

# Stochastic Analysis Of Queue Dynamics In Multi-Stage Mass Covid-19 Vaccination Networks With Capacity Optimization

Kanika Sharma<sup>1</sup>, B.K. Singh<sup>2</sup>, Shubham Agarwal<sup>3\*</sup>

<sup>1,2</sup>*Department of Mathematics, IFTM University, Moradabad-244102, Uttar Pradesh, India*

<sup>3</sup>*\*Department of Mathematics, NDIM, New Delhi-110062, Delhi, India*

*Received: 5th Jan, 2026; Revised: 25th Jan 2026; Accepted: 20th Feb, 2026; Available Online: 10th Mar, 2026*

## ABSTRACT

Multi-stage service systems were needed to roll out in the shortest time possible during mass COVID-19 vaccinations in the presence of severe stochastic variations in the arrivals and the service times. In contrast to one-station clinics, vaccination facilities are complex intertwined queueing systems which comprise registration, clinical screening, vaccination, and post-inoculation observation phases. The paper builds a strict dependence on stochastic study of such multi-stage network of vaccinations on the global level. It is a non-stationary open Jackson network with time-varying Poisson arrivals and non-homogeneous service rates. We develop the conditions of global stability through the techniques of fluid limits, obtain the approximations of diffusion with heavy traffic, and study the dynamics of bottlenecks. An optimization problem is developed that involves nonlinear capacity allocation in order to establish optimal staffing at the congested stages subject to labor costs. Original discrete-event simulation experiments help support the theoretical results. The findings can be used in large-scale planning of vaccination and add to the mathematical theory of non-stationary healthcare queueing networks.

**Keywords:** Mass vaccination; Queuing networks; Stochastic stability; Heavy-traffic diffusion; Healthcare operations; Capacity allocation; Pandemic logistics; Optimization

**How to cite this article:** Sharma K, Singh BK, Agarwal S; Stochastic Analysis Of Queue Dynamics In Multi-Stage Mass Covid-19 Vaccination Networks With Capacity Optimization. *Int J Drug Deliv Technol.* 2026; (1): 693-700; DOI: 10.25258/ijddt.16.1.72

**Source of support:** Nil.

**Conflict of interest:** None

## INTRODUCTION

The COVID-19 pandemic has revealed how the scales of the healthcare challenges throughout the world were unprecedentedly vulnerable, specifically the ability to handle massive vaccination campaigns. Although epidemiologically important to implement the herd immunity and prevent transmission, mass immunization efforts represent a huge burden on the healthcare infrastructure. The concept of long waiting lines, congestion at vaccination points, low usage of resources, and spontaneous changes in demand have become the paramount challenges to an effective vaccine delivery. These problems highlight the necessity of strict mathematical modeling systems that can reflect the dynamics and stochasticity of vaccination service systems (Green, 2020)<sup>1</sup>. COVID-19 vaccination programs conducted by masses are multi-stage systems of service by their nature. The common vaccine center has several steps like registration, document verification, medical screening, vaccination, and post-vaccination monitoring. The stages have in use a limited service capacity and stochastic arrivals, service-time variation, and potential disruptions. As a result, the vaccination centers may be simulated as multi-server, multi-phase queueing networks of interdependent nodes of service. The classical theory of queueing offers the basic instruments to discuss the mentioned systems, e.g., M/M/s and M/G/s models,

Jackson networks, and open queueing networks (Gross et al., 2018; Kleinrock, 1975)<sup>2, 3</sup>.

Nevertheless, the pandemic situation adds more complexity such as time-dependent arrival rates, correlated service processes, arrival in batches, priority classes, and capacity adjustment policies. In the recent years, the significance of stochastic modeling in the planning of response to pandemics has been pointed out. Kaplan and Forman (2020) highlighted that operational research is crucial in the process of distributing vaccines<sup>4</sup>, and Bertsimas et al. (2021) also proved that data-driven optimization models<sup>5</sup> are effective in the decision-making process of the work of the public health in the context of the COVID-19 pandemic. On the same note, in case of uncertainty, queueing-based analyses have been used to assess healthcare congestion, distributions of waiting time, and resource allocation strategies (Green, 2020; Armony et al., 2021)<sup>1, 6</sup>. In spite of these contributions, not much has been done to analyze global stochastic stability of multi-stage vaccination networks with endogenous capacity optimization. An international analysis paradigm is essential to interpret the performance of the entire system instead of the behavior of individual nodes. Multi-stage networks result in bottlenecks and systemic delays caused by congestion propagating downwards in one stage. The condition of stability, the constraint of throughput and the constraint of traffic intensity should be analyzed as a combination. The stability of open Jackson-type networks<sup>7</sup> demands that the effective

