International Journal of Allied Sciences (IJAS)
Volume.16, Number 10; October, 2025;
ISSN: 2836-3760| Impact Factor: 8.13
https://zapjournals.com/Journals/index.php/Allied-Sciences
Published By: Zendo Academic Publishing

A MULTI ECHELON QUEUING THEORETIC APPROACH TO TIME SENSITIVE SERVICE LEVEL OPTIMIZATION FOR MEDICAL SUPPLY CHAIN DISTRIBUTION DURING PANDEMIC OUTBREAKS

[1]Ominigbo Ojay Jacob and [2]Okedoye Akindele Michael

Article Info

Keywords: Pandemic Outbreaks, Multi Echelon, Queuing Theory, Lateral Trans-shipment, Service Level Optimization, Time Sensitive, Medical Supply Chain, and Queuing Theoretic.

DOI

10.5281/zenodo.17466506

Abstract

During a pandemic, the distribution of time-sensitive medical supplies, such as vaccines, poses a critical public health challenge characterized by stochastic demand and stringent response time constraints. This study presents a novel multi echelon queuing theoretic model for the integrated optimization of facility location, inventory allocation, and distribution logistics. We model the supply network as a system of interconnected facilities where demand follows a Poisson process and incorporate two key flexibilities: lateral trans-shipments between distribution centers and pipeline stock management from central depots. The objective is to minimize total system costs including fixed facility, inventory holding, pipeline, and lateral transportation costs while ensuring that a target demand fraction is fulfilled within a maximum allowable response time. The problem is formulated as a Mixed Integer Linear Program MILP and solved using a Lagrangian relaxation scheme coupled with a sub gradient optimization algorithm. Numerical simulations, based on a case study of 40 demand locations in Delta State, Nigeria, demonstrate the model's efficacy. The results show that the optimized network consistently achieves high Time Based service levels $SL_i^T > 0.7$ for all locations, with many exceeding 0. 95 by effectively balancing inventory across echelons and leveraging lateral trans-shipments to mitigate local shortages. This approach provides a quantifiable decision support tool for designing resilient, responsive, and cost effective medical supply chains for future public health emergencies.

[1],[2]Department of Mathematics, Federal University of Petroleum Resources, Effurun, Nigeria.
E-mail: ojayominigbo@gmail.com, okedoye.akindele@fupre.edu.ng

1ntroduction

The COVID-19 pandemic has revealed the vulnerabilities of global medical supply chains. The surge in demand for critical, often perishable, items, such as vaccines and personal protective equipment (PPE), exposed profound inefficiencies in traditional distribution models, leading to critical shortages in some areas and oversupply in others. A key lesson is the need for an integrated approach that simultaneously optimizes the strategic placement of facilities location the quantity of supplies held at each inventory point and the dynamic routing of goods transshipment under highly uncertain, time sensitive conditions.

The development of COVID-19 vaccines was a monumental scientific achievement. However, its public health impact was contingent on an equally complex logistical challenge: the rapid, global distribution of a perishable commodity under intense demand uncertainty. For instance, an mRNA vaccines have strict and short thermal stability outside cold storage, creating a use it or lose its reality. Bottlenecks plagued the initial phases of the rollout, leading to the risk of vaccine spoilage and critical treatment delays. Traditional SC models, which often focus on deterministic flows and treat location and inventory problems separately, are ill equipped to handle this stochastic, time sensitive environment.

This study addresses this gap by developing a comprehensive multi echelon queuing theoretic framework that integrates these decisions. Our model is specifically designed for the distribution of time sensitive medical supplies, where exceeding a maximum response time T can render a vaccine ineffective or lead to a life threatening shortage. Building on the foundational work of Yang et al. 2013 we model each distribution center as an M (n)/D/∞ queuing system. This allows us to accurately capture the stochastic nature of demand arrival and service times. The key innovations of the model are the incorporation of lateral transshipment**s**, which allows stock to be shared between peer facilities to quickly address local surges and the explicit management of pipeline stocks from central suppliers.

The primary research question is: How can a multi echelon medical supply network be designed to minimize total cost while guaranteeing a high probability of meeting strict, time sensitive delivery windows during a pandemic?

2. Literature Review

The coronavirus disease 2019 (COVID-19) pandemic served as a stark stress test for global medical supply chains, exposing critical vulnerabilities in the logistics of distributing time sensitive medical supplies. A significant body of research has documented the widespread disruptions. For instance, Moosavi et al. (2022) [3] analyzed potential disruption management strategies, highlighting the need for resilience in the face of such shocks. The literature specific to COVID- 19 vaccine distribution consistently highlights several interconnected challenges: stringent and unbroken cold chain requirements Lee et al., 2021), profound demand uncertainty during rollout phases, and the logistical nightmare of last -mile distribution to vaccination centers.

