**با سلام این سوال رو استادم ازم پرسیده و باید بهش دقیق پاسخ داده بشه. خودم فایل مقاله هم در پایین صفحه است همه بخشهای اصلی و مدل قرار داده شده.**

After examining its content, I am afraid that the paper lacks significant contributions that advance the theory or practice of production economics.  It is just a reformulation of classic problems with integration of queueing models and interger linear programming appraoch. In addition, some modelling details have not been presented clearly, for instance the ILP details, as well as the assumptions in the queueing network (should the departure process in one node still be Poisson, if capacity is contratinted or there is a splitting of flows)

**Healthcare Service Supply Chain Network Design using Queuing theory**

**Abstract**

The development of the health supply chain network has a significant effect on the efficiency of the system. Therefore, it is necessary to simplify and control materials' flow by designing an efficient healthcare supply chain network. The novelty of this article is related to the novel queuing model. Patient flow analysis is one of the essential scientific topics that advance healthcare service chains. In this research, the proposed decision-making model resolves the patient selection problem for effective DRG treatments. An integer linear programming (ILP) model was proposed to determine the number of patients accepted in the hospital according to the capacity constraint. In the ILP formulation, the queue model considered that patients' arrival is Poisson distributed, and G/G/1 or G/M/1 queuing model is used to complete the overall therapeutic service chain. Moreover, a case study in a People's Hospital in cancer center is supposed, and the results are reported. The results show that patients' satisfaction with patients' priority under different modules (regular, injured, or emergency patients and VIP patients) has increased.

**Keywords:** healthcare service chain network; diagnosis-related groups; response time; Poisson distribution, supply chain.

**1. Introduction**

In recent years, researchers and organizations have paid more attention to supply chain management services in the health care system (Ellram and Cooper, 2014, Karimi et al., 2014; Haddadsisakht and Ryan, 2018; Nagurney, 2021). The Healthcare supply chain is used to integrate suppliers, health centers efficiently, and emergencies center considering reducing the system cost, improving the health care services, and expanding the service distribution to the correct number at the right time and place (Masoumi et al. 2017; Mahdavi et al., 2018). Based on these facts, small firms and companies are in high competition in designing, modeling, and optimizing supply chain systems. Activities such as procurement of materials and service planning, supply and demand planning, maintenance of goods, distribution, inventory control, delivery, and customer service all were previously done at a company, although at present, it is done at the supply chain level. They incorporate discrete probabilistic reasoning into supply chain network performance and propose the equilibrium conditions as a multivariate logit-dependent nonlinear optimization problem, assuming clients are complex in time-cost bi-criteria decision making. For model representation and operational insights such as benefit maximization for a firm that engages in this supply chain network, numerical examples are given ( Ma et al. 2020). Investors and supply chain executives in the healthcare sector always require convenient and reliable strategies for making decisions. Such decisions have a strong influence on all configurations and future operational aspects of the health management system like supply network design, inventory control system, distribution and logistics, and transportation system (Wang et al., 2015; Ramezanian, R., & Behboodi, 2017). Zahiri and Pishvaee (2017) indicated that decision models in the supply chain healthcare sector could considerably reduce the entire supply chain cost. For addressing, such a complex optimization problem, various modeling approaches and solution approaches can be used (Shakibayifar et al., 2017b, Zahiri et al., 2014; Zokaee et al. 2017). In addition, this decision is the most influential factor in the efficiency of investment in the healthcare service chain with the cost-cutting mechanism, resulting in increased profitability and increased efficiency and competitiveness. Khalilpourazari et al. (2019) This study introduces a new multi-objective computational model for designing a disaster-resilient health supply chain network. Five multi-objective and decision-making approaches with Pareto optimal solutions were used to solve the proposed multi-mixed integer linear programming model. Then, to show the most significant parameter, sensitivity tests are conducted to investigate any changes in essential parameters on the optimization problem.

Talaei et al. (2016) believe that increasing the reliability of the supply chain through an optimization model in network design corresponds to long term and tactical strategies of modern companies. Goodarzian et al. (2020) They develop a general multi-objective multi-echelon multi-product multi-period pharmaceutical supply chain network and the production–distribution–purchasing–ordering–inventory holding–allocation–routing issue in the presence of difficulty. To deal with uncertainty variables, we formulate a Mixed-Integer Non-Linear Programming model and create a novel robust fuzzy programming process. The findings support the algorithm's ability to find a near-optimal solution.

The proposed query mathematical model considers the health service supply chain of various patient groups in this study. The objectives of the mathematical model are to answer decisions related to the patient satisfaction of receiving services during treatment and how the decision-making model decides on patient selection for effective DRG treatments. An objective ILP (Integer Linear Programming) formulation assumes that patient arrivals are Poisson distributed. The G/G/1 or G/M/1 queuing models are used to complete the overall therapeutic service chain function to minimize total healthcare supply chain costs and total transportation time. Also, the constructive effect of the healthcare service chain is considered in the queueing model to determine the number of patients accepted in the hospital according to the capacity constraints robust and resistant service supply chain during the disaster.

The present study addresses a supply chain network design in the healthcare sector utilizing queuing theory through a case study in a hospital in People's Hospital. The second section points out the recent study reviews in this area and draws the research contribution. It has several sub-sections in order to indicate a comparison with our model. Section 3 develops the proposed model for supply chain network design, its characteristics, and details. In section 4, the solution methods are explained at the end sections 5 and 6 present the results and discussion of the model implementation model and conclusion, respectively.