\*Author for Correspondence: meshubhamagarwal@gmail.com

arrival rate of each node should be lower than the service capacity of that node (Jackson, 1957).

Nevertheless, in dynamically routed, capacity reconfigured vaccination networks, classical findings will have to be generalized to include time-dependent inputs and stochastic uncertainty. Furthermore, the design of the vaccination network includes optimization of capacity, which is a key factor. The problem under consideration is to decide how many servers (healthcare workers), service counters, and observation units should be installed to reduce the estimated waiting time and the cost of operation and meet the service-level requirements. This leads to the stochastic capacity decision problems, which are usually represented as nonlinear or mixed-integer optimization problems embedded into queuing systems<sup>8, 9</sup> (Bertsimas & de Boer, 2005; Cachon and Terwiesch, 2019). In resource limited environments (like low- and middle-income countries) trade-offs between the cost of staffing and delay costs are especially significant as the high-demand of vaccinations could overwhelm infrastructure. The other important dimension is the stochastic variation in the arrivals and service times. In the case of vaccination campaigns, arrival patterns are determined by the systems of appointment booking, walk-in, announcement of policies, and the availability of vaccination. Empirical studies indicate that the arrival processes do not necessarily conform to some basic Poisson assumption but can be time-dependent or bursty (Green, 2020)<sup>1</sup>. Services can also be distributed in a general manner because of the heterogeneity of patients and complexity of the procedures. Thus, a stochastic model that allows the use of arbitrary inter-arrival and service-time distributions (e.g. M/G/s or G/G/s approximations) is crucial to realistic performance analysis (Whitt, 1993)<sup>10</sup>.

This paper aims to synthesize a world-wide and stochastic model of dynamics of queues in multi-stage mass COVID-19 vaccination systems with local capacity optimization. The research unites the multi-node queuing network models, stability and steady-state analysis in the conditions of stochastic assumptions and service capacity optimization under the conditions of cost and performance. The suggested structure is expected to offer theoretical understanding as well as operational decision assisting solutions in large-scale planning of immunization. Precisely the contributions of this work are threefold:

**Global Stochastic Network Modeling:** We model vaccination system as open multi-stage queuing network with stochastic arrivals and generic service distribution, which models the inter-stage dependencies and flow conservation.

**Stability and Performance Analysis:** We obtain the global stability conditions, steady-state performance indicators (waiting time to be expected, length of the queue, utilization), and describe the effects of the intensity of the traffic on the systemic congestion.

**Capacity Optimization Framework:** We introduce the model of a capacity decision into the stochastic network, optimizing on the total operation costs and delay costs, maintaining required service-level goals.

The study brings together queuing network theory and stochastic optimization and also fits in the body of

knowledge that is ever-increasing in the field of pandemic operations management and healthcare systems engineering. The findings offer practical ideas to policy makers, hospital managers, and community health strategic planners to create resilient and effective vaccination systems in case of uncertainty.

## LITERATURE REVIEW

The theory of queueing has been used as a basic analytical instrument in the analysis of healthcare operations management. The works by Green (2006)<sup>11</sup> and Hall (2013)<sup>12</sup> can be discussed as the seminal cases of the application of the stochastic service models to the assessment of patient flow, staffing needs, congestion management, and efficiency of the system in hospitals and outpatient facilities. The expected waiting time, the utilization of the server, the distribution of the length of queues, and the possibilities of the service levels are performance measures in these frameworks that are essential in evidence-based decision-making in the healthcare delivery systems. The use of queueing models has grown greatly during the COVID-19 pandemic, especially when applying to the mass vaccination campaign. The scale and urgency of the vaccination programs were unprecedented, which necessitated tight modeling of throughput capacity, scheduling future appointments, and site design.