Studies in low resource settings further emphasize systemic weaknesses. In their assessment of Delta State, Nigeria, Taigbenu et al. (2025) [2] explicitly detailed how cold chain limitations, workforce shortages, and community hesitancy led to a decline in immunization coverage, demonstrating how pre-existing system frailties are exacerbated during a pandemic. Similarly, Haidari et al. [2016] used simulation to show that optimizing the location of storage facilities could dramatically improve vaccine availability in low and middle income countries. Although these studies excellently describe the problems and some strategic level solutions, a recognized gap exists in quantitative, prescriptive models that can dynamically optimize inventory and transshipment decisions across a multi -tiered network under the real world constraint of a strict, time sensitive usable window for prepared vaccines.

However, the application of queuing theory has largely been confined to the service delivery poin**t** within a healthcare facility. Research that extends queuing models upstream into the supply chain and inventory

management processes that feed these service points is scarce. Berman and Larson (2001) introduced spatial queuing for emergency vehicle deployment, but its application to medical supply chains is limited. The use of queuing theory to model a *network* of distribution centers and vaccination hubs as an integrated system of interconnected queues, where the "service" is the fulfillment of a vaccine order, remains underexplored. Recent studies, such as Arora et al. (2021**)**, have begun applying queuing models to vaccination centers for capacity planning, but they stopped at the clinic door. This paper bridges that gap by modeling the entire supply network using multi echelon queuing systems like M(n)/D/∞ as inspired by Yang et al. [2013], to capture the stochastic nature of demand and replenishment lead times across all levels, not just the final service point.

This research builds directly upon the literature on integrated inventory-transshipment. The work of Yang et al. [2013] on Service Parts Inventory Control with Lateral Transshipment and Pipeline Stock Flexibility is particularly relevant because it provides a queuing based analysis for a similar base stock (S- 1, S) record point inventory control policy system.

This study aims to address four critical gaps in the existing literature on supply chain management and health care logistics. First, it confronts the lack of integrated, quantitative optimization models for pandemic vaccine distribution that can dynamically handle the pervasive stochastic demand and supply. Second, it extends the traditional application of queuing theory within clinical settings by extending it to model the entire upstream medical supply network as a multi echelon queuing system. Third, it introduces explicit, hard time based service level constraints intoSL_L^T the supply chain model, a feature that is critical for managing perishable commodities like vaccines but often absent from existing integrated models. Finally, it specifically investigates the underutilized potential of lateral transshipment as a dynamic strategy for mitigating shortages of highly perishable medical supplies during a pandemic outbreak.

3.0 Problem formulation and mathematical model

3.1 Problem formulation

We consider a multi echelon supply network in Delta central senatorial district, Delta state, Nigeria comprising selected 40 healthcare facilities used as vaccination centers J (33 Primary Healthcare Centres and 7 Hospitals) and 8 distribution hubs (locations) L, (five Hospitals and three Primary Healthcare Centtrs) Table 1.

Table1. Health Care Facilities for COVID -19 immunization services and Vaccine Distribution Hubs across Delta Central Senatorial District in Delta State, Nigeria.

S N	Local government area	HFC	Vaccination Center and Distribution Hub
	ETHIOPE EAST		
1	Isiokolo	PHC	Both
2.	Eku	PHC	Vaccination Center
3.	Ovu	PHC	Vaccination Center
4.	Abraka	PHC	Vaccination Center
ETHIOPE WEST			
1.	DELSUTH		Both
2.	Jesse	PHC	Vaccination Center
3.	Mossogar	PHC	Vaccination Center
4.	Evade Ovadje	PHC	Vaccination Center
5.	Ogharaeki	PHC	Vaccination Center
6.	Ogharaefe	PHC	Vaccination Center
SAPELE			