**2. Literature review**

Management of the healthcare service supply chain is one of the latest scientific topics that has influenced recent progress in an advanced industrial management system's attitude in developing countries (Zhou et al., 2014: Chong et al. 2015; Govindan et al. 2020).

***2.1. Queuing theory in healthcare systems***

Several enterprises, such as businesses, airlines, telephone corporations, and police forces, frequently use queueing models to assess the levels of ability required to respond in a timely method to experienced requirements. Although queueing analysis has been employed in hospitals and other health situations, it is not widely used in this field. However, given the ease at which healthcare problems are readily available and the fact that most healthcare facilities are willing to satisfy rising demand with severely reduced capital, Queuing models can also help in the creation of more efficient bed allocation and staffing procedures, as well as evaluating other approaches to improve service allocation and staffing (Nawusu, 2020). The queuing theory is a mathematical representation of waiting entities in which indicators, such as mean queue length, mean waiting time, maximum queue length, and mean waiting time, are evaluated in the long term to analyze the system's status efficiency. An accurate estimation of such indicators provides a correct decision basis for the service and manufacturing industries (He and Hu, 2014, Radman and Eshghi, 2018).

In the previous literature, the queuing models were mainly used to analyze the trade-off between servers and customer service time. The service expense will be raised if the number of servers is high. However, the customers' waiting time is low (Bhaskar and Lallement, 2010, Onggo et al., 2018). The queuing systems compute the optimal number of the server to minimize the operating costs. It considers the average entry rate for service requests, average service time, loss caused by the customer waiting (hidden credit cost), and serving operating costs. Queuing models are used to obtain functional efficiency criteria such as queue length, serving time, and waiting time and determine the probability of a specified delay. Determining the required number of servers to satisfy anticipated demand and determining the reliability and number of various kinds of server facilities are two examples of queuing issues (Bandi et al., 2018). The application of queuing models has been established in the area of healthcare supply chainanalysis. The past research in healthcare service models has addressed the queuing models at both the policy and operational levels. A comprehensive review of the queuing models at the policy level was presented (Green 2006, Lakshmi and Iyer, 2013). At the policy level, the aim is to account for the legal perspectives and consider policies such as capacity sharing and demand management by monetary incentives to decrease patients' waiting time. On the other hand, they were queuing models at the operational level address operational problems such as appointment planning, congestion, and human resource decisions. Siciliani and Hurst (2005) provided a comparative study to get insights on policies to reduce excessive waiting times for noncompulsory surgery in twelve countries. They indicated that patient waiting times could be decreased by finding a trade-off between the surgery's demand and supply. Guo et al. (2014) introduced a two-tier queuing model as two parallel M/M/1 servers, with free and toll service options. The model was examined for the potential of toll service that maximized healthcare. For modeling and preparing Outpatient Appointment Systems, Ahmadi-Javid et al. (2017) presented a systematic analysis of recent computational and numerical queuing theory models and optimization methods (OAS). Ewing et al. (2017) created a method to integrate the Real-Time Locating System (RTLS) with other EMRs to show how the merged data can be utilized further to comprehend patient flow, bottlenecks, and patient-provider relationships to enhance Emily Couric Clinical Cancer Center (ECCCC) facilities.

Wan et al. (2017) developed a two‐level healthcare service model to analyze the cost of waiting for patients. This queue model provides a new approach to classified patients based on their waiting cost by price discernment mechanism. The objectives were 1) minimizing total waiting cost for all patients, 2) maximizing public gain for the service provider, and 3) maximizing the profit for the private service provider. Zhong et al. (2017) studied similarities between healthcare systems and production systems in design, performance evaluation, and continuous improvements. They investigated the differences and difficulties caused by variability, constraints, dynamics, and human behavior. Accordingly, some recent potential advances for care operation planning, patient flow management, and the health care delivery system were identified. Hu et al. (2018) use queueing theory in modeling Emergency Departments (E.D.s) and assessing the strengths and limitations of health care systems. The model's outputs were compared with the result of discrete‐event simulation, and the advantages were reported. Bai et al. (2018) conducted a study of the Intensive Care Unit (ICU) and focused on the ICU management challenge from a variety of angles, including decision horizons, problem settings, modeling, and solution strategies.

***2.3. Integrated approaches***

Many researchers have proposed queuing theory to determine quick solutions to the healthcare service chain problems. In service chain modeling, by using a queue, each service system's primary process is considered the behavior of orders. This makes it possible to evaluate the efficiency of a supply chain management system. Its inputs are requests for patient service, and its outputs correspond to the level of patient satisfaction. The overall challenge in the process-oriented approach is the correct initialization of each process. This initialization is similar to the time delay setting or operation time with a safety margin. Existing statistics (average response time) or the system's current state may be used to derive these values (waiting time and service time in the nodes). The most critical challenge in developing countries' health service supply chain is improving efficiency and reducing waiting costs. This issue can be visualized as a physical network of services with a queuing paradigm, with each source containing a server and waiting events in the queue. Each event includes indexes such as the number of requests or referrals process, the amount of service, and delay. Using the traditional efficiency criteria that have been provided for queuing systems (such as the average queue length, average serving time, etc.), the minimum delivery time of an order can be estimated. The increasing development of integrated healthcare supply networks decreases annual costs and increases efficiency for public health services. In practice, a range of essential factors influences the proposed planning and growth of healthcare logistic networks. Based on the amount of delay, establishing the best policy to launch a process is computationally challenging. Also, the efficiency criterion can be estimated by calculating the number of methods that have reached their goals. Also, all procedures need to be controlled. In general, the control method acts as an intermediary feature between the outcome and the target during the life cycle of a process between the result and goal. Since the operations compete to have access to the limited resources, some resource conflicts or deviations may occur due to the resource breakdown, such as malfunctions of computed tomography (C.T.) scans or Magnetic resonance imaging (MRIs).