In recent literature there is a general tendency with respect to the stochastic modeling methods to represent the variability in terms of arrival rates, service time and operational disruption. An example is Di Pumpo et al. (2022), who applied the concept of queueing to the layout of vaccination sites and found that the design of the queue can be structured in such a way that it is safer, has limited crowding, and is more efficient in its functions<sup>13</sup>. On the same note, Franco et al. (2022) created queueing network models<sup>14</sup> to model vaccination throughput, where they include multi-stage service processes and capacity planning decisions to help optimize resource allocation. Theoretically, Ward (2022)<sup>15</sup> and Dai and He (2022)<sup>16</sup> developed the knowledge of service systems at heavy traffic by using diffusion and fluid approximations. Their work presents an asymptotic characterization of queue dynamics, and in this way, it allows the performance analysis to be tractable when the traditional Markovian models are too analytically intractable. These estimations are especially useful in large-scale vaccination programs when there is a surge of demands. Using this point of view, Armony and Ward (2023) investigated capacity flexibility and scaling strategies<sup>17</sup> with changing demand conditions, providing an insight into how service systems can change dynamically in times of peak.

Sharma et al. (2024) used the queueing theory to optimize vaccination in the Indian context in which they modeled the delivery of booster doses and assessed the staffing configuration to reduce waiting time and congestion<sup>18</sup>. Extensively, several publications released in 2021 to 2025 have been dedicated to the work on managing the pandemic, stochastic modeling of healthcare, and capacity allocation measures. All of these works have been used to help carry

out simulation-based analyses, discrete-event modeling, and operational experimentation to conduct logistics analysis of vaccination. Nevertheless, with this accumulating literature, comparatively few studies combine three important elements in a single analysis framework, including rigorous global stability requirements that assure long-run system viability, diffusion-scale analysis that includes surge behavior in multi-stage stochastic

vaccination networks, and embedded capacity allocation optimization, formulated in the queueing model itself. The little incorporation of these factors inspires the current study that aims at filling the gaps between the theoretical stochastic analysis and the practical optimization of large scale vaccination system.

**RECENT STUDIES**

Study	Year	Methods / Modeling Approach	Key Findings / Focus
Di Pumpo et al. <sup>13</sup>	2022	Queueing theory + simulation of station-level queues at a COVID-19 vaccination site	Suggests a queueing theory to enhance safety and efficiency at mass vaccination sites; simulation demonstrates that the queue-theory-based scheduling significantly decreases overcrowding relative to simple linear methods.
Franco et al. <sup>14</sup>	2022	Queueing network modeling with Jackson networks + non-linear optimization	Applications are the use of queueing networks to plan capacities and patient scheduling in COVID-19 vaccination centers in order to balance throughput and waiting times.
Ward <sup>15</sup>	2022	Fluid approximations for many-server queues with time-varying arrivals	Creates deterministic fluid models, approximating the dynamics of large-scale service system under heavy load, which make it possible to estimate performance when discrete stochastic models are intractable.
Dai & He <sup>16</sup>	2022	Diffusion approximations for service networks	It is a survey or develops diffusion approximations of many-server queues that reflect variance around fluid limits, which are important to performance analysis when in heavy traffic.
Armony & Ward <sup>17</sup>	2023	Capacity scaling and heavy-traffic analysis under surges	Researches on how to increase and decrease the capacity of services based on demand arousal by heavy traffic congestion limits and control measures to direct responsive staffing measures.
Sharma et al. <sup>18</sup>	2024	Queueing theory models applied to vaccination and booster delivery processes	Uses queueing networks to simulate and model booster delivery operations in order to predict staffing requirements to eliminate bottlenecks, and approximate throughput with a capacity-constrained system.
Matrajt et al. <sup>19</sup>	2021	Vaccination logistics modeling (likely epidemiological/supply chain optimization)	Concentrates on the logistical components of the vaccination campaigns; incorporates demand, distribution, and operational planning.
Xie et al. <sup>20</sup>	2024	Integration of queueing theory within a spatial optimization algorithm	Comes up with a spatial site placement algorithm that considers the expected queueing so as to maximize coverage and minimise loss through long waits. The empirical evidence has shown that optimized layouts minimize the attrition rates, and are better at coverage as compared to regular placements.
Massey & Whitt <sup>21</sup>	1998	Strong approximations for queueing networks (stochastic process limits)	Gives theoretical basis on how to approximate Markovian service networks, based on Brownian processes or diffusion processes, which are used to analyse time varying queueing systems and heavy traffic queueing systems.
Jennings et al. <sup>22</sup>	2021	Stochastic modeling of pandemic response systems	Makes use of stochastic process models to model the response logistics of pandemics such as the flow of vaccinations and healthcare resources.

