Journal of Management World **2025**, 1: 718-727
DOI: 10.53935/jomw.v2024i4.767

A Queueing Theory Approach to Analyze the Impact of COVID-19 Pandemic on Hospitals System Capabilities: A Lesson for Future Pandemic Preparedness

Balveer Saini[1*], Dharamender Singh[2], Kailash Chand Sharma[3]

[1,2]Department of Mathematics, M. S. J. Govt. P. G. College, Bharatpur affiliated to Maharaja Surajmal Brij University, Bharatpur, Rajasthan, India; veer.coiaf786@gmail.com (B.S.); dharamender.singh6@gmail.com (D.S.).
[3]Principal (Retd.), Rajasthan Higher Education, India; ks.pragya@gmail.com (K.C.S.).

Abstract. Pandemics are life-threatening and resource-intensive events. Hospitals with limited resources experience a surge in patients during such crises. These situations can lead to a high number of casualties within a short period. The "golden hour" after infection is critical for treatment and can significantly influence survival rates. It is essential to treat victims within this timeframe to prevent serious loss of life. Our goal is to improve the capacity of healthcare systems to treat victims during the golden hours in pandemics. This article evaluates patient arrivals and compares them with hospital treatment capabilities. We have developed a system called the Transient Act System (TAS), which detects hospital overcrowding. The hospital's treatment capacity is modeled using an M/M/s/GD/c/ multi-channel queueing system to estimate patient handling capabilities. Capacity calculations are based on the waiting time probability distribution function, with patient waiting times following an Erlang distribution. The proposed method has been validated through numerical analysis, including probability distributions and graphical representations. The study concludes with recommendations for future research and strategies to enhance hospital efficiency during pandemics. This research demonstrates the feasibility of the approach. The findings can assist policymakers and healthcare planners in making informed decisions during health crises, ultimately improving response effectiveness and saving lives.

Keywords: COVID-19 Pandemic, Erlang distribution, Hospitals Treatment Capability, Multi-Server Queueing Model, Priority Queue, Transient Act System.

1. INTRODUCTION OF THE STUDY

Throughout history, we all have seen various health crises and pandemics, such as Spanish flu, Ebola virus, Hong Kong flu, Zika virus, SARS (Severe Acute Respiratory Syndrome), AIDS, and influenza, etc., studied by Qiu et al. (2017). The most recent health crisis is the coronavirus pandemic. The coronavirus is a contagious disease that arose in Wuhan, China, in December 2019. The coronavirus, or Covid-19, pandemic has become a major issue worldwide, affecting numerous nations by infecting a large number of people. The COVID-19 pandemic has highlighted the challenges hospitals face in managing waiting lines. During the peak of the pandemic, most hospitals were overwhelmed with patients and faced lengthy waiting lines. The hospital's system was on the verge of collapse due to the increase in patient flow.

Queuing theory provides a powerful tool for addressing these problems relating to hospital management. By using different queuing models, service process control mechanisms, and optimizing system utilization, hospitals can better manage patient flow and improve their overall performance. Queuing theory can be used for designing, capacity planning, performance evaluation, and optimization of healthcare systems.

1.1. The Patients Scenario during COVID-19

The Coronavirus disease originated from a wet market in Wuhan, China, and a large number of infected people were found in December 2019, analyzed by Chen et al. (2020). Initially in Wuhan, approximately 300 people were infected by this virus, and 6 people died, described by Baker (2020). Despite early detection and intervention by the Chinese government, the virus has spread to many countries around the world. Thailand reported the first case of covid outside of China on January 13, 2020, with the patient being a returnee from Wuhan published in Aljazeera news (2020).

Due to its contagious nature, the disease has the potential to infect many people very quickly. As of February 2020, the number of infected cases increased from a few hundred to approximately 81,000, with 78,191 recorded in China. The rapid increase in corona cases led the World Health Organization to declare this health emergency a pandemic. As of April 2020, more than one million Covid cases had been recorded in the whole world, with around 62,000 deaths. Since then, the number of people infected with the disease has risen significantly, with over 170 million cases and 3.6 million deaths as of June 2021. However, more than 100 million people have recovered from the virus. These data stated in WHO press conference on June, 2021.

The pandemic has hit several countries across the world, involving the USA, the UK, Italy, France, Spain, South Korea, the Netherlands, Iran, Malaysia, and Thailand. The pandemic has also severely affected India, with 30,411,634 recorded cases as of 1 July 2021, leading to the discharge of 29,488,918 people and the occurrence of 399,459 deaths. As of 1 July 2021, there were 523,257 active cases in India published by the Govt. of India in the Covid report (2020).

© 2025 by the authors; licensee Academia Publishing Group
* Correspondence: veer.coiaf786@gmail.com

In such situations, Public Health and Social Measures (PHSM) have been implemented worldwide since the beginning of the COVID-19 pandemic. These measures aim to suppress the SARS-CoV-2 virus transmission, decrease morbidity and mortality from COVID-19, and prevent the healthcare system from overloading. PHSM's primary mission is to reduce the risk and extent of transmission, thereby reducing morbidity and mortality and reducing the burden of disease on the healthcare system. Some PHSM directly impacts the healthcare system; such measures of PHSM include partial or complete lockdowns, social and physical distancing, restrictions on travel and gatherings, and mask-wearing mandates described by Warren (2020). The Government of India, for instance, implemented a 21-day nationwide lockdown on March 24, 2020.