1. Central Hospital Sapele		Both
2. Gana	PHC	Vaccination Center
3. Amukpe	PHC	Vaccination Center
4. Etamu	PHC	Vaccination Center
5. Urban	PHC	Vaccination Center
OKPE		
1. Orerokpe	PHC	Both
2. Adeje	PHC	Vaccination Center
3. Ughotor	PHC	Vaccination Center
4. Egborode	PHC	Vaccination Center
5. Okwokoko	PHC	Vaccination Center
UDU		
1. Ekete	PHC	Vaccination Center
2. Emadadja	PHC	Vaccination Center
3. Orhuwhorun	PHC	Vaccination Center
4. General Hospital, Udu		Both
5. Ovwian	PHC	Vaccination Center
6. Opete	PHC	Vaccination Center
UGHELLI NORTH		
1. Ekiugbo	PHC	Vaccination Center
2. Central Hospital Ughelli		Both
3. Government Hospital, Orogun,		Vaccination Center
4. Agbarho Government Hospital		Vaccination Center
5. Uwheru	PHC	Vaccination Center
UGHELLI SOUTH		
1. Otu Jeremi	PHC	Vaccination Center
2. Okpare	PHC	Vaccination Center
3. Ewu	PHC	Vaccination Center
4. Usiefurun	PHC	Vaccination Center
UVWIE		
1. Ugbomro	PHC	Vaccination Center
2. Ogboroke	PHC	Vaccination Center
3. Ekpan General Hospital		Both
4. Ogborikoko	PHC	Vaccination Center
5. Enerhen-1	PHC	Vaccination Center

(Source: Asaba, Delta State Primary Health Care Development Agency)

3.2 Mathematical Model Formulation

3.2.1 Sets and indices

- $l \in L$ index for location of demand vaccination centers)
- $j \in$ Jindex for supply sources distribution hubs)

3.2.2 Demand and Service Parameters

- λ_lDemand arrival rate at location l, assumed to follow a Poisson process.

- μ service rate at a vaccination point at the center.
- s: number of identification vaccination points at a center
- $p = \lambda/(s\mu)$: Utilization factor of the queuing system; stability requires $\rho<1$.
- T: Maximum response time threshold (6 hours for thawed vaccine). Demand fulfilled after T is late.

3.2.3 Inventory policy and stock variables

- S_l : Base stock level at demand location *l* under continuous review$S_l - 1$ policy.
- N_lNumber of outstanding orders or backorders at demand location *l, respectively*
- IK_l Inventory$S_l - N_l$ at demand location *l.*
- I_HExpected average inventory holding level at the demand location *l.*
- I_PExpected average pipeline inventory between supplier j and location *l.*

3.2.3 Service level and performance metrics

- SL_l^0Instantaneous service level at location *l* fraction of demand fulfilled immediately from local stock.
- SL_l^TTime -based service level at location *l* —friction of demand fulfilled within the maximum time threshold T.
- ω_lFraction of demand at location *l* fulfilled from pipeline stock within time T.
- α_{lj}:Fraction of demand at local *l* fulfilled by lateral transshipment from location j.
- θ_lFraction of demand at location *l* not fulfilled within threshold T. So SL_l^T= 1- θ_l.

3.2.4 Cost parameters

- C_{lj}Unit cost of lateral transshipment from location j to *l.*
- c_lUnit inventory holding cost at location *l.*
- c_{pl}Unit pipeline stock cost for location *l.*
- F_lFixed cost of the operating facility at location *l.*

3.2.5 Decision variables

- $Y_l \in$Binary$\{0,1\}$ variable indicating whether a facility at demand location *l* is open if closed,$Y_l = 1$ if $Y_l = 0$
- $X_{lj} \in$Binary$\{0,1\}$ variable whether demand location *l* is assigned to supply source or distribution hub j, if assigned $X_{lj} = 1$otherwise $X_{lj} = 0$.

3.2.6 *Queuing* and Inventory Model Parameters

- δ_lDemand process at location *l.*
- γ_l:Replenishment process at location *l.*
- p_nSteady state probability of having n orders in the system at location *l* at location *l.*
- K_lLead time or replenishment time.
- p^0Normalizing constant or empty system probability.

3.2.7 Lagrangian Relation Parameters

- $\lambda_1, \lambda_2, \lambda_3, \lambda_4$Lagrangian multiplier for complicating constraints.
- L_λLagrangian function.
- g_λDual function.
- α k Step size at iteration k in the subgradient algorithm

3.2.8 Other parameters and constants

- B: capacity of demand location *l.*

- M: Number of node locations.
- d_{lj}Distance between demand location *l* and distribution hub j.
- S_{maxl}Maximum permitted stock level at the demand location *l.*
- ψ_lMaximum required service level for SL_l^0andSL_l^T.

3.3.1 Queuing Model (Vaccination Center as an M/M/s Queue):

- Arrivals: Patient arrivals follow a Poisson process with mean rate λ.
- Service: The center has identical vaccination stations. The service time per patient is exponential with mean rate μ. The service rate μ is a decision variable representing resource intensity.
- Stability: The system is stable if $\rho = \lambda / (s\mu)<1$.

3.3.2 Model Assumptions:

- Demand: The demand for medical supplies e. g., orders from clinics at each location *l* follows a rate-based Poisson process. λ_l

Service**:** Each location operates under a continuous-review (S_l-1,S_l) inventory policy.