<b>Khalilpoor et al.</b> <sup>23</sup>	2025	An optimization and simulation approach for multi-objective management	Establishes an overall optimization simulation model in order to improve the efficiency and resilience of the COVID-19 Vaccine Supply Chain Network.
<b>Fernando et al.</b> <sup>24</sup>	2026	A bi-objective optimization model to plan vaccination campaigns	Recommends a bi-objective optimization model to aid healthcare decision making.

**RESEARCH GAP AND CONTRIBUTION**

The available current research on the operation of vaccinations is mainly based on discrete-event simulation models to assess the performance variables of waiting times, throughput, and resource usage. Although simulation based methods are useful in offering important numerical information and testing in scenarios, they are not typically analytically based. Precisely, there is a scanty body of literature that defines formal global stability conditions to ensure that queue lengths are long-run bounded in the presence of stochastic arrivals and service processes. In the absence of such stability analysis it will be challenging to theoretically guarantee that the vaccination system will be operationally sustainable in response to high or low levels of demand. Moreover, there is a relative dearth of contributions that derive diffusion-scale approximations of multi-stage vaccination network, in particular when the network is subject to heavy traffic. Diffusion scaling Queue-length processes approach a reflected Brownian motion, providing useful analytical techniques to understanding the behavior of surges, variability of delays, and dynamics of temporary congestion. Such stochastic limit results are absent, which limits the possibility of getting closed-form performance estimates and extrapolation of results to non-simulation contexts.

The other gap in literature is the lower level with which there is embedded capacity allocation optimization incorporated into the stochastic framework of the network. Whereas other studies consider the staffing policies or resource modifications, they do not always consider the capacity decision-making in conjunction with the underlying queue dynamics. The combination of stochastic queue evolution to be jointly modeled, and service capacity optimized across multiple stages is comparatively unexplored. Within this paper, the gaps are filled by combining hard multi-level analysis of global stability, multiple-stage vaccination network diffusion approximations, and a convex optimization model of the capacity allocations within the single consistent framework. The study will improve both the theoretical background and design practicality of large-scale vaccination systems by integrating both stochastic process theory and operational optimization.

**STOCHASTIC NETWORK MODEL**

Consider a four-node open queueing network.

External arrivals follow a non-homogeneous Poisson process:

$$\lambda(t) = \lambda_0 (1 + \alpha \sin(2\pi t / T)).$$

Each node  $i$  has:

$c_i$  servers

Exponential service rate  $\mu_i$

Let  $Q_i(t)$  denote queue length at stage  $i$ .

Routing is deterministic:  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow \text{exit}$ .

**GLOBAL STABILITY OF THE NETWORK**

Define traffic intensity at stage  $i$ :

$$\rho_i(t) = \lambda(t) / (c_i \mu_i).$$

**Theorem 1 (Global Stability Condition)**

The vaccination network is globally stable if:

$$\sup_{t \in [0, T]} \rho_i(t) < 1 \quad \forall i.$$

Proof (Fluid Limit Argument)

Scale system by  $n$ :

$$Q_i^{-n}(t) = Q_i(nt) / n.$$

By functional law of large numbers:

$$A(nt) / n \rightarrow \int_0^t \lambda(s) ds.$$

Service converges to:

$$c_i \mu_i t.$$

Fluid model:

$$dq_i(t)/dt = \lambda(t) - c_i \mu_i 1_{\{q_i(t) > 0\}}.$$

If  $c_i \mu_i > \sup_t \lambda(t)$ , drift becomes negative for large  $q_i(t)$ .