1.2. Queueing Theory in Emergency Healthcare Management

Queues, or waiting lines, are a common occurrence in daily life, including in the medical sector. Unfortunately, hospitals often view queuing as a symbol of inefficiency. During a pandemic, many victims must rush to the nearest hospital and wait for treatment. Treating the victims during "golden hour" can significantly influence their prognosis. Not treating the victims during this period could potentially result in a rise in casualties. The waiting time for treatment depends on the hospital's capacity and the length of the queue. Therefore, we need to approximate the victims arriving at the hospital and the hospital's treatment capacity at each point in time. To determine the number of victims who arrived at the hospital, we have created a "Transient Act System (TAS)." This helps us determine if the hospital's capacity is sufficient to handle the number of victims arriving. The classification of patients is also important for determining the order of treatment. The determination of a patient's health plays an important role in quality control.

This study is carried out to develop a queuing theory's transient non-steady state system. The study focuses on situations where there are varying numbers of servers and non-homogeneous arrival rates. In this paper, we describe the $M/M/s/GD/\infty/\infty$ type multi-server queueing model to calculate the hospital treatment capacity. The hospital treatment capacity calculation method is based on the waiting time probability distribution function, where the patient waiting times follow an Erlang distribution. In this model, each doctor is represented by a server, and the hospital is represented by the system.

Here's a summary of this paper: In the first part of the introduction, we discussed the pandemic situation, its impact on human lives, and the initial actions taken by the global community. In the second part of the introduction, we have described the importance of queueing theory in emergency hospital management. In the second section of this article, we present a literature review of prior research that is relevant to our current work. We provide the assumptions and notations that frame the proposed queueing model in the third section. In the fourth section, we have discussed "The Proposed Work." This section's first part calculates the number of patients who arrived in transient states. The second part outlines the severity level at which patients should begin treatment. We also provide a flow chart to better understand patient flow and the emergency treatment process. In the third section, we discussed the capacity analysis of a hospital system using the provided queueing model. The fifth section provides a numerical analysis of the proposed work, complete with a probability distribution and graphical representation. The sixth section concludes the overall findings with suggestions and guidance for future research work. Finally, we presented the importance and limitations of this study.

2. LITERATURE OVERVIEW

There is a comprehensive literature on the use of queueing theory in healthcare sectors. Researchers have conducted extensive research on emergency situations and related issues. This present work has significantly benefited from the literature on disaster management and capability analysis. It gives a clear direction for our research work and provides a basic structure to support the decision-making process in disaster management.

It is worth noting that the static models may not be effective in managing the sudden and rapidly changing demand for rescue or health services in pandemic situations. Wang et al. (2008) highlight these features in their research. They describe a single-class, multi-server queuing model with time-dependent arrival rates as well as varying numbers of servers. They estimate medical treatment capacity by dividing it into two cases, depending on whether the current rate of arrival is less or greater in comparison to the rate of service. In the first case, they utilize the model of M/M/s when the system is in a steady state. In the second case, when the medical capacity is insufficient, they limit the storage capacity to prevent overloading and maintain the system in a stationary state. This work deals with the non-homogeneous rate of arrivals along with the varying numbers of servers.

The studies by Caunhye et al. (2012) describe detailed information correlated to pre/post-disaster planning. It focuses on accommodation, distribution of relief, as well as transportation of casualties. In this study, mathematical programming, the theory of probability, simulation, and statistics are the most commonly utilized approaches. Kilic et al. (2013) have discussed how to determine the optimal rate of treatment after a disaster. Pradhan et al. (2015) conducted a study on the distribution of queue length in a batch service queue, where the capacity and service depend on the batch size. Hajnal Vassa et al. (2015) have applied queuing models to patient flow in emergency departments to evaluate the system's performance and manage emergency situations effectively. Agrawal and Singh (2017) proposed an analytical study of queues in the medical sector. They also presented a statistical analysis of related data. Yaduvanshi et al. (2019) have proposed some queuing theory applications to optimize the waiting time in hospital operations. Mittal and Sharma {(2020a), (2020b)} have described an assessment of Delhi NCR traffic using the queueing model. The authors proposed a probabilistic model to assess the queuing time of COVID-19 patients using a queuing paradigm. We outlined a sequential queuing model to estimate the time it takes to detect and identify infections in high-demand situations. In a different study, Mittal and Sharma (2022) proposed a simulation-based method to decrease waiting times at AIIMS, Delhi, through the use of a queuing model. Saini et al. (2024) have analyzed the role of queueing theory in the way of implementation of the Right to Health Act.

The study of Li et al. (2024) covered the issue of predicting the $M_t/G/\infty$ queue's service time distribution based on departure epoch observation. They studied the behavior of the minimax pointwise risk over a suitable family of service time distribution functions and created minimax optimal estimators of G. Hillas et al. (2024) investigated the functioning of multi-class, multi-server bipartite queueing systems, in which each arriving customer can only work with a subset of servers. They concentrated on the system's performance under the service management discipline of a first-come, first-served, longest idle server assignment. Aalto (2024) considers the optimal scheduling problem in a multi-server queue, which includes eager clients from various classes. They assumed that each customer experiences a random moment of desertion, and if they don't receive their service before then, they will exit the system. They used the Whittle index approach to solve complex problems. The discrete-time multi-server queueing problem with discounted costs is approached using the Whittle index method. Muller et al. (2023) characterized the supply chains as ad hoc networks designed to meet particular, urgent, and time-sensitive needs in the context of COVID-19. Through the use of a special sampling, they investigated how businesses achieve supply chain agility when creating ad hoc supplier chains. They proposed dynamic capabilities in their emergent theoretical model, which an entrepreneurial orientation moderates, in response to a specific need.