- Time-sensitivity: A maximum response time threshold T is defined e. g., 6 hours for a thawed vaccine. Demand fulfilled after T incurs a high penalty, representing the cost of a spoiled dose or a critical treatment delay.
- Flexibility**:** Lateral transshipments between locations are permitted if the transfer can be completed within T.

3.4.1 Queuing-Based Inventory Analysis:

Following the methodology in the provided paper, the inventory behavior at each location is modeled as an M (n)/D/∞ queue. The steady state probabilities of this queue are used to derive the expressions for SL_l^0, ω_l and α_{lj}, which are then embedded within the optimization model.

3.4.2 Key performance metrics (service levels)

The model tracks two critical service levels for each location *l*:

1. Instantaneous Service Level (SL_l^0): The SL_l^0 fraction of demand fulfilled immediately from the local stock.
2. Time Based Service Level (SL_l^T): The SL_l^T fraction of demand fulfilled within the maximum time threshold T. This is the primary metric for time sensitive supplies and is calculated as follows:

$$SL_l^T = SL_l^0 + \omega_l + \sum_{j \in J} \alpha_{lj}$$

Where, ω_l,is the fraction fulfilled from the pipeline within T, and $\alpha_{lj}$is the fraction fulfilled via lateral transshipment from location j.

3.4.3 Computation of Performance Metrics

Through the queueing analysis, we compute the key performance metrics needed for the optimization. Namely, using the steady-state probabilities of the M(n)/D/∞ queue Yang et al.[2013, Section5] hence for an inventory base stock $S_l - 1, S_l$ control policy, the inventory at demand location l satisfies. $IK_l = S_l - N_l$ Modeling the inventory level distribution at each location l for any j:$\leq S_l$

$P(IK_l = j) = P_0(S_l - j, \Lambda_l K_l)$, the formula for the poison probability mass function:

$$P_0(k_l, \square) = \sum_{i=0}^{k} p_0(i, \beta) \text{ provides the likelihood of obtaining that value k, given the underly}$$

The Poisson rate represented□ by the demand arrival rate.Λ_l Therefore, by applying the PASTA property, the fraction of the demand immediately satisfied from on hand stock can be determined as follows:

$$p^0 = \frac{1}{\sum_{n=0}^{S_l} \frac{(\delta_l K_l)^n}{n!} + \left(\frac{\delta_l}{\gamma!}\right)^{S_l} \sum_{n=S_l}^{\infty} \frac{(\gamma_l k_l)^n}{n!}}$$

Furthermore,

$$SL_l^0 = P(N_l<S_l) = p^{0e^{\delta_l K_l}} \sum_{n=0}^{S_{l-1}} p^0 (n_i \delta_l K_l) = p^{0e^{\delta_l K_l}} P_0(S_l - 1, \delta_l K_l) = p^{0e^{\delta_l K_l}} \left(\sum_{n=0}^{S_l} \frac{(\delta_l K_l)^n}{n!} e^{-\delta_l K_l} \right)$$

The fraction of L's demand fulfilled by the pipeline within time T is given as follows:

$$\omega_l = P(0<W_l \leq T) = p^{0e^{\gamma_l K_l}} \left(\frac{\delta_l}{\gamma_l}\right)^{S_l} [P^0(S_l - 1, \gamma_l(K_l - T)) - P_{0(S_l} - 1, \gamma_l K_l)]=$$

$$p^{0e^{\gamma_l K_l}} \left(\frac{\delta_l}{\gamma_l}\right)^{S_l} \left(\left[\sum_{n=0}^{S_l-1} \frac{(\gamma_l K_l)^n}{n!} e^{-\gamma_l K_l} \right] - \left| 1 - \sum_{n=0}^{S_l-1} \frac{\left(\gamma_l (K_l - T)\right)^n}{n!} e^{-\gamma_l (K_l - T)} \right| \right)$$

The fraction of location L's demand fulfilled by j can be estimated as follows:

$$\alpha_{lj=(1-SL_l^0-\omega_l)\left(1-SL_{lj}^0\right)\ldots\left(1-SL_{lj-1}^0\right)SL_{lj}^0}$$