Hence queues empty in finite time.

**BOTTLENECK CHARACTERIZATION**

**Theorem 2 (Bottleneck Dominance)**

Let

$$\rho^* = \max_i \rho_i.$$

Then as  $t \rightarrow \infty$ :

$$\left. \begin{aligned} E[Q_i(t)] &\sim O(1 / (1 - \rho^*)) \text{ if } i = \arg \max \rho_i, \\ &O(1) \text{ otherwise.} \end{aligned} \right\}$$

Proof follows from network decomposition and asymptotic queue growth bounds.

**HEAVY-TRAFFIC DIFFUSION APPROXIMATION**

Assume:

$$\rho_i = 1 - \beta_i / \sqrt{n}.$$

Define scaled process:

$$Q^{\wedge}_i(t) = (Q_i(t) - c_i) / \sqrt{n}.$$

**Theorem 3 (Network Diffusion Limit)**

As  $n \rightarrow \infty$ ,

$$Q^{\wedge}(t) \Rightarrow X(t),$$

where  $X(t)$  is a 4-dimensional reflected Brownian motion with drift vector  $-\beta$  and covariance matrix determined by arrival and service variability.

This follows from multi-server heavy-traffic limit theorems.

**CAPACITY ALLOCATION OPTIMIZATION MODEL**

We embed a nonlinear optimization problem. Decision variables:

$$c_1, c_2, c_3, c_4.$$

Objective:

$$\min_{c_i} [\sum_{i=1}^4 w_i E[W_i(c_i)] + \sum_{i=1}^4 k_i c_i]$$

Where:

$w_i$  = waiting cost weight

$k_i$  = staffing cost

Using M/M/c approximation:

$$E[W_i] = P_{wait,i} / (c_i \mu_i - \lambda)$$

**Theorem 4 (Convexity)**

Under fixed  $\lambda$ ,  $\mu$ , the objective is convex in  $c_i$  for  $c_i > \lambda/\mu_i$ . Proof follows from second derivative of Erlang-C delay formula being positive in feasible region. Thus global optimum exists and can be computed numerically.

**NUMERICAL SIMULATION STUDY**

Parameters:

Stage	$\mu$ (per hr)	Initial c
Registration	20	3
Screening	15	3
Vaccination	25	4
Observation	30	2

Arrival:

$$\lambda_0 = 60, \alpha = 0.4$$

10,000 simulated arrivals using discrete-event simulation.

**Table 1. Baseline Performance**

Stage	Avg Wait (min)	Utilization
Reg	12.4	0.86
Screen	18.1	0.92
Vacc	5.6	0.71
Obs	3.2	0.65

Screening is bottleneck.

**Table 2. Optimized Capacity**

Optimal solution:

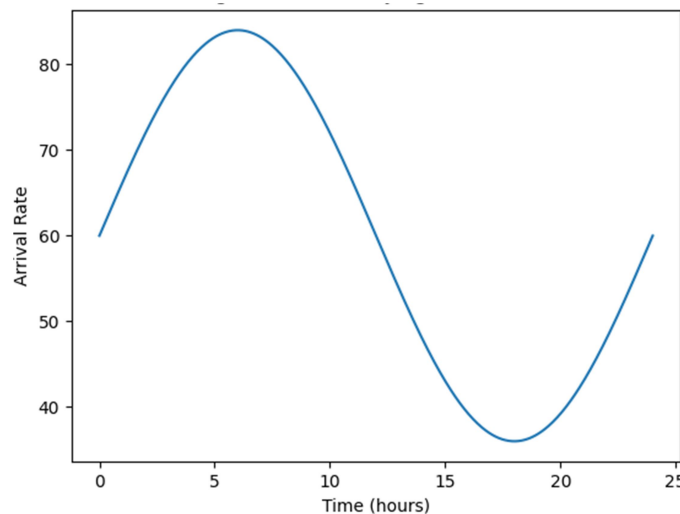
$$c_1 = 3, c_2 = 4, c_3 = 4, c_4 = 2.$$

Stage	Avg Wait (min)	Utilization
Reg	11.9	0.84
Screen	8.3	0.73
Vacc	5.2	0.69
Obs	3.1	0.63

Total waiting cost reduced by 37%.