Moreover, preparation times and shared activities (such as a patient's transportation outside the healthcare service center) may cause resource conflicts. The probability of re-routing the actual flow from one node to another resource for the next operation is one of the decision variables. This problem is called a routing problem. Despite the importance of integrated approaches to healthcare service chain management, there are few scientific efforts in this area. Maghsoudlou et al. (2016) proposed a bi-objective optimization model for a three-echelon multi-server supply-chain network in distribution systems. The model is based on M/M/m queuing system for the design of cross-docks facilities. Because the operation's timing is significantly different for each case, recipients must wait in line before the procedure. Another critic is that they could consider queuing theory with Modeling and analyzing the problem based on G/G/m queuing models and consider another heuristic algorithm to compare with that to see the strength and quality of the research. Also, in Maghsoudlou et al. 2016, prioritization is based on each member's importance (based on demand and critical time for each member). Because each member's hospitalization operation has a unique operating time, other patients' waiting times for hospitalization should be considered. This waiting time directly affects the severity of the disease or the patient's chances of recovery and should be significantly increased. In the proposed model, we try to address the above problems and build a much reliable one.

***2.4. Research motivation & contribution***

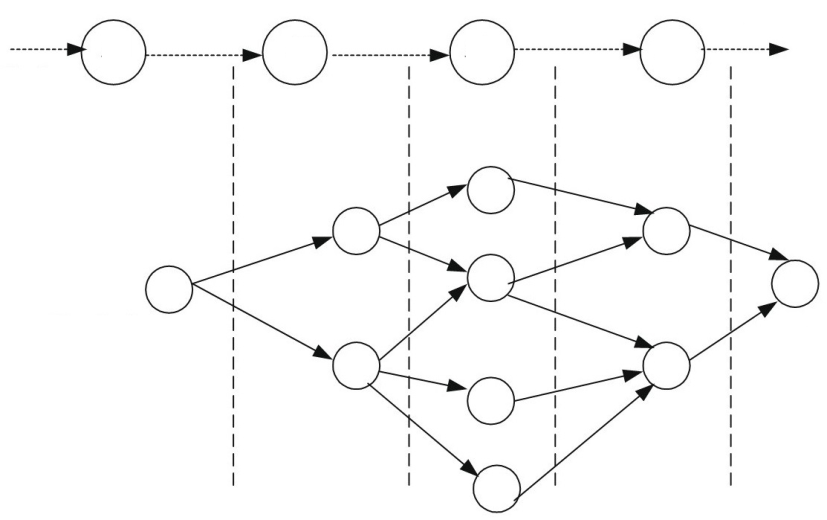
One of the primary *motivations* of this research is applying an integer programming model to determine the number of patients who should be accepted in the hospital according to the hospital capacity using queuing theory in the healthcare service supply chain. In this study, the health supply chain problem is formulated using the queuing approach. The aim is to have the highest benefit and patient satisfaction with the healthcare system. Typically, hospitals need a combination of Diagnosis Related Groups (DRGs) to classify patients who are accepted. DRGs is a patient classification scheme that provides consistent prospective hospital payments and promotes cost containment initiatives. For formulating this problem, patients are categorized based on their types of diseases, and every patient is modeled as a unique random process. The proposed model assumes that patients' arrival follows Poisson distribution while treatment time follows a general probability distribution. The significance of DRGs is also considered to be constant, and medical care facility is limited. The percentage of such a disease is described as a random process after the type of infection is diagnosed. Also, to validate the health supply chain model, a case study was implemented in a cancer center of People's Hospital,and the research findings are interpreted. The main contributions of the present study are twofold:

* **First**; A novel queue model is developed to determine which patients should be accepted and when the reward of efficiency should be increased due to the capacity constraints. A queueing model is created in order to forecast queue lengths and waiting times. Since the findings are often used when creating management decisions on the services required to deliver a service, queuing strategy is mainly known as a branch of operations science. A network of queues is a structure in which a variety of queues are linked via patient networking. When a patient is served at one node, they can either move to another node and wait for service or abandon the system.
* **Second**; A Mixed Integer Programming-based heuristic solution approach is developed to solve the real-world instances of the problem. The novelty of this article is related to the specific queuing model. The researcher wants to make a clear pattern and a realistic view of the health care network, especially in the cancer clinic. A realistic case study is used to demonstrate the strength of the proposed model, including a variety of features and challenges. The proposed integer programming model's validity is often evaluated using a real case study. Our findings show that a platform's solution productivity and efficacy can be optimized by patient satisfaction. This research looks into how power, illnesses, and medical routing choices in healthcare systems can impact patient satisfaction in both preventive and curative situations. We assumed that patients' arrival follows a Poisson distribution, while treatment time follows a general probability distribution in the proposed model. A case study of cancer care in People's Hospital, was used to validate the model's validity. Based on the current capacity constraints, the built model will decide which patients should be admitted to the hospital.