The instantaneous service level achieved for location L, specifically considering the total demand filled from the local inventory stock, is obtained as follows:

$$SL^0 = \sum_{l \in J} SL_l^0 \frac{\Lambda_l}{\sum_{l \in J} \Lambda_l}$$

The fraction of demand satisfied within a maximum response time threshed T is given by the following:

$$SL_l^T = SL_l^0 + \omega_l + \sum_{j \in J} \alpha_{lj}$$

The total demand of location L satisfied within a maximum response time threshold T can be estimated as follows:

$$SL^T = \sum_{l \in J} SL_l^T \frac{\Lambda_l}{\sum_{l \in J} \Lambda_l}$$

The expected average inventory level held at location L in the long run is as follows:

$$I_H = E[(S_l - N_l)^+] = \sum_{n=0}^{S_l-1} (S_l - n) p_n$$

Wherep_n, the steady-state probabilities are derived as

$$\begin{cases} p^{0\frac{(\delta_l K_l)^n}{n!}} & \text{if } n < S_l \\ p^0 \left(\frac{\delta_l}{\gamma_l}\right)^{S_l \frac{\left(k_l \gamma_l\right)^n}{n!}} & \text{if } n \geq S_l \end{cases}$$

The long run expected average inventory held in the pipeline between supply j and location L is then estimated as follows:

$$\sum_{n=0}^{\infty} n P_n$$

Finally, the long run expected average demand at location L fulfilled by lateral transshipment from depot distribution hub j is derived as follows:

$L_{AT}= \alpha_{lj\lambda_j}$

These metrics were iteratively estimated following Yang et al. (2013,) to enable the evaluation of the objective function and constraints for any candidate solution to the optimization model. Note that for location L:

$$SL_l^0 + \omega_l + \sum_{j\in J} \alpha_{lj\lambda_j} + \theta_l = 1$$

Furthermore, $SL_l^{T=}1\text{-}\theta_l$

3.5 Integrated Optimization Model

With the problem context and formulation defined. We present a mathematical model for a multi echelon queuing theoretic approach to time sensitive service level optimization for medical supply chain distribution during pandemic outbreaks with the objective of guaranteeing a high probability of meeting strict, time sensitive delivery windows during a pandemic subject to service level constraints.

Objective Function:

$$\min \sum_{l\in L}\left(\sum_{j\in J} c_{lj} L_{AT} X_{ij} + \left(c_l I_H + c_{pl} I_p\right) Y_l + f_l Y_L\right) \qquad (1)$$

This includes Lateral Transshipment Cost InventoryL_{AT} Holding Cost PipelineI_H Stock Cost andI_p Fixed Facility Costs.f_l

3.5.1 Model constraints

Subject to

- Service Level Constraints or Time Sensitivity Constraints: Ensuring that minimum service standards are met within the time of the trench hold. They mathematically define the time sensitive goal by mandating that a minimum fraction of demand must be fulfilled within the critical time window T.

$$SL^0 \geq \varphi_l Y_l \qquad \forall l \in L \qquad (2)$$

$$SL^T \geq \tau_l Y_l \qquad \forall l \in L \qquad (3)$$

- Single-source constraint: Each demand point is assigned to one supply source.

$$\sum_{j\in J} X_{lj} = 1 \qquad \forall\, l \in L \qquad (4)$$

- Capacity constraint: The total demand assigned to a demand location l (if it is open) should not exceed its capacity.

$$\sum_{j\in J} X_{lj} \leq B Y_j \qquad \forall j \in J \qquad (5)$$

- Lead Time Distance constraint**:** This constraint ensures that the average distance and thus, transportation time between echelons is bounded by the replenishment lead time.N_L During a pandemic, a vaccine shipment cannot travel an unrealistic distance that would cause spoilage. This constraint rules out inefficient and dangerous solutions.

$$\frac{\Sigma d_{lj}}{M} \leq K_l \qquad \forall \quad l \in L, j \in J \qquad (6)$$

- Maximum stock level or real-world resource bound constraint: This constraint limits the number of vaccines that can be distributed at a distribution hub. The cost or quantity of the base stock level at S_l location l must not exceed a maximum allowable values S_{maxl} but only if the location is open.

$$c_l S_l \leq S_{maxl} Y_l \qquad \forall l \in L \qquad (7)$$

- Emergency Sharing Limit or Lateral transshipment proportion Constraint: This condition states that the total help a vaccination center receives from others Lateral transshipment cannot exceed 100% of its needs.

$$\sum_{j \in J} \alpha_{lj} \leq 1 \qquad \forall\, l \in L \qquad (8)$$

- Logical linkage constraint: This constraint links the assignment variable to X_{ij} the facility opening decision variable. Y_l It states that vaccines can only be assigned to a vaccination center if the center is open. If L is closed, the vaccine cannot assign Y_l=0.

$$\sum_{j \in J} X_{lj} \leq Y_l \qquad \forall\, l \in L \qquad (9)$$

- The binary nature of location constraint: This specifies that if the variable Y_l which indicates whether a facility is opened at location l, must be either 0 closed or 1 open.