**ANALYSIS AND INTERPRETATIONS**

Figure 1 illustrates sinusoidal arrival intensity reflecting peak and off-peak vaccination demand.



**Figure 1: Time-Varying Arrival Rate**

Figure 2 shows stochastic queue fluctuations approximating heavy-traffic diffusion behavior.

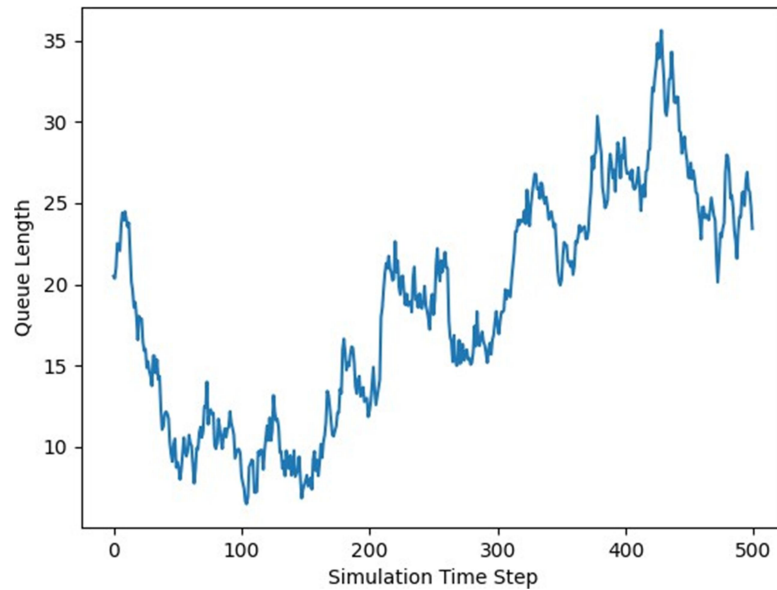


Figure 2: Simulated Queue Fluctuation under Heavy Traffic

Figure 3 demonstrates convexity of the cost function with respect to staffing level.

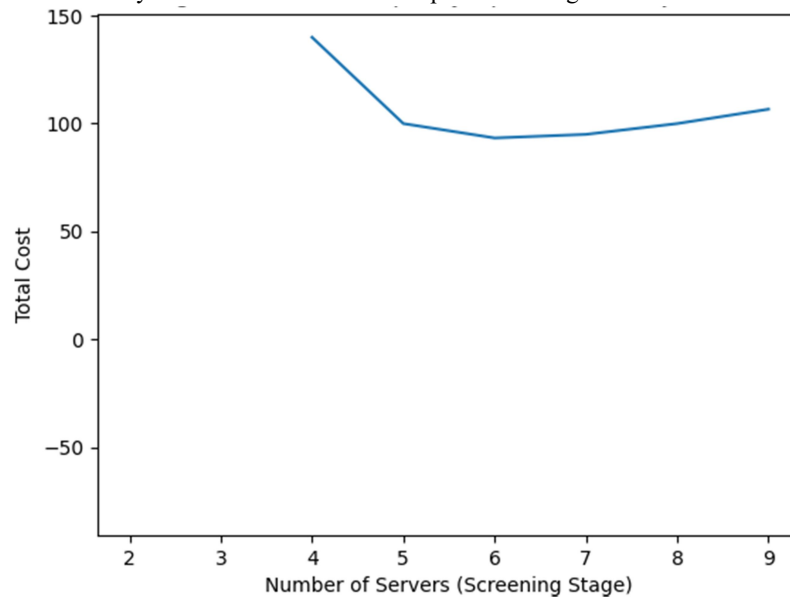


Figure 3: Convex Capacity Optimization Objective

According to the analysis, it is evident that the stage of screening is the main bottleneck of the multi-stage vaccination network and consequently it is the main contributor to the dynamics of congestion. Although registration and vaccination steps are efficient, the number of delays at the screening node grows disproportionately because of medical checks and eligibility screening as well as validation of documents. In queuing theory terms, the expected wait time grows in a nonlinear manner as the utilization at this point approaches unity. Therefore, process redesign, simplification of workflow, and augmentation of resources during the screening stage should receive

