CHAPTER 2
LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1. Literature Review

In health care planning, the application of operations research (OR) is a well-known tool of maximizing the utilization of resources and reducing waiting times (Brandeau *et al.*, 2004). These researches offer valuable methods for decision-making by the use of mathematical models and the simulation of current capability processes, such as in the outpatient installation section. Such analysis and simulation techniques are based on historical occurrences, statistical simulation, and projection of potential future events with an appropriate proportion of inaccuracy. These sorts of experiments are utilized to develop treatments based on such setups since the processes are so complicated that their actions cannot be predicted easily. The more intricate the scheme, the higher the number of potential elements of input and interactions are generated.

One kind of study is called queuing theory analysis (Takakuwa *et al.*, 2007). It is used in areas where it is very important to sequence occurrences and waiting times. In reality, queuing theory is a statistical approach for assessing planned timeframes and preparing to offer capacity and service depending on the results. From the spontaneous admission of patients to the outpatient installation section, waiting queues typically emerge. The cost of waiting time lost by patients and the cost of the facilities rendered by the implementation of this approach should be balanced. With the use of quantitative models and statistical distributions, this principle can support administrators by evaluating postponements and waiting times throughout the procedures to enhance their system functions.

2.2. Queuing Theory

In 1904, the queueing principle was established by A.K. Erlang to help assess the Danish telecommunications system's power requirements (Brockmeyer *et al.*, 1948). It has since been extended both to a wide variety of service sectors and emergency systems. In assessing appropriate levels of workers, services, and housing, as well as in deciding on the allocation of resources and the introduction of new initiatives, queueing models can be very beneficial.

Queueing models need fewer data, unlike simulation methodologies, and are easy and quick to use, making it a very strong and practical method. Because of this simplicity and speed, different options for service delivery can be easily assessed and compared. In addition to the most fundamental problem of deciding how much capacity is required to reach a defined level of service, queueing models may also be helpful to gain insights into the necessary level of specialization or versatility to be used in the arrangement of resources or the effect of different priority schemes on the determination of patient service sequence.

On the contrary, the form of operational data required in healthcare environments as input for the queueing model is often inaccessible while queueing models only need very little data. While the data of demand or arrival are sometimes registered, there is typically no documentation of service times. However, this type of data might become readily accessible as in healthcare, the information management systems are becoming more widespread.

2.3. Characteristic of Queue

According to Heizer and Render (2006), there are three components in a queuing system, namely:

2.3.1. Characteristics of Arrival or System Input

Input sources that bring in customers for a service system have the following main characteristics:

a. Population size

This is a source of consumers who are unlimited or limited population. An infinite population is where only a percentage of all possible arrivals are the number of arrivals at any given time. Whereas when there are only a limited number of potential service users in a queue is a limited population.

b. Arrival behavior

The behavior of each consumer is different in obtaining service. There are three characteristics of arrival behavior, namely, patient customers, customers who refuse to join the queue, and customers who defect.

c. Arrival pattern

It describes how customer distribution enters the system. The arrival distribution consists of constant arrival distribution and random arrival pattern. The customer who arrives every certain period is the constant arrival distribution, while the customer who arrives unexpectedly is the arrival pattern random.

2.3.2. Queuing Discipline

Queuing discipline is a queuing rule that refers to the rules of customers who are in line to receive services consisting of:

a. First In First Out (FIFO) is the queuing system for customers who arrive first will be served first. For example, a queuing system at a bank, gas station, etc;

b. Last In First Out (LIFO) is the queuing system for customers who come last will be served first. For example, the queue system in an elevator lifts for the same floor;

c. Service in Random Order (SIRO) is the queuing system in which calls are dependent on random odds regardless of who comes first to be served;

d. Shortest Operation Times (SOT) is a queuing system where the one who gets the first service is the one who has the shortest service time.

2.3.3. Service Facilities

Two important things in the characteristics of service facilities are:

a. Service system design

Services are generally categorized based on the number of channels available and the number of stages. According to the number of channels, there are single-line queuing systems and multiple line queuing systems. Meanwhile, according to the number of stages is a single-stage system and a multiple stage system.

b. Distribution of service time

The service pattern is like the arrival pattern where this pattern can be constant or random. The time required to serve each patient is the same if the service time is constant. Meanwhile, the random service time is when the time is random or not the same for serving each customer.

2.4. Model of Patient Flow

When assessing the current or planned service systems, five standard metrics must be considered: the total number of waiting patients; the average wait period for patients; capacity utilization; the cost of a given capacity level; and the possibility that a patient entering the system might have to wait for service.

The measure of usage of the device represents the degree to which the servers are occupied instead of idle. Using 100 percent utilization may not be feasible in normal circumstances; the hospital should strive for a strategy that minimizes the amount of time spent waiting and the expense of the ability. During queue modeling, the healthcare must also guarantee that total arrival and service rates remain constant, implying that the system is in a stable condition, which is a core assumption.

The key features of the queuing model are (Yasara, 2009): the number of servers; the source of the population; the arrival and service patterns; and the discipline of queues.