$$Y_l \in \{0, \}, \text{for each} \qquad \forall\, l \in L \qquad (10)$$

- Assignment Decision Constraint's Binary Nature: This constraint specifies that the assignment variable X_{lj} which indicates whether demand location l is assigned to depot j, must be binary 0 or 1.

$$X_{lj} \in \{0,1\}, \text{for each} \qquad \forall\ l \in L\ j \in J \qquad (11)$$

4.0 Case Study: The COVID-19 Vaccine Distribution Problem

4.1 Distribution of Covid-19 Vaccines in Delta State, Nigeria

We frame our model around a multi-echelon, large-scale vaccination hub during the initial mass inoculation phase of the Johnson and Johnson Covid-19 vaccine distribution of 87,038 doses of Moderna for the first dose vaccination and 16,080 doses of Oxford Astrazenica for the second dose vaccination across the three senatorial districts in Delta state, Nigeria Table 2.

Table 2. Distribution of the first and second doses of Covid-19 Vaccines across Delta State Senatorial Districts

Senatorial District	Moderna Doses	Oxford AstraZenica Doses
Delta North	30,012	5,428
Delta Central	29,009	6,012
Delta South	28,017	4,640

(Source: Asaba, Delta State Primary Health Care Development Agency)

4.1.2 Distribution of COVID-19 Vaccines in the Delta Central Senatorial District

This study focused on the distribution of the Johnson and Johnson vaccines in the Delta central senatorial district for the first (Moderna) and second (Oxford Astrazenica) doses across the 40 selected healthcare facilities (HCFs) which comprised (32 Primary Healthcare Centres) and (8 Hospitals) which served as both primary and secondary healthcare facilities . Table3.

Table3. Distribution of the Johnson and Johnson COVID-19 Vaccines for 1st and 2nd doses in the Delta Central Senatorial District of Delta State, Nigeria

S /N	PHC	Local Govt. Area	Moderna (Doses)	Oxford Astrazenica (Doses)
1.	Isiokolo PHC	ETHIOPE EAST	821	149
2.	Eku PHC		796	125
3.	Ovu PHC		781	117
4.	Abraka PHC		1012	131
5.	DESULTH	ETHIOPE WEST	852	208
6.	Jesse PHC		563	112
7.	Mosogar		688	115
8.	Ovadje PHC		634	102
9.	Oghareki PHC		672	125
10.	Ogharefe PHC		659	131
11.	Gana PHC	SAPELE	692	136
12.	Amukpe PHC		782	153
13.	Central Hospital Sapele		897	222
14.	Etamu PHC		633	126
15.	Urban PHC		657	132
16.	Orerokpe PHC	OKPE	875	207
17.	Adeje PHC		671	101
18.	Egborode PHC		563	121
19.	Ughotor PHC		651	127
20.	Okwokoko PHC		801	133
21.	Ekete PHC	UDU	654	163
22.	Emadadja PHC		532	109
23.	Orhuwhorun PHC		637	148
24.	General Hospital, Udu		578	142
25.	Ovwian PHC		655	166
26.	Opete PHC		516	115
27.	Ekiugbo PHC	UGHELLI NORTH	784	208
28.	Central Hospital Ughelli		998	233
29.	Government Hospital, Orogun,		785	181
30.	Agbarho Government Hospital		887	217
31.	Uwheru PHC		661	139
32.	Oto Jeremi PHC	UGHELLI SOUTH	876	176
33.	Okpare PHC		532	118
34.	Ewu PHC		478	101
35.	Usiefurun PHC		858	127
36.	Ugbomro PHC	UVWIE	627	159
37.	Ogboroke PHC		763	171
38.	Ekpan General Hospital		1071	229

39.	Ogborikoko PHC	702	164
40.	Enerhen-1 PHC	715	173

(Source: Asaba, Delta State Primary Health Care Development Agency)

3.1 Characteristics of the Problem

Time Sensitivity: The Johnson and Johnson vaccine, for instance, had to be used within 6 hours once thawed and diluted. This creates a strict maximum allowable time (T_{max}) between the preparation of a vaccine dose and its administration.

Stochastic Demand: Appointments, walk in, and group arrivals lead to highly variable and unpredictable individual arrival patterns, best modeled as a Poisson process with rate λ patients hour.

Controllable Service: The service process registration, screening, injection, and observation can be accelerated by adding more staff vaccinators, administrators and stations, increasing the service rate μ patients /server).