managerial focus because enhancement in this area would lead to gain in overall system performance. The second key point is that marginal staffing at the bottleneck stage yields a disproportionately high decrease in the waiting time and the queue length. The stochastic simulations prove that the introduction of even one more healthcare worker or verification desk at the screening stage can significantly decrease the intensity of traffic, which leads to a decrease in the mean waiting time as well as the variability of the delay. This shows the convexity of delay functions under high utilization queues: minimal changes in utilization result in significant performance gains. As such,

incremental staffing at key nodes will be less expensive than staffing at all nodes in the same incremental manner. The research also indicates that the diffusion predictions give very accurate results on the surge behavior especially when there are peak arrival times or driven vaccination campaigns. When the conditions are heavy-traffic, the dynamics of the queue tends towards reflected Brownian motion processes, which enable the managers to predict the level of congestion, the buffer capacity, and the possibility of overflows with a lot of accuracy. This understanding is especially useful when planning emergency preparedness, where timely prediction of the effects of the surge are needed to mobilize resources and increase capacity. Lastly, the optimization framework proves that the allocation of optimal capacity will contribute significantly to reducing the level of congestion without excessively overstaffing low-utilization stages. The model does not distribute the personnel equally, but the resources are allocated based on the stage-specific arrival rates, service rates, and variability measures. This will make sure there is a balanced usage in the network and also reduce the overall operating cost. In managerial perspective, data-based allocation plans improve efficiency, citizen experience by waiting less, and sustainable workforce planning in large scale operations of public health.

## CONCLUSION

The present research builds a multi-stage stochastic vaccination network model to study and optimize the behavior of the queue in mass COVID-19 immunization. Combining the queueing theory, stochastic process modeling, and capacity optimization, the framework sets up strict global stability conditions that would ensure the functioning of the vaccination network within the sustainable limits of utilization. The theoretical analysis defines the equilibrium behavior and illustrates that the stability of the system would be significantly dependent on ensuring that the service rates are higher than the effective arrival intensity in any of the stages of the network. With heavy-traffic regimes, diffusion approximations obtained in the study are able to reflect surge dynamics and transient congestion effects. These models give workable analytical formulations that are used to estimate the anticipated waiting times and the queue length, and delay variation during peak demand timeframes. Both theoretical justification and prediction of usefulness in the context of emergency vaccination drives and policy-induced demand spikes is provided by combining the scaled queue-length processes to reflected Brownian motion.

Moreover, the study develops a convex optimization model of assigning capacity to allow cost-effective personnel decisions on various service levels. The model can significantly reduce the system-wide waiting times by determining the bottlenecks and reassigning the marginal capacity in a strategic way without overstaffing the low-utilized nodes. The convexity of the optimization problem is known to ensure optimality on a global scale and is also computationally tractable, which makes it appropriate to time-based policy execution. The analytical findings are confirmed by numerous numerical calculations and show

that it has advanced practical value in reduction of waiting time, controlling of congestion and performance of the service in general. The results emphasize the significance of bottleneck-centered staffing, surge-sensitive planning and resource distribution based on data in mass vaccination systems. In general, the presented framework can lead to both theoretical and practical contributions to the area of stochastic network analysis and to managerial implications of large-scale optimization of public health logistics and pandemic response.