3. ASSUMPTIONS AND NOTATIONS

3.1. Assumptions

The following assumptions are used in the proposed method:

In this paper, we analyze two problems:

(i) Analysis of the number of victims requiring immediate treatment when a virus-spreading situation occurs, and

(ii) Analysis of medication capacity and time of waiting for victims in desired hospitals within the "golden hour" during virus outbreaks.

3.2. Notations

The following notations are used in the proposed method:

(i)	λ	=	Rate of patient Arrivals
(ii)	μ	=	Service rate per server (doctor)
(iii)	S	=	Parallel servers (doctors)
(iv)	C	=	System limit of patients (hospital capacity)
(v)	N(t)	=	Patients presence in the system at t time
(vi)	$P_{ij}(\Delta t)$	=	Probability of j patients in time Δt when i patients are already present in the system at the time t
(vii)	q_j	=	"Steady-state probability of j patients when patients enter into system
(viii)	π_j	=	Probability with queue size j at any moment [$\lim t\to+\infty\ P_{ij}(t) = \pi_j$]
(ix)	L_q	=	Average waiting queue length when the system in a steady state
(x)	L	=	Average queue length when the system in a steady state
(xi)	W(t)	=	Distribution function of victims staying time when the system is in a steady state
(xii)	W_q	=	Average waiting time of patients when the system is in a steady state
(xiii)	W	=	Patients staying time when the system is in steady state
(xiv)	W_s	=	Average service time per patient when the system is in a steady state
(xv)	$W_q(t)$	=	Distribution function of the waiting time in the system.
(xvi)	ρ_s	=	Intensity of patient flow of S servers
(xvii)	π_s	=	Probability with queue size S at any time [$\pi_s = \lim t\to\infty\ P\{N(t)\} = S$]

4. THE PROPOSED WORK

4.1. Calculation of Number of Arrived Patients in Transient States System

The current pandemic situation is characterized by low probability and high uncertainty. Therefore, we cannot predict the flow of people to hospitals during a virus outbreak using traditional queuing systems. Panic can drastically increase hospitalizations, complicating the determination of treatment capacity. It acknowledges the probability distribution of the transient state. As a result, the characteristics of the queuing system constantly change over time, placing it in a non-steady state. To accurately estimate the number of victims arriving at the hospital at any given time, we need a 'transient acts system (TAS).' It considers factors like the number of infected people and their proximity to the hospital to change the queuing system parameters over time. By estimation of victims arriving at the hospital, we can determine whether the number of victims exceeds the hospital's capacity to treat them within the "golden hour."

Assume that victims arrive at the closest hospital at a specific time. The arrival time interval, which follows an exponential distribution with a rate of arrival of λ(t). At time t, s(t) parallel servers are available with a service rate of μ(t), and the service time is exponential. The hospital can accommodate a maximum of N victims.

To determine transient probabilities, we choose the inter-arrival time such that at most one procedure (an arrival or treatment completion) can happen in the interval of time Δ(t). If k victims are now existing in the system at t time, then:

(i) The probability of one arrival in the time interval Δt is λ(t)·Δt.
(ii) The probability of more than 1 arrival in the interval of time Δt is o(Δt).
(iii) The arrivals in various time periods are independent.
(iv) The probability of one treatment completion in the interval of time Δt is defined by min(s(t), k)·μ(t)Δt.
(v) The probability of more than one treatment completion in the time interval Δt is o(Δt).

If the patient's arrival follows the first three assumptions, it is considered a non-homogeneous Poisson procedure. The Δt value could be made small enough to decrease the error in the final results.
Now let $P_i(t)$ represent the probability of the i victims present in the system at 't' time.
Therefore, $P_0(0) = 1$ *and* $P_0(i) = 0$, *if* $i > 0$.
Then, using the $P_i(t)$ knowledge, we can calculate the probability $P_i(t+\Delta t)$ as given below:

$$P_0(t+\Delta t) = [1-\lambda(t)\Delta t]P_0(t) + \mu(t)\Delta t P_i(t) \tag{1}$$

$$P_i(t+\Delta t) = \lambda(t)\Delta t P_{i-1}(t) + [1-\lambda(t)\Delta t - min(s,\, i)\mu(t)\Delta t]P_i(t) + min(s,\, i+1)\,\mu(t)\Delta t P_{i+1}(t)\,;\; 1 \le i \le N-1 \tag{2}$$

$$P_N(t+\Delta t) = \lambda(t)\Delta t P_{N-1}(t) + [1-min(s,\, N)\mu(t)]\Delta t P_N(t) \tag{3}$$

From this relation, we can estimate the probability of COVID-19 patients arriving or receiving care at 't+Δt' time. Then the patient's average number

$$E(t) = \sum_{i=1}^{n} \lambda P_i(t) \tag{4}$$

With this equation, we can calculate the average number of casualties waiting and receiving care at the same time.