4.0 Solution and numerical simulation methods

4.1 Solution Method

After establishing the problem context and motivations, this section outlines the proposed solution methodology. Given the integrated, multi echelon nature of the problem with strategic decisions around facilities, inventories, and their alignment, a queueing inspired optimization approach is well suited. Specifically, a mixed integer linear program (MILP) formulation is developed to optimize service network configuration and dynamic inventory allocation policies. This approach represents facilities as nodes in a network with decision variables determining optimal location sizes and capacities. Manufacturer plants and final customers markets are designated as special nodes to model the supply and demand boundaries. A Lagrangian relaxation scheme is applied to decompose the large MILP into tractable sub problems for solution efficiency. The network structure is exploited to derive optimality conditions that enable iterative policies to reach coordination. Computational experiments across demand scenarios were conducted to assess balancing costs and fill rates under various flexibility settings. To apply Lagrangian relaxation, we identify the following complicating constraints: (2, 3, 4, and 9) that couple the facility location (Y)) and assignment (X) decisions. We introduce non-negative Lagrangian multipliers $\lambda_k \forall\ k = 1, 2,3,4$corresponding to these constraints. Where

$$\eta_{lj} = \sum_{l\in L}\left(\sum_{j\in J} c_{lj} L_{AT} X_{lj} + \left(c_l L_H + c_{pl} I_p\right) Y_l + f_l Y_l\right) \qquad (12)$$

The partial Lagrangian function is form by moving the complicating terms to the objective function with their multipliers as follows:

$$L_\lambda = \eta_{lj} + \lambda_1\left(\sum_{l\in I}\sum_{j\in J} X_{lj} - 1\right) + \lambda_2 \sum_{l\in L}\left(\sum_{j\in J} X_{lj} - Y_l\right) + \lambda_3 \sum_{l\in L}(\varphi_l Y_l - SL^0) + \lambda_4 \sum_{l\in L}(\tau_l Y_l - SL^T) \qquad (13)$$

Subject to

$$\sum_{j\in J} X_{il} \le B Y_j \qquad \forall j \in J \qquad (14)$$

$$\sum_{j\in J} \alpha_{lj} \le 1 \qquad \forall\ l \in L \qquad (15)$$

$\frac{\Sigma d_{lj}}{M} \leq K_l \quad \forall l \in L, j \in J$ (16)

$c_l S_l \leq S_{maxl} Y_l \forall l \in L$ (17)

$Y_l \in \{0,\}, \text{for each} \quad \forall l \in L$ (18)

$X_{lj} \in \{0,1\}, \text{for each} \quad \forall l \in L\ j \in J)$ (19)

Taking the derivative of L_Λ with respect to each primal variable and $Y_l X_{lj} SL^0 SL^T$ setting it to zero, we derive the optimality conditions, we obtained $\Lambda_3 \Lambda_4 = 0$), so that

$$g(\Lambda_j) = L_{\Lambda 3.\Lambda 4} = \eta_{ij} + \Lambda_1 \left(\sum_{l \in L} \sum_{j \in J} X_{ij} - 1 \right) + \Lambda_2 \sum_{l \in L} \left(\sum_{j \in J} (X_{lj} - Y_l) \right) \quad (20)$$

This quantity $g(\Lambda)$ represents the minimum value achieved by L_Λ hence the partial Lagrangian dual function. It forms the basis for the iterative sub gradient method to obtain the optimal primal solution.

4.1.1 Sub-gradient Optimization

Having derived the dual function $g(\Lambda_j)$ to develop the iterative algorithm for maximizing g and solving the dual problem, we apply a sub gradient optimization method. Where the partial derivatives of g with respect to the Lagrange multipliers Λ_1 and Λ_2 is obtained, these partial derivatives represent sub gradients of g because g is non smooth due to the inequality constraints involved in the formulation of the primal problem. Specifically, the dual problem is as follows:

$max\ g(\Lambda)$ (21)

Subject to

$$\frac{\partial g}{\partial \Lambda_1} = \sum_{l \in L} \sum_{j \in J} X_{lj} = 1 \quad (22)$$

$$\frac{\partial g}{g \Lambda_2} = \sum_{l \in L} \left(\sum_{j \in J} X_{lj} - Y_l \right) \quad (23)$$

Here, the gradients are:

$$s_1 = \nabla \Lambda_1 g \sum_{l \in L} \sum_{j \in J} X_{lj} - 1 \quad (24)$$

$$s_2 = \nabla \Lambda_2 g = \sum_{l \in L} \left(\sum_{j \in J} X_{lj} - Y_l \right) \quad (25)$$

To solve this via the sub-gradient method, we initialize the multiplier vector as follows:

$\Lambda_1(0) = 0$ (26)

$\Lambda_2(0) = 0$ (27)

The sub-gradient algorithm then iteratively updates the multipliers at iteration k as follows:

$\Lambda_1(k+1) = \Lambda_1(k) + \alpha(k) . s_1(k)$ (28)

$\Lambda_2(k+1) = \Lambda_2(k) + \alpha(k) . s_2(k)$ (29)

Where $\alpha(k) = \frac{c}{\sqrt{k}}$ $\forall k \ddagger 0$ k is the step size, where c is a constant (Polyak, 1987).