2.5. Queue Formula for Single Path Model

There are several queue formulas with several symbols used for the single-path model, which are:

x = average number of arrivals per unit time

y = the number of individuals served per unit time

a. The system's average number of patients (who waits to be served)

$$Ls = \frac{x}{y - x} \quad (2.1)$$

b. The length of time spent in the system on average (time to wait plus the duration of the service)

$$Ws = \frac{1}{\mathrm{y} - \mathrm{x}} \quad (2.2)$$

c. The average number of waiting patients in the line

$$Lq = \frac{\mathrm{x}^2}{\mathrm{y}\,(\mathrm{y} - \mathrm{x})} \quad (2.3)$$

d. Average time spent in the line for waiting before being served

$$Wq = \frac{\mathrm{x}}{\mathrm{y}\,(\mathrm{y} - \mathrm{x})} \quad (2.4)$$

e. System utilization

$$\rho = \frac{\mathrm{x}}{\mathrm{y}} \quad (2.5)$$

f. The probability that there are 0 units in the system

$$P0 = 1 - \frac{\mathrm{x}}{\mathrm{y}} \quad (2.6)$$

g. The likelihood that there are more than *k* patients being in the system, where n is the total number of patients

$$Pn > k = (\frac{x}{y})^{k-1} \qquad (2.7)$$

2.6. Scheduling

Scheduling is the process of allocating work to resources for a specific period of time. Schedules are present in everything from operating systems and networks to industrial plants and medical visits. Many of these issues are solved by manual action or heuristics specifically suited for the situation (Gomes, 2017).

Patient scheduling is a dynamic activity which is critical to the continuity of treatment. There are many ways of patient scheduling, ranging from distributing resources to patients in need of examinations and allocating surgical rooms to scheduling on-demand appointment for Family Doctors operating at Primary Care Clinics. An effective scheduling of appointments would prevent the disappointment of patients and doctors, and as such, is a significant determinant of treatment.

The appointment scheme is planned to connect real patient appointment requests to a predetermined arrival phase. The interarrival hours, the number of arrivals at each arrival period, and the number of arrival periods over the booking horizon may all be used to identify the mapping arrival process. The reservation horizon is the time between when a provider opens reservations for a certain session (block of time) and when that session begins. According to Gupta and Denton (2008), the arranging of patient appointments can be categorized into three:

a. Single Batch Process

Decisions are postponed in this appointment scheduling procedure until all appointment requests for a specific time have been received. In surgery starting times, this model is widely used and allows full details to be scheduled such that an ideal solution can be sought by discrete optimization or heuristic approaches.

b. Unit Process

The system is expected to arrive one at a time and is scheduled at the moment the application arrives in this appointment scheduling process. An optimal solution is impossible to be uncovered by this method, but it can be approximated by studying the distribution of styles of appointment request.

c. Periodic Process

In this scheduling process for appointments, the requests are kept in a fixed-size buffer and are scheduled when the buffer is complete. This provides a closer approximation to the optimal solution by evaluating optimal or near optimal solutions at each point.

2.7. Government Regulation (Indonesia)

2.7.1. Minimum Hospital Service Standards

The standard of outpatient services is following the decree issued by the Ministry of Health of the Republic of Indonesia (129/Menkes/SK/II/2008) that, the waiting time for outpatients is ≤ 60 minutes.

2.7.2. Medical Practice License

As stated on PMK RI 512/MENKES/PER/IV/2007, the Ministry of Health of the Republic of Indonesia set the regulation that doctors are only allowed to practice a maximum of three places (hospitals).

2.8. Simulation

Simulation is the process of creating a stochastic model of a real-world event and then conducting sample tests on it. The stochastic model is the property that separates a simulation from a simple sampling experiment in the traditional sense. A simulation requires first and foremost the creation of an abstract model of the system to be investigated, whereas a traditional sampling experiment in statistics is most commonly done directly on raw data.

Before modifying a current system or building a new one, simulation is used to decrease the possibilities of failing to meet requirements, eliminate unanticipated bottlenecks, avoid under or over-utilization of resources, and improve system performance. Simulation, in its widest definition, is a technique for evaluating the performance of an existing or prospective system under many configurations of interest and over lengthy periods of time.

2.9. Research Gap

Many researches on the functioning of a healthcare system have been conducted using simulation models. Santos *et al.* (2013) developed a computational framework that simulated a healthcare professional's decision-making process and gave a way for evaluating team performance in order to enhance patient safety. Their work, in particular, provides a greater knowledge of and more help to individuals in making better decisions for enhanced patient safety.

Furthermore, simulation is extensively used to examine healthcare services and patient flows, and it has received a lot of academic interest in the last ten years. For example, Wijewichrama and Takakuwa (2005) utilize simulation to establish an appointment schedule that reduces both patient waiting time and doctor idle time, as both are equally important. According to Gosavi (2015), the simulation optimization technique has been shown to be effective in determining optimum scheduling via simulation. In addition, Lailomthong and Prichanont (2014) built a simulation model to investigate appointment scheduling in order to decrease patient waiting time, and discovered that if the patient is late, a larger appointment interval is proved to be more successful.

The former researches made significant contributions to lowering patient waiting times in a variety of contexts, but few addressed the issue of patient unpunctuality, particularly in situations where patients come earlier than their planned appointment time. Patient unpunctuality is a serious problem in healthcare facilities, and it causes considerable hospital overcrowding, resulting in patient discontent and physician overtime, which is the impetus for this study.