**3. Proposed Model Interpretation & Mathematical formulation**

Another feature of the model is the integrated facility, in which medical centers and hospitals are integrated into one place, saving on fixed costs and increasing access to resources. By combining facilities, patients are less likely to move, resulting in a lower risk of death and patient transmission complications. We must jump to conclusions about When choosing a queueing model, consider the probabilistic existence of the arrival and service systems. When it comes to arrivals, the most general belief is that a Poisson procedure is used. The term is assumed to be derived from a Poisson distribution, the number of arrivals at any moment. Then if N(t) is the number of people arriving, t and N(t) has a Poisson distribution over t time. The Erlang or M/M/s type of delay is the most often used method of queueing. This model uses a particular queue that feeds into its equivalent servers with an infinite waiting room. Customers arrive at a steady rate, including a Poisson method, and there is an exponential distribution of the service period (e.g., LOS or provider time linked with a patient). When the two "M's" have been used in the model notation, they are sometimes referred to as Markovian. The same mean for the normal distribution of the exponential distribution in M/M/s implies that CV is about the same time. The M/M/s would also have reasonable time predictions whether the actual operation CV is small or close to one. Suppose the CV is much smaller than one. True delays, in this case, will be significantly underestimated or overestimated by the M/M (recall that in bad uncertainties, the model will overestimate delays); conversely, if variability is large, the model will underestimate delays). In this case, the average delay for any service delivery is already calculated when Poisson is in the arrival process. There is only one server, using the following equation, defined as the M/G/1 method (Raj, 2020).

In integrated approaches, a method is a combination of resource-consuming and system operations. Fig. 1 shows how, as a physical process, the logical concepts are drawn. The functions represent the steps in the complete explanation in this figure. Each activity is carried out with available resources in one location. In general, the output of a specific activity can be of high relevance to the next operation. If the locations of the different parts of the healthcare system are geographically dispersed, a "service" means "treatment" or the convergence of physical flows of serving to a final point.



**Input**

**Logical Mode**

**Activity 1**

**Recovery**

**Activity 2**

**Activity 3**

**Activity 4**

**End of Process**

**Start node**

**Physical Flow**

**Final delivery**

**Fig. 1.** Physical flow in the healthcare service supply chain (**own elaboration**)

***3.1. Problem classification***

In this section, before problem formulation, the taxonomy of different queuing models is provided. The following are several other queuing models that can be considered to develop supply chain management systems:

* The M / M / 1 model, in which the Poisson input process and service time are exponential. The M / M / 1 queuing model has a clear definition: the number of customers in the system at any given time. For M / G / 1 queuing models in which the input process is Poisson. Still, the service time does not follow a particular distribution; The number of consumers in the system and the number of services that have been provided to them so far determine the system condition.
* The G / M / 1 queue model is a hybrid of the M / G / 1 and the G / M / 1 queue models, in which the input process does not require a specific distribution, but the service time is cumulative. The state definitions are expressed under the equilibrium conditions at a given moment before a task is triggered. Based on the Ergodic theorem, the distribution of the states at random time moments is unacceptable. It means that the input arrival process does not follow a Poisson distribution.
* For M / Ek / 1 or Ek / Ek / 1 models, Ek is the Erlang distribution with stages; various premises can be used to calculate the likelihood of packet loss delay. It refers to Erlang B if the blocked orders are discarded. In the case of the secured orders remaining in the queue until they are done, the process is a type of Erlang C. It is worth noting that Erlang B and Erlang C models were applied mainly in traffic and transport systems ( Hassannayebi et al., 2017, Shakibayifar et al., 2017a).
* The waiting time distribution is almost exponential for G / G / 1 and G / G / m queue models (where m is the number of servers) only when the suggested traffic is too large (efficiency parameter is close to one). When ρ≈1, the waiting times are incredibly long. The waiting time is not allocated in any specific way for other traffic flow values; thus, the state definitions are unacceptable at an arbitrary time. Therefore, M / M / 1 queue models are suitable for supply chain management problems due to simplicity and responsiveness (Gradshteyn and Ryzhik, 2014).

In addition to the taxonomy mentioned above, the application of the queuing theory in supply chain management can be classified as follows:

* From a mathematical point of view, queuing theory applications can be divided into two categories. In the first category, the arrival process and service times follow a known probability distribution. Second, some queues have standard, empirical, or hypothetical distributions studied using simulation modeling approaches
* From the structural characteristics of the service chain's point of view, the main parameters that are used in the strategic design of the service network are (He and Hu, 2014):
* Single-tier or multiple tiers queuing network
* Single-server or multiple servers on the network
* Single period or multi-period planning horizons
* Deterministic or probabilistic parameters
* The extended capacity of appointment centers and servers (limited or unlimited)

***3.2. Problem description***

The supply chain healthcare system is a combination of some primary and supportive activities to meet patient demands. These preliminary activities in a healthcare system, including appointment, visit, radiology, and hospitalization, are modeled to evaluate supply chain performance with a three-stage process. Physical, operational resources support these activities. This model's concept is the following: The hospital needs a combination of DRG cases that determine which patients should be accepted and when the reward of efficiency should be increased due to the capacity constraints. The significance of DRGs is presumed to be known, and the hospital's support is given. The set of inputs of that type of cancer is modeled as a random process after the disease is diagnosed. The patient is treated based on the known distribution of probabilities upon admission.