4.2. Priority Level in Emergency Department

Hospital specialists collaborate with emergency departments to diagnose, monitor, and either admit patients to the hospital or discharge them if their situation does not require inpatient care. The first interaction between patients and staff takes place in the emergency care area, where the emergency staff examines the patient's status, decides its severity, and assigns the appropriate triage level or code from A to E as shown in Table 1. The emergency staff immediately directs patients with Level A to the intensive care room and sends those with Level B to the immediate care ward. Patients with Levels C, D, and E proceed to the waiting room unless a bed becomes available right away.

Table 1: Level of triage (or codes of triage) in emergency department.

Level / Code of Triage	Time of waiting in triage area
Level A / Red	0 min.
Level B / Pink	Max. 15 min.
Level C / Yellow	Max. 30 min.
Level D / Green	Max. 60 min.
Level E / White	Max. 120 min.

One can describe the emergency department's working system as an absolute queueing process. Red code patients are handled on a priority basis, and patients are assessed and treated and then discharged or transferred to a ward. Based on the number of infected victims, queuing theory is utilized to approximate the number of human resources needed for the ED department and to determine the average waiting time. These results are crucial for emergency department management to make informed decisions and to best organize the emergency care system.

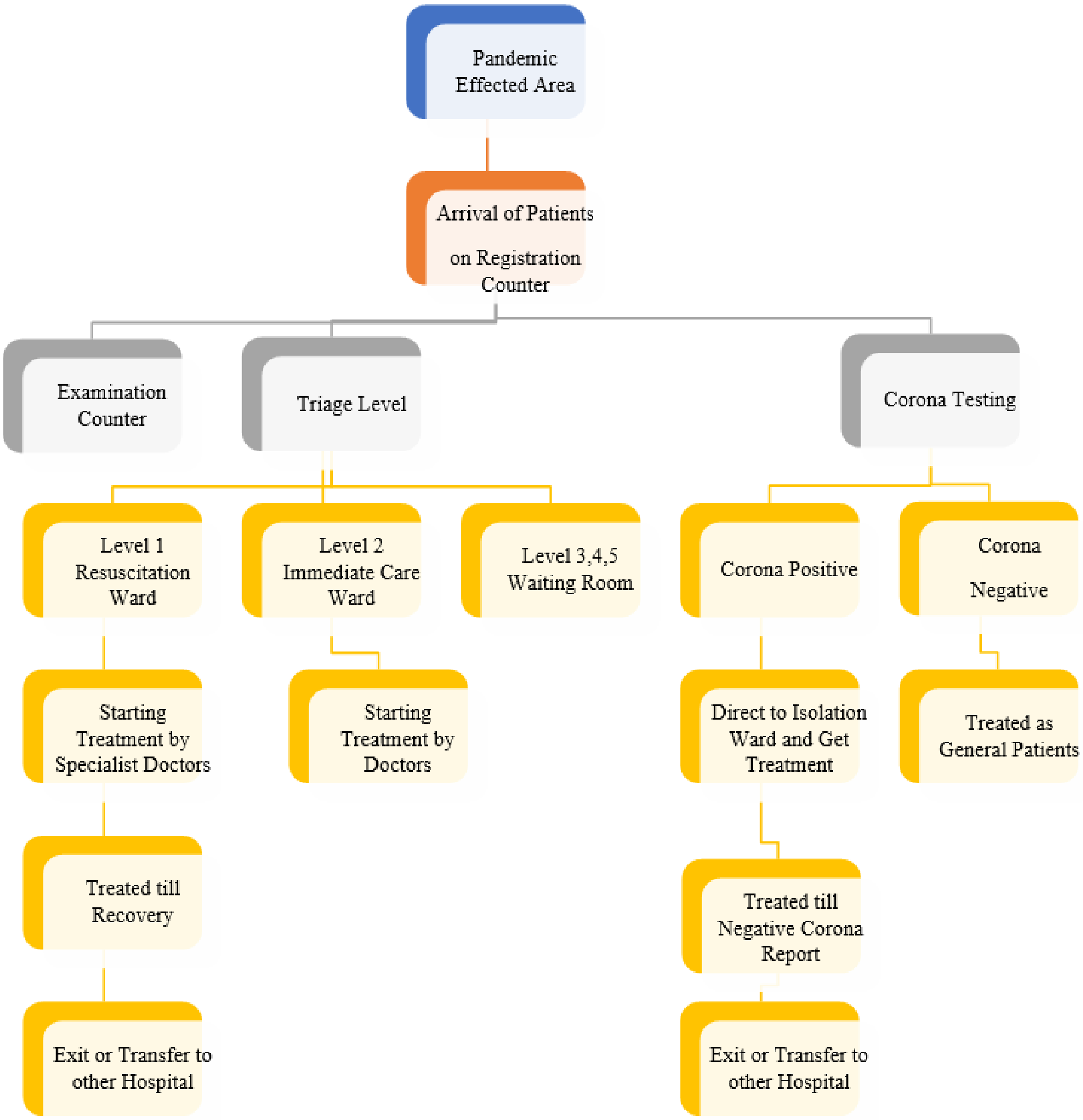

Figure 1: Flow Chart of Hospital Treatment Process.

