bib.ubp.edu.ar/cgi-bin/koha/opac-detail.pl?biblionumber=16647.

[70] J. A. Barra Montevechi, A. T. Campos, G. T. Gabriel, and C. H. dos Santos, "Input Data Modeling: An Approach Using Generative Adversarial Networks," *Winter Simulation Conference*, IEEE, pp. 1–12, Dec. 2021. doi: 10.1109/WSC52266.2021.9715407.

[71] Y. Anacona-Mopan, J. Segura-Dorado, and H. Paz-Orozco, "Optimización de la distribución en planta con formulación QAP y simulación de eventos discretos," *Informador Técnico*, vol. 87, no. 1, pp. 13–28, Sep. 2023, doi: 10.23850/22565035.4814.

[72] J. González C, D. Galvis S, and L. Hurtado T, "La distribución Beta Generalizada como un modelo de sobrevivencia para analizar la evasión universitaria," *Estudios pedagógicos (Valdivia)*, vol. 40, no. 1, pp. 133–144, 2014, doi: 10.4067/s0718-07052014000100008.

[73] S. V. Prasad, P. S. K. Patra, R. Donthi, and M. K. Challa, "The optimization of the health

care system by M/M/S queueing model," *AIP Conference Proceedings*, vol. 2707, no. 1, May 2023, doi: 10.1063/5.0148445/2889666.

[74] Y. D. V. Torres, Y. D. V. Torres, P. L. Alfonzo, and S. I. Mariño, "Teorí¬a de colas. Propuesta de un simulador didáctico," *Revista Publicando*, vol. 4, no. 13 (2), pp. 1–20, Dec. 2017.

[75] R. Poloczek and B. Oleksiak, "Modeling and simulating production processes with the use of the FLEXSIM method," *Metalurgija*, vol. 62, no. 3–4, pp. 484–487, Jul. 2023.

[76] C. E. Mendoza-Macías and Y. I. Ramos-Alfonso, "Evaluación de la calidad del servicio con la utilización de indicadores cualitativos y cuantitativos en el sector bancario," *Dominio de las Ciencias*, vol. 6, no. 3, pp. 950–966, Jul. 2020, doi: 10.23857/dc.v6i3.1327.

[77] M. M. Kembe, E. S. Onah, and Iorkegh, "A study of waiting and service costs of a Multi-Server Queuing model in a specialist Hospital," *International Journal of Scientific & Technology Research*, vol. 1, n. 8, pp. 19–23, 2012.

[78] M. Mohammadi and M. Shamohammadi, "Queuing Analytic Theory Using WITNESS Simulation in Hospital Pharmacy," *International Journal of Engineering & Technology*, vol. 12, no. 6, pp. 20–27, 2012.

[79] M. Bahadori, S. M. Mohammadnejhad, R. Ravangard, and E. Teymourzadeh, "Using queuing theory and simulation model to optimize hospital pharmacy performance," *Iranian Red Crescent medical journal*, vol. 16, no. 3, 2014, doi: 10.5812/ircmj.16807.