As presented in Fig. 2, these resources are distributed geographically, and health service channels connect them.

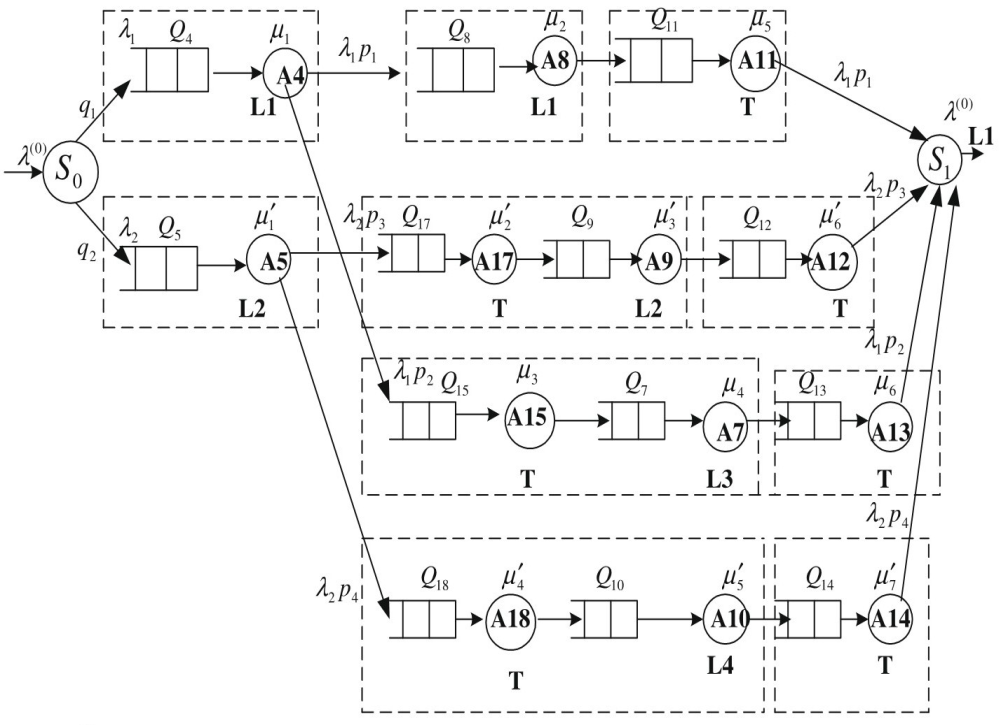
* The patient's entrances are in L1 (triage) and L2 (Emergency).
* After diagnosis, transition places are in L1, L2, L3 (hospitalization units), and L4 (surgical teams).
* Initial tests are taking place in the L1 unit.

Here, there are two selections for the path of material flows in two stages of the processes:

* Entrance: from L1 or L2.
* Hospitalization: From L1 to L1 or L3 or from L2 to L2 or L4.

As shown in Fig. 2, each source is modeled as the queue in which accumulators are waiting for the process to be performed. Path selection can be made by considering the total delay estimated from to. This delay consists of health care delay (dependent on cumulative waiting time) and entire waiting time in downstream queues.

The mater problem can be easily broken down into the example raised in this study. In other words, the service request that follows a particular distribution at the entrance rate is formed as a network. As a result, this queueing model was examined by looking at how the input and output processes followed a particular pattern (Mehmood et al., 2017). In stage 1, each output from L1 and L2 has two potential connections or routes. L1 output in Stage 1 is related to L1 and L3 nodes in Stage 2, while the production of L2 in Stage 1 is related to L2 and L4 nodes in Stage 2. Finally, the A11 and A13 servicers' output enter into S1, and the A12 and A14 outputs also enter into the S1. Therefore, nodes do not interfere in different paths. Dealing with this particular case is essential for matching "urgent orders" ("Urgent patients") and "regular orders." Urgent orders need quick service, and on the other hand, regular orders require regular assistance. Requests can be sent for processing to L1 and L2 depending on stage 1 and subsequently to L1, L2, L3, and L4 in stage 2.



**Exit**

**Exit**

**Exit**

**Fig. 2.** Block diagram for modeling a processes network queue (two inputs network). : Index of resources, S1: Entrance site index (triage), L1: site in triage, L2 and L3: hospitalization units, L4: Surgical sections, and T: Transport index **(Own elaboration)**

The serving time and the rate of input orders can have distinguishable distributions. For example, serving time can have a normal logarithmic distribution considered for the supplier's delay G/G/1 or G/M/1 queuing model can model a complete and overall therapeutic service chain. The whole challenge in this work is to estimate an efficiency criterion that can propose a good service quality (such as the lowest customer response time). In a chain network, the value of estimating the minimum response time is highlighted in the following. The collaboration between activities in various systems should occur in a given time window to deliver a real-time system, such as the healthcare service chain. The unavoidable communication delay is a combination of uncertainty in transmitting a message in a communicative device and a delay in delivering information to the destination. A simple delivery approach is to consider a lag when entering an order (patient) called an "on-demand" approach. Upon entering an order in the source node, this order is routed through all the central nodes to reach the end node. The goal is to discover a path between the origin and destination nodes that takes the least amount of time. The path in the complex chain number of health care services relates to providing an order to the final node from the source node with the minor overall delay and the network. This problem has a stable and fixed structure and meets the various capacities of serving and maintenance tasks. The ideal way links the begin and finishes nodes with a few intermediate nodes; it is the shortest path because it takes the least amount of time to react. It is important to note that all entrance patients (inputs) are not placed in the quickest way of serving path due to overcrowding. Therefore, it is recommended that the serving rate should be increased or doing time should be reduced to allow orders to be re-distributed for the system's load balance in all other routes.