4.3. Capability Analysis of a Hospital System by Queueing Model

This section analyzes the therapeutic potential of hospitals based on the relationship among λ and μ. It can be divided into two cases:

Case (I). The hospital treatment capacity for $\lambda \leq \mu$.

Case (II). The hospital treatment capacity for $\lambda > \mu$.

Case I. The Hospital treatment capacity for $\lambda \leq \mu$

For $\lambda \leq \mu$, the hospital's treatment capacity can be determined using the queuing mechanism $\{M/M/s/GD/\infty/\infty\}$. In this mechanism, patient arrival and service time are monitored by exponential distribution, and each patient's service time is independent. Parallel servers (involving doctors as well as healthcare personnel) are 's'. The system has an unlimited capacity, and patients are unlimited.

Since $\lambda \leq \mu$, the system is at steady state. This allows us to determine the average time of waiting (W_q).

$$W_q(t) = \frac{\rho_s}{\lambda(1-\rho_s)^2}\pi_s \tag{5}$$

The average time spent by the patients in the hospital is

$$W = W_q + \frac{1}{\mu} \tag{6}$$

In this case, all the patients can be treated in the critical time period.

Case II. The Hospital treatment capacity for λ>μ

If $\lambda > \mu$, the hospital's treatment capacity is insufficient, then to avoid explosive situation, we limit the storage capacity by c in "Case I model". Then, no more patients can be entered into the system, and the system remains in a steady state.

In this situation, the $\{M/M/s/GD/c/\infty\}$ queuing system can be used to calculate the hospital's treatment capacity. In this system, both patient arrivals and service time are monitored by an exponential distribution, and each patient's service time is independent. parallel servers (involving doctors as well as healthcare personnel) are 's', the system's limit is 'c', and patients are unlimited.

The analysis of the system characteristics is given below:

The system admits the service discipline 'first come, first serve', and q_j represents the steady-state j patient's probability of being present in the system.

Then $$q_j = \frac{\pi_j}{1-\pi_c},\ j = 1,\ 2,\ \ldots,\ c-1 \tag{7}$$

The waiting time of patients ruled by *j−c+1-* Erlang distribution. Then, the distribution function of the waiting time can be calculated by

$$W_q(t) = P\{W_q \le t\} = \begin{cases} \sum_{j=0}^{s-1} q_j & t = 0 \\ \sum_{j=0}^{s-1} q_j + \sum_{j=s}^{c-1} q_j \int_0^t \frac{s\mu(s\mu x)^{(j-s)}}{\lfloor j-s} e^{-s\mu x}, & t > 0 \end{cases} \tag{8}$$

Then the patient's average time of waiting (W_q) can be determined using

$$W_q(t) = \sum_{j=0}^{c-1} \frac{(j-s+1)}{s\mu} q_j \tag{9}$$

The average time spent by the patients in the hospital can be determined using

$$W = W_q + \frac{1}{\mu} \tag{10}$$

Since not all patients can receive treatment during the critical hours, we attempt to utilize the waiting time distribution to identify those who can. We can then use the results in these cases to ascertain whether the patients receive treatment during the critical hours.

Therefore, we can deduce (11) from (8), which indicates the probability that the waiting time is less than one hour.

$$W_q(t) = P\{W_q \le t\}$$
$$= \sum_{j=0}^{s-1} \frac{\pi_j}{1-\pi_c} + \frac{\pi_0}{(1-\pi_c)\lfloor s} \begin{bmatrix} \rho^s\left(1-e^{-s\mu t}\right) + \frac{\rho^{s+1}}{s}\left(1-e^{-s\mu t} - s\mu t e^{-s\mu t}\right) + \\ \frac{\rho^{s+2}}{s^2}\left(1-e^{-s\mu t} - s\mu t e^{-s\mu t} - \frac{1}{\lfloor 2}(s\mu t)^2 e^{-s\mu t}\right) + \ldots \\ + \frac{\rho^{c-1}}{s^{c-s-1}}\left(1-e^{-s\mu t} - s\mu t e^{-s\mu t} - \ldots - \frac{1}{\lfloor c-s-1}(s\mu t)^{c-s-1} e^{-s\mu t}\right) \end{bmatrix} \tag{11}$$

The expansion of (5) to (11) can be found in the book by Wayne and Winston (2006).

5. NUMERICAL ANALYSIS OF PROPOSED WORK

In this segment, we will be validating the described method using simulation techniques to calculate the arrival rate and hospital treatment capacity.

5.1. Simulation of Arrival Rate

We can determine the number of patients who reached the nearest hospital on time using equation (4). Let's assume that a corona pandemic has occurred in Jaipur, and the nearest hospital is SMS Hospital, Jaipur. External factors such as crowd density, traffic, and road conditions can determine the value of victims arriving at the hospital at various times. Table 2 presents these empirical values, modified from equation (5).

Table 2: The hospital arrival rate of victims λ at various times.

Internal Time (Min)	Arrived rate λ (Patient/Hr)
0—5	0
5—10	720
10—20	900
20—25	600
25—30	180
30—60	0

We can calculate the probability distribution of the arriving patients at various times using equations (1), (2), and (3). Without going into detail, we provide a preview of how to compute this distribution. Equation (4) then allows us to determine the average number of waiting victims and those receiving treatment within one hour. Table 3 presents the partial results.