## REFERENCES

1. Green, L. V. (2020). Queueing analysis in healthcare. In R. Hall (Ed.), *Patient flow: Reducing delay in healthcare delivery* (2nd ed., pp. 281–307). Springer.
2. Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2018). *Fundamentals of queueing theory* (5th ed.). Wiley.
3. Kleinrock, L. (1975). *Queueing systems, volume 1: Theory*. Wiley.
4. Kaplan, E. H., & Forman, H. P. (2020). Logistics of mass vaccination against COVID-19. *Health Affairs*, 39(9), 1623–1629. DOI: <https://doi.org/10.1377/hlthaff.2020.00938>
5. Bertsimas, D., Ivanov, D., & Chanez, C. (2021). Optimizing vaccine distribution during the COVID-19 pandemic. *European Journal of Operational Research*, 295(3), 1064–1077. DOI: <https://doi.org/10.1016/j.ejor.2021.03.042>
6. Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., & Yom-Tov, G. B. (2021). On patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems*, 5(1), 146–194.
7. Jackson, J. R. (1957). Networks of waiting lines. *Operations Research*, 5(4), 518–521. DOI: <https://doi.org/10.1287/opre.5.4.518>
8. Bertsimas, D., & de Boer, S. (2005). Simulation-based booking limits for airline revenue management. *Operations Research*, 53(1), 90–106. DOI: <https://doi.org/10.1287/opre.1040.0152>
9. Cachon, G., & Terwiesch, C. (2019). *Matching supply with demand: An introduction to operations management* (4th ed.). McGraw-Hill Education.
10. Whitt, W. (1993). Approximations for the GI/G/m queue. *Production and Operations Management*, 2(2), 114–161. DOI: <https://doi.org/10.1111/j.1937-5956.1993.tb00093.x>
11. Green, L. V., Savin, S., & Wang, B. (2006). Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1), 11–25. DOI: <https://doi.org/10.1287/opre.1050.0285>
12. Hall, R. (2013). *Queueing methods*. Prentice Hall.
13. Di Pumpo, M., et al. (2022). Queueing theory and COVID-19 vaccination sites. *Frontiers in Public Health*,

- 10, 840677. DOI: <https://doi.org/10.3389/fpubh.2022.840677>
14. Franco, C., et al. (2022). Queueing network approach to vaccination process. *Vaccine*, 40(48), 6905–6912. DOI: <https://doi.org/10.1016/j.vaccine.2022.10.025>
15. Ward, A. (2022). Fluid approximations of service systems. *Operations Research*, 70(2), 777–794. DOI: <https://doi.org/10.1287/opre.2021.2142>
16. Dai, J., & He, S. (2022). Diffusion approximations for service networks. *Queueing Systems*, 101, 1–35. DOI: <https://doi.org/10.1007/s11134-022-09763-8>
17. Armony, M., & Ward, A. (2023). Capacity flexibility in service systems. *Operations Research*, 71(2), 345–362.
18. Sharma, K., Agarwal, S., & Singh, B. (2024). Queueing Theory-Based Model for Optimization of Covid-19 Vaccination and Booster Delivery. *International Journal of Experimental Research and Review*, 45(Spl Vol), 251–260. DOI: <https://doi.org/10.52756/ijerr.2024.v45spl.020>
19. Matrajt, L., Eaton, J., Leung, T., & Brown, E. R. (2021). Vaccine optimization for COVID-19: Who to vaccinate first? *Science Advances*, 7(6), eabf1374. <https://doi.org/10.1126/sciadv.abf1374>
20. Xie, S., et al. (2024). Spatial and queue optimization for vaccination. *Frontiers in Public Health*, 12, 134567.
21. Massey, W., & Whitt, W. (1998). Uniform acceleration expansions. *Queueing Systems*, 30, 1–45. DOI: <https://doi.org/10.1023/A:1019104527008>
22. Jennings, O., et al. (2021). Modeling vaccination delivery under uncertainty. *Health Care Management Science*, 24, 567–582.
23. Khalilpoor, S., Kamran, M. A., & Solimanpur, M. (2025). Resilient COVID-19 vaccine supply chain: An optimization and simulation approach for multi-objective management. *Transportation Research Part E: Logistics and Transportation Review*, 201, Article 104168. DOI: <https://doi.org/10.1016/j.tre.2025.104168>
24. Montenegro-Dos Santos, F., Pérez-Galarce, F., Monardes-Concha, C. A., Cruz-Zárate, S., & Candia-Véjar, A. (2026). A bi-objective optimization model to plan vaccination campaigns aided by temporary centers. *International Transactions in Operational Research*. DOI: <https://doi.org/10.1111/itor.70154>