***3.3. The mathematical model***

Bring up a set of patient types and possible treatments, the probability that an n-type of patient needs treatments of subset m is . The period of treatment has a density distribution with mean . The number of patient inputs of the type in a time interval with a random variable is determined. We consider a data model that describes the number of correct type treatments, at time of t, the hospital has the resources to handle it.

Let denotes the distribution of the number of patients treated under m-type treatment at time *t* and is the distribution of completed m-type treatment at time *t*. and take values according to the type of accepted patients. The preferred DRG case combination has been shown with the weight and represents the relative importance of treatment *m*. The issue is identifying who a patient is for the first time they are accepted, and it should be noted that any time after, the n-type patient should also be recognized. The objective is to determine how the number of weight treatments (weight-based on the ) increases, while it must be sure that the number of remedies does not exceed a fraction of the time. Also, a hospital with capacity aims to maximize the number of treatments it provides where they can exceed the time fraction's allowable capacity ().

The value of an input parameter shows the hospital's unwillingness to perform surgeries beyond their capacity. The high value means that the order is more than the total (order congestion), during the low value implies that demand is less likely to exceed capacity. Suppose be the vector of selected times for accepting patients of type *n* and the resulting rewards by the collective weight of the treatments completed for the decision. The cost of computing is the discount factor that ensures the upstream costs have also been considered. Finding optimal results in the following mathematical formulation:

|  |  |
| --- | --- |
|  | (1) |

Where;

|  |  |
| --- | --- |
|  | (2) |

The rewarding mechanism in Eq. (2) is based on the number of completed treatments and the motivation behind the hospital's financial structure that is being studied. Other options are also possible. Relevant alternatives that the decision process decides include:

1. in which its reward is based on the deduction and fraction of patients over the capacity.
2. Its compensation is based on cost overcapacity.
3. Its reward is based on the expected number of patients treated with *m*.

Capacity parameter () refers to the amount of current treatment *m* that a hospital can handle at time *t*. This decision has been made based on how much time is allocated for treatment. Envisage an MRI machine that is used for 2000 minutes a week for type A and B treatments. Suppose that each type of A and B treatments require 10 and 20 minutes, respectively. One method for dividing 2000 minutes into types A and B is to allocate 1200 minutes to type A and 800 minutes to type B, which leads to type A treatment and to type B treatment per week.

There are various ways to distribute 2000 minutes between the two types, and consequently, there are different probabilities for and. However, in the proposed model, the decision is how much time must be devoted to any treatment. It means that the goal is to find a unique value for and. In other cases, it is desirable to include the resource allocation decisions as a part of the model. Thus, in this case, the demand for resources is determined by multiplying the number of current treatments by the treatment time . In conclusion, to solve Eq. (1), we need to decide how random variables and can be determined for the vector .

***3.3.1. Queue model***

A Queue Analysis represents a numeric summary of the queuing process, including some premises on the probabilistic existence of the arrival and operation tasks, server number and form, and queue organization. There are many queueing patterns to choose from, but we will focus on the more popular ones in this section. Many other models offer formulas for a simple assessment of various performance indicators that can help build or enhance an internal service structure. A simple queue system is a service system where "customers" come into a service center and order one of the workers to provide a service. It should be noted that all parties waiting for service is a 'customer' and must not be an individual. In a "back office" scenario, the "patients," for example, will interpret radiological images, perhaps read pictures.