Table 3: Probability distribution of the arrived victims at various times.

Increasing time		The probability of the arrived victims at various times: Prob(person)					
"Sec	Hour	Prob (0)	Prob (1)	Prob (2)	Prob (3)	Prob (250)	Prob (251)
0	0	1	0	0	0	0	0
10	0.00	1	0	0	0	0	0
300	0.08	1	0	0	0	0	0
600	0.17	9.64E-18	5.5E-17	2.53E-16	1.04E-15	0	0
900	0.25	1.78E-39	1.17E-38	6.26E-38	3E-37	1.73E-44	3.44E-45
1200	0.33	2.95E-60	1.78E-59	9.24E-59	4.42E-58	3.47E-07	5.67E-07
1500	0.42	9.77E-71	5.3E-70	2.42E-69	1.03E-68	0.035609	0.180140
1800	0.50	9.48E-70	3.21E-69	1.12E-68	3.93E-68	0.129262	0.350354
2100	0.58	1.26E-64	2.83E-64	9.14E-64	2.93E-63	0.012013	0.002168
2400	0.67	3.74E-60	7.55E-60	2.27E-59	6.77E-59	0.000144	1.34E-05
2700	0.75	4.27E-56	7.86E-56	2.22E-55	6.22E-55	1.32E-06	8.3E-08
3000	0.83	2.23E-52	3.76E-52	1E-51	2.66E-51	1.08E-08	5.14E-10
3300	0.92	5.95E-49	9.25E-49	2.35E-48	5.94E-48	8.33E-11	3.18E-12
3600	1	8.88E-46	1.28E-45	3.11E-45	7.51E-45	6.17E-13	1.97E-14"

Conditions:
1. Arrivals admit a non-uniform Poisson process;
2. The hospital has 10 doctors and can treat 60 patients per hour;
3. The hospital can serve up to 250 patients;
4. The first patient, taking 5 minutes, arrived at the nearest hospital,

5.2. Graphical Analysis

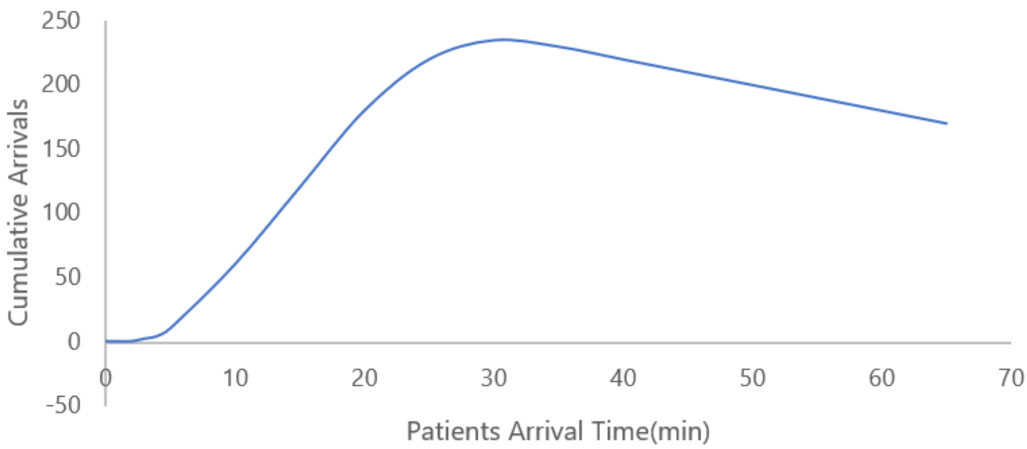

Figure 2: Average number of patients awaiting and in treatment.

The graph in Figure 2 indicates that the number of victim arrivals at the hospital has the fastest growth after the first victim arrived in the fifth minute. Within the first 30 minutes, around 250 victims had arrived at the hospital. Simultaneously, rescue efforts led to a decrease in the number of victims. These simulation outcomes have been consistent with our experience. We could leverage this approach to find the victim's number at any given time in a hospital or on the streets, providing valuable information for decision-making support systems.

5.3. Simulation of Hospital Treatment Capacity

An estimated, approximately 300 victims are currently awaiting treatment at the hospital. The exponential distribution of service time allows each doctor to treat 6 patients per hour, but the time it takes to treat each victim can vary. The waiting area can accommodate and hold a specific number of patients, with a patient traffic density limit of 50. The "golden hour" refers to the crucial 60-minute window during which patients should receive treatment. Table 4 displays the number of patients treated, the average waiting time, and the likelihood of receiving treatment during the "golden hours."

Table 4: Average time of waiting and probability of patients.

Availability of doctors	Treated patients	Average time of waiting (Hours)	The treatment probability in 60 minutes	Availability of doctors	Treated patients	Average time of waiting (Hours)	The treatment probability in 60 minutes
	50	0.66	0.99		110	0.74	1.00
	53	0.71	0.99		115	0.79	0.99
	55	0.75	0.98		120	0.83	0.88
10	58	0.80	0.95	20	125	0.87	0.93
	60	0.83	0.92		130	0.91	0.85
	65	0.91	0.77		135	0.95	0.71
	70	1.00	0.53		140	0.99	0.54

Table 4 reveals that when the treatment probability during the "golden hours" exceeds 0.95, 10 doctors can treat 58 patients, while 20 doctors can treat 125 patients. However, if one hundred victims are present in the system, then after the fifty-ninth victim waiting for the 10 doctors, a victim will have two options:

(i) Referrals to other hospitals for immediate treatment are necessary. The treatment capacity of the referred hospital can also be leveraged in this way.

(ii) We need more doctors (including paramedical staff) to serve the victims.

For example, if the initial hospital employs only 10 doctors, and an additional 10 doctors join the team, it can treat 125 victims effectively. Theoretically, based on the calculation in (2), we can recruit a sufficient number of doctors from other hospitals to treat the necessary victims. We also need to consider the limitations in hospitals, such as emergency rooms and required equipment. In such cases, transferring the casualty to another medical facility would be the best option.

6. CONCLUSIONS

In this article, we analyzed the rate at which patients arrived at a preferred hospital and the hospital's treatment capacity during the coronavirus pandemic. We have developed a "Transient Act System (TAS)" using queuing theory techniques to address emergency conditions. This system helps to analyze the number of victims requiring immediate treatment within the "golden hour" during virus outbreaks. The patient's prioritization is crucial for emergency department management to make informed decisions and to best organize the emergency care system. We created a flow chart to simplify the process of patient flow to hospitals during virus outbreaks.

We also described a method to adapt the TAS system to a transient non-steady state. We calculated the number of patients arriving at the hospital during a given time period and their waiting times using the relationship between the arrival rate and the service rate. Specifically, we examined the calculation of hospital treatment facilities during the periods when the arrival rate was higher than the service rate; that situation was closely relevant to the overcrowding in hospitals during virus outbreaks. The method for calculating hospital capacity is based on the waiting time probability distribution function, where the patient waiting times follow a j-c+1-Erlang distribution. To validate the proposed model, we performed numerical analysis and simulation techniques using modified empirical values. The presented graph indicates that the number of victim arrivals at the hospital has the fastest growth after the first victim arrived in the fifth minute. Within the first 30 minutes, around 250 victims had arrived at the hospital. Simultaneously, rescue efforts led to a decrease in the number of victims.

We could leverage this approach to find the victim's number at any given time in a hospital or on the streets, providing valuable information for decision-making support systems. This research work examines the actual operation of the hospital system's capacity and suggests modifications to the system to decrease patient waiting times. We can use the database to enhance the decision support system in emergency management. This research serves as an example that justifies the feasibility.

6.1. Future Research Directions

Unnecessary hospital wait times can worsen a patient's health complications and, in some cases, even result in death. There are various possible ways to avoid unnecessary wait times. We can generate a system for pre-registering patients using ICT and have staff members serve them directly, which can reduce patient wait times. To reduce the initial service time before doctors, treat patients, we can deploy more paramedical staff. This will reduce the overall service time for doctors to treat patients, which will present an immediate improvement in services. We can also deploy more doctors and paramedical officers in emergency patient flow situations. The future research will include features such as the functions of λ and μ, reverse queueing systems, systems with multiple queues and multiple servers, and even multiple hospitals, etc. Additionally, we can simulate the arrival rate of victims based on their location, arrival time, environment, and other incident circumstances to more accurately reflect the real situation.

7. IMPORTANCE OF THE STUDY

- Improving Hospital Systems Capabilities: The importance of this study in assisting managers and organizational healthcare officials in understanding areas affected by COVID-19 in order to plan for hospital architecture and the ability to handle future pandemics is one of its possible consequences.
- Optimizing Resource Allocation: The queueing system included in the research means that the kind of resource constraints that most hospitals have during a pandemic will not be as noticeable, instead improving the standard of treatment that patients receive.
- Enhancing Pandemic Preparedness: All things considered, the data gathered throughout the study contributes to improving the suggestions for dealing with upcoming pandemics in medical facilities and lessening their consequences.
- Informing Healthcare Policy: This study provides specific recommendations on how healthcare policy might be improved further in order to prepare for the next pandemic.

8. LIMITATIONS OF THE STUDY

- Model Assumptions: The queuing hypothesis employed in this study might not be solely based on some factors, such as the rate of arrivals, particularly the amount of time spent caring for these patients, and other factors that might actually differ.
- Data Limitations: This data classification could introduce bias into the study by affecting how well healthcare systems are able to quantify the impact of COVID-19.
- Generalizability: These findings might not apply to other hospital systems worldwide or to subsequent pandemics because the current research focuses on the experience of a particular pandemic—COVID-19—as well as the hospital system architecture.
- Simplicity of Complex Systems: The queueing theoretical model used in this case study makes some assumptions about the real world, such as not accounting for all of the feedback present in a hospital system.

Availability of Data and Material:

This study uses qualitative research techniques to analyze the capacity of hospitals to care during the COVID-19 pandemic. For this study, we obtain data from secondary sources, specifically government reports, magazines, newspapers, and online sources like WHO Reports [5], the NITI AYOG website [6], etc. The numerical data are empirical and have been modified from equation (5). Numerical analysis of the data will be used in this study depending on the purpose of the study.

Competing Interest:

The authors have no competing interests to declare that are relevant to the content of this article.