The direction is determined by a server,' which is a person or object. Inpatient beds will be required servers when determining waits for patients in the emergency department (E.D.) awaiting admission to the hospital. When a client enters, if all servers are filled, a queue must be entered. Queues are also described as precise lines of people or things, but they may often be invisible, such as waiting in line for a phone call. The concept which determines the order of queued customers is recognized as the queue protocol. The best-established approach is the FCFS law, but other strategies are also used to make consumers more time-sensitive or reduce their wait time. A good illustration of priority queue discipline is the triage procedure in an emergency department. Priority areas can be preemptive or non-preemptive, depending on whether a higher priority client arrives and interrupts a service in progress. In most queueing models, there is no limit on the number of customers waiting for service, implying a large waiting area. It is a rational assumption when patients do not physically reach a queue, such as in a mobile call center, or where the real location where customers wait is high relative to the number of customers that typically wait for service. Even if there are not waiting room capacity restrictions, newcomers who see a long line can "balk" and refuse to join the line. It could happen in a walk-in clinic. Reneging is a similar feature used with some queueing systems, which happens as customers feel friendly and exit the line before being served. A description of this action is seen in some E.D.s were "left without being seen" is frequently referred to by patients who reach an agreement (Nagariya et al., 2020). In this study, the health supply chain network is formulated as a queuing system. The main objective is to determine the number of patients receiving treatment () and the number of patients who have undergone treatment , since the population is large and people get sick on their own, Poisson's arrival process is a normal belief. Furthermore, since seasonality and inclination are common in many diseases, it is common to think of a variable disease mechanism. For general distribution functions, the Model can be considered proportionate to the amount of time the patient spends on care. The consistent method allows different treatment types to be taken into account quickly and accurately. Numerous studies have shown that there is a real-world queuing theory in the health system. They have also introduced the Poisson distribution as a good distribution for the entry rate. In this paper, due to changes in the quality of transport between two entries, environmental conditions, and traffic load, the allocation for modeling the entry rate of donated organs to transplant centers is considered. Also, according to peak hours, it is assumed that the entry rate. The median and service rates are constant, and patients entering transplant centers follow the Poisson distribution. The model is presented as a G/M/1 system based on the queue formed by the entry of donated members into transplant centers. Thus, in the model, several service providers with infinite capacity are considered without any entry barrier. In the proposed system, the priority of each member is based on its index number. A member with a zero index has the highest priority and based on the ascending order of the index number; the stresses are reduced. The entry rate is calculated based on the number of members transferred to each center and the total service time from the waiting time in the queue, and the surgery duration.

Suppose patients become ill due to the Poisson process . This process indicates that the patients of type *n* who have been referred to the hospital receive a treatment distribution, which is the average value and is the fraction of the population that had selected the hospital when the patient admitted . Otherwise, . Without loss of generality, we assume that. With the probability ofa patient of type needs treatment, the parameter of the patient's arrival process of a kind equals to. The number of patients of type *n*, who require *m* treatment at time t, is modeled as unstable queue customers in which the rate of input is and distribution of congestion service time is given by. This condition is defined in Eq.3. The reader is referred to Ahmadi-Javid et al. (2017) for further details.

|  |  |
| --- | --- |
|  | (3) |

So that,

|  |  |
| --- | --- |
|  | (4) |

Where is the cumulative distribution function for which he is serving

(5)

Let be the sum of random variables of the Poisson process. Therefore, the Poisson distribution is modeled with the rate parameter. Also, it is referred to as the Poisson distribution in the unstable queuing model. Thus, the average number of patients with completed treatment *m* of type *n* disease at time *t* is calculated as and .

***3.3.2. Model Features***

Some findings on the endless server queues are listed in this subsection to illustrate the related optimization issues. The author (Tran and Do, 2000) is referred to for some proof. When resolving the problem, these findings are later used.

First, it is presumed that the current patients are separate in the unlimited server queue model. It implies that patients or their arrival orders are not influenced by the time in the system. It suggests that, because these patients are not impacted by existing patients and are thus not impacted by subsequent decisions, the overall advantage of accepting a different kind of patient is seen to be the admission of time.

Therefore, the reward for adding a patient of type *n* by calculating time is as follows:

(6)

Also, the function time is formulated as:

(7)

In this formulation, the rewards are calculated . The second observation is about Eq. (1) and parameter. The percentage of distribution can be accurately estimated. If represents the percentage of the distribution of *x,* then:

(8)

Based on this observation, Eq. (1) can be rewritten as follows:

(9)

**4. Solution methodology**

In this section, the solution methodology is described in detail. The objective is to improve the patients' satisfaction concerning the priority of patients with different types (regular, injured, or emergency injured, and VIP patients). Fig.3 illustrates the process of solution.

**Fig. 3. The flowchart of the solution method**

Accordingly, in this section, two Integer Linear Programming (ILP) methods and an Approximate Solution Approach (ASA) are proposed for solving the model. The validity of the model is tested using a case study and some numerical examples. An actual instance of the People's Cancer Center's oncology department is used to evaluate the ASA's efficiency. Also, a sensitivity analysis of the parameters is conducted. The purpose of sensitivity analysis is to identify the cases in which the ASA provides similar results compared with the ILP method. As another example, Emergency Oncology in People's Hospital is keen to expand and provide new treatments with new equipment. To give a financial budget for such equipment, managers have recognized the most valuable DRGs and prioritized them. For the growth of such DRGs, hospitals should determine what type of patients should be accepted more and what time. In what follows, the proposed model is applied to support such a decision for specific surgeries. Because the treatment options for each disease are limited, it is the hospital's responsibility to decide which patients should be accepted. For example, the hospital's historical records show 37 different diseases and 109 other cancer diagnosis methods. In this case, some types of conditions do not require any special treatment, so these 37 types of diseases could be divided into 109 different treatments to have sub-problems with smaller sizes. These sub-problems are called a subset of the "special surgery" problem. It includes all the surgeries and related types of disease, resulting in six conditions and 34 treatments. Seven of these 34 are composed of 90% of the total treatment, and only seven cases were considered in this study. The planning horizon is 11-years, and every month new diseases are accepted (). The management of special surgeries requires rejecting new patients.

On the other hand, it is expected that the volume of existing diseases will increase while the hospitals absorb new patients. It is analogous to the same issue in which an improvement in the rate of arrival with a form of illness is a different type of disease. The objective is to decide whether this new form of illness is to be recognized. This decision is based on the priority and severity of the patients. Management has predicted how much disease will increase, according to the incidence of a new disease type. All data, except for these estimates, are derived from the historical records of the hospital. All types of diseases described elsewhere in this section represent a new type of disease. The ILP was solved with the ILOG CPLEX solver 12.2, and the queue model's parameters were calculated with Microsoft Visual Basic. The complex model utilized in ASA is resolved by employing the Dijkstra algorithm. According to the obtained results, all test problems were solved by ILP and the ASA method in less than 10 seconds, which is better than previous research. The percentage of n-type patients treated under treatment for both regular and emergency conditions are shown in Tables 1 and 2. Since a patient can use multiple treatments, each row's total probability is more than 100%. Equally, this sum can be less than 100% because only seven types of complete 34 treatments have been considered.

The relative importance of DRG () is related to the hospital's amount of reward for each case. Accordingly, the relative importance of DRG is ordered as: .

Given the fairness policy, the predicted arrival rate (patient per month) for each n-type patient in both ordinary and emergency cases is provided in Tables 3 and 4. The average service time () for each treatment of m-type patients in both the standard and emergency conditions is also given in Tables 5 and 6. Moreover, the Capacity () for providing m-type treatment for both the standard and emergency states in the planning horizon, Tables 7 and 8.

**Table 1:** Percentage of using n-type patients from m-type treatment in the normal state (the percentage of applied m-type treatment by the n-type patient)

**Fig. 4.** The frequency of use of n-type patients from m-type treatment in normal mode

We can understand that we have considered two types of cases (Normal modes) based on the figure. In this case, the frequency percentage of patients and treatment percentages on series 1,4 and series 3, 6 have more effect than the other series.

**Table 2:** Percentage of n-type patients from m-type treatment in emergency mode

**Fig. 5.** The frequency of use of n-type patients from m-type treatment in an emergency state

Based on the figure, we can understand as we have considered two types of cases (emergency modes). In this case, the frequency percentage of patients and treatment percentages on series 2,3,6 and series 3, 6 has more effect than normal patient. Hence, based on those two figures, we can realize that in both conditions and due to patients' disease, those in emergency conditions have been treated with enough resources and service to get satisfactory results. However, the cost will be the same for the two schemes.

**Table 3:** Average input rate (patient/month) for the n-type patient in a normal state

**Table 4:** Average input rate (patient/month) for n-type patient in an emergency state

**Table 5:** Average service time for m-type treatment in a normal state

**Table 6:** Average service time for the m-type treatment in emergencies

**Table 7:** Treatment capacity of patients with m-type in the 11-year plan in the normal state

**Fig. 6.** The frequency of m-type patients in the 11-year plan in normal mode

Fig. 6 and Fig. 7 concluded that type 1 (lung cancer) and type 2 disease (membrane lung cancer) have the highest therapeutic capacity in the hospital. According to Fig. 6, the therapeutic capacity trend from 2009 to 2020 regularly increases in the hospital. Also, according to Fig. 7, the therapeutic capacity trend does not follow the same direction.

**Table 8:** Patients with m-type treatment in the 11-years plan in the emergency state

**Fig. 7.** The frequency of m-type patients in the 11-year plan in emergency mode

**5. Result and discussion**

The model outputs, including the optimal reward and the time of accepting various diseases, are provided in Table 5. The values of tn for both regular and emergency cases (based on the fairness policy) related to the optimal time to add a patient of type *n* are summarized in Tables 9 and 10. In other words, type 1 and type 2 patients should be immediately accepted, and patients of type 3-6 should be accepted a few months later. This trend is set based on the fairness policy. By comparing the number of patients treated with existing capacity, it was found that the number of treated patients is less than the total for all types of treatments (except the 7th treatment type). According to the obtained result, the inequality holds. As a result, the 7th treatment type is the source of the in the healthcare system. We use the queuing model to examine the incentive shift by increasing the 7th treatment form power to solve this dilemma. Thus, the treatment capacity of type 7 has been increased by 10, 20, 30, and 40 percent, and the corresponding rewards have been analyzed. According to the results, enhancing the capacity up to 30% enhances the reward by more than two factors. Moreover, increasing the total does not increase up to 30% of the prize, resulting in a problem. Management is responsible for determining whether the high investment reward supports the capacity. The model has been solved several times to see how a further expansion would help the hospital manage supply and demand. The model can also be used to evaluate potential capacity limitation scenarios for particular therapies. The model outputs provide a reasonable level of resources as well as the final solution. The ILP solution is provided in Tables 9 and 10. The solution shows the amount of time and the reward values for accepting each patient. By applying this information, the hospital can establish a new approach to admitting the patient to the hospital. It can be inferred that the hospital should rely on the absorption of type 2, 5, and 6 patients in the short term (0-10 months). The hospital should concentrate on type 1 and type 4 for the next 10 to 30 months. Finally, in the long term, the focus must be on type 3 patients. The queuing model can also be used as a policy review by knowing the absorption rate of new patients to provide a unique patient treatment method in the long term. As our model is extended by mathematical modeling to implement some constraints related to the healthcare supply chain network, we have compared that to recent related investigations shown in Table 10.

**Table 9:** The initial solution for a case study in normal mode

**Table 10:** An initial solution for a case study in an emergency mode (monthly)

**Table 11: C**ompared some other methods

**6. Conclusion & Future research**

Today, healthcare centers have various problems in delivering services to patients with an acceptable level of satisfaction.

**References**