Authors' Contributions: All authors have made substantial contributions to the conception or design of the work. Material preparation, data collection, and writing work were performed by Balveer Saini [first author]. The supervision and analysis of the work were performed by Dr. Dharamender Singh and Dr. Kailash Chand Sharma. All authors read and approved the version to be published.

Acknowledgement:

We would like to thank the M. S. J. Govt. P.G. College, Bharatpur, Rajasthan, India, for providing the facilities for this research and their valuable support.

REFERENCES

Aalto, S. (2024). Whittle index approach to multi-server scheduling with impatient customers and DHR service times. *Queueing Systems* 107: 1–30. https://doi.org/10.1007/s11134-024-09902-5.

Agrawal, R., & Singh, B.K. (2018). An analytical study of queues in medical sector. *OPSEARCH* 55(2): 268-287. DOI: 10.1007/s12597-017-0324-7.

Aljazeera (2020). How the new Corona virus spread. https://www.aljazeera.com/news/2020/01/timeline-china-Coronavirus-spread-200126061554884.htm.

Baker, S. (2020). Everything we know about the mysterious, deadly Wuhan virus sweeping across China. https://www.pulse.com.gh/bi/tech/everything-we-know-about the-mysterious-deadly-wuhan-virus-sweeping-across-china/zd3sgq4.

Caunhye, A. M., Nie, X., & Pokharel, S. (2012). Optimization models in emergency logistics: A literature review. *Socio-economic Planning Science* 46(1): 4–13. https://doi.org/10.1016/j.seps.2011.04.004.

Chen, N., Zhou, M., Dong, X., Qu, J. et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel Corona virus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 395(10223): 507—513. https://doi.org/10.1016/S0140-6736(20)30211-7.

Government of India (2020). #IndiaFightsCorona COVID-19. Available from: https://www. mygov.in/COVID-19.

Hillas, L. A., Caldentey, R., & Gupta, V. (2024). Heavy traffic analysis of multi-class bipartite queueing systems under FCFS. *Queueing Systems* 106: 239–284. https://doi.org/10.1007/s11134-024-09903-4.

Kilic, A., Dincer, M.C., & Gokce, M.A. (2014). Determining optimal treatment rate after a disaster. *Journal of the Operational Research Society* 65(7): 1–15. https://doi.org/10.1057/jors.2013.52.

Li, W., & Goldenshluger, A. (2024). Adaptive minimax estimation of service time distribution in the $M_t/G/\infty$ queue from departure data. *Queueing Systems* 108: 81–123. https://doi.org/10.1007/s11134-024-09921-2.

Mittal, H., & Sharma, N. (2020a). Assessment of Delhi NCR Traffic Using Queuing Model. *International Journal of Advanced Science and Technology* 29(7): 5418-5424.

Mittal, H., & Sharma, N. (2020b). A Probabilistic Model for the Assessment of Queuing Time of Coronavirus Disease (COVID-19) Patients using Queuing Model. *International Journal of Advanced Research in Engineering and Technology* 11(8): 22-31.

Mittal, H., & Sharma, N. (2022). A simulation-based approach for minimizing waiting time in AIIMS, Delhi using Queuing model. *International Journal of Health Sciences* 6(S5): 7037–7054.

Muller, J., Hoberg, K., & Fransoo, J.C. (2023). Realizing supply chain agility under time pressure: Ad hoc supply chains during the COVID-19 pandemic. *Journal of Operations Management* 69: 426–449. https://doi.org/10.1002/joom.1210.

Pradhan, S., Gupta, U.C., & Samanta, S.K. (2016). Queue-length distribution of a batch service queue with random capacity and batch size dependent service: $M/GY_r/1$. *OPSEARCH* 53: 329-343. https://doi.org/10.1007/s12597-015-0231-8.

Qiu, W., Rutherford, S. et al. (2017). The pandemic and its impacts. Health, culture and society 9:1-11. https://doi.org/10.5195/HCS.2017.221.

Saini, B., Singh, D., & Sharma, K.C. (2024). Analysis of Right to Health Using Queueing Theory: A Critical Overview. *IRJAEM* 2(3): 332-338. https://doi.org/10.47392/IRJAEM.2024.0048.

Vassa, H., & Szabo, Z.K. (2015). Application of Queuing Model to Patient Flow in Emergency Department. *ELSEVIER- Procedia Economics and Finance* 32: 479-487. https://doi.org/10.1016/S2212-5671(15)01421-5.

Wang, J., Yu, H., Luo, J., & Sui, J. (2008). Medical treatment capability analysis using queuing theory in a biochemical terrorist attack. *ISORA*: 415–424.

Warren, K. (2020). Countries that are on lockdown because of the Corona virus. https://www.pulse.ng/bi/lifestyle/countries-that-are-on-lockdown-because-of-Coronavirus/f28yp94.

Wayne L. Winston (2006). Queueing Theory. In 4th Ed., *Operations research: Introduction to probability models* (pp. 1051-1131). Tsinghua University Press.

World Health Organisation (2021). Virtual press conference on COVID-21 June 2021, *World Health Organisation.*

Yaduvanshi, D., Sharma, A., & More, P.V. (2019). Application of Queuing Theory to Optimize Waiting Time in Hospital Operations. *OPERATIONS AND SUPPLY CHAIN MANAGEMENT* 12(3):165 – 174. http://doi.org/10.31387/oscm0380240.