This satisfies the conditions of being positive, summation to infinity, and square summable. Repeat until *convergence* of Λ_j.

4.1.2 Algorithm for the Subgradient Method for Solving the Dual Problem

1: Initialize: k ← 0, chose α(k) step size rule $\Lambda_1(0)$, $\Lambda_2(0)$
2: while not convergence do
3: Compute the subgradients: $s_1(k) \leftarrow \nabla\Lambda_1 g = \sum j\, X_{ij} - 1\ s_2(k) \leftarrow \nabla\Lambda_2 g = \sum j\, X_{ij} - Y_l$
4: The gradient ascent step: $\Lambda_1(k+1) \leftarrow \Lambda_1(k) + \alpha(k)s_1(k)$ $\Lambda_2(k+1) \leftarrow \Lambda_2(k) + \alpha(k)s_2(k)$
5: k← k+1
6: if max($|s_1(k)|$, $|s_2(k)|$) < ϵ then
7: break
8: end if
9: Update step size: compute new α(k)
10: end while
11: Primal variables are recovered: find x^*, y^* from Λ_1^*, Λ_2^*
12:
13: Return $\Lambda_1^*, \Lambda_2^*, x^*, y^*$

4.2 Numerical Simulation Results

The numerical simulation results for the performance metrics in the distribution of Moderna for 1st dose and Oxford Astrazenica for 2nd dose of COVID-19 vaccine across the vaccination centers in Delta central senatorial district 8 (eigh)t local government area in Delta state, Nigeria, are shown in Table 4, with performance metrics for demand locations shown in Figure 1. To conduct computationally intensive modeling and optimization, the University s high-performance computing HPC cluster was used. Each compute node contained:

• Intel Core i5-9700K 3.60 GHz processor
• 32GB DDR4 2666 RAM
• 1 TB solid-state drive for data storage

The cluster comprised 40 identical nodes interconnected through an InfiniBand EDR 100 Gbps fabric. Workloads were scheduled on the Slurm resource manager using a message passing interface (MPI) for parallelization. Simulations were implemented in Python 3.8 using NumPy, Pandas, and Matplotlib. The Pyomo optimization modeling language, along with CBC and SCIP mixed integer programming solvers, was employed to formulate problems and obtain optimal solutions. This HPC configuration was essential for systematically exploring large solution spaces across many random problem instances within reasonable time frames. Parallelization facilitates solving industry scale cases with hundreds of nodes to achieve scalability of the benchmark algorithm.

Table4. Numerical Simulation Results for the Distribution of the Moderna (1st dose and Oxford Astrazenica (2nd dose of COVID-19 Vaccine

S/N	SL_l^0	SL_l^T	ω_l	$\sum \alpha_{lj}$
1	0.71	0.90	0.40	0.20
2	0.90	0.94	0.02	0.21
3	0.81	0.95	0.04	0.88
4	0.78	0.81	0.04	0.51
5	0.81	0.84	0.08	0.22
6	0.79	0.97	0.03	0.61
7	0.82	0.95	0.07	0.59
8	0.83	0.87	0.05	0.71
9	0.83	0.87	0.06	0.24
10	0.74	0.82	0.02	0.87

11	0.94	0.93	0.02	0.49
12	0.76	0.98	0.02	0.16
13	0.79	0.87	0.01	0.54
14	0.93	0.91	0.07	0.56
15	0.76	0.94	0.08	0.61
16	0.94	0.87	0.02	0.81
17	0.87	0.98	0.01	0.51
18	0.88	0.93	0.03	0.68
19	0.81	0.95	0.04	0.70
20	0.85	0.98	0.07	0.36
21	0.91	0.88	0.07	0.54
22	0.82	0.89	0.05	0.67
23	0.75	0.83	0.06	0.37
24	0.83	0.91	0.05	0.76
25	0.77	0.96	0.07	0.63
26	0.91	0.97	0.07	0.87
27	0.78	0.82	0.07	0.31
28	0.85	0.96	0.04	0.44
29	0.89	0.85	0.02	0.89
30	0.83	0.99	0.05	0.72
31	0.94	0.93	0.06	0.64
32	0.76	0.96	0.02	0.63
33	0.97	0.96	0.04	0.47
34	0.75	0.82	0.07	0.69
35	0.71	0.82	0.03	0.53
36	0.75	0.91	0.02	0.21
37	0.76	0.91	0.07	0.80
38	0.98	0.81	0.06	0.54
39	0.79	0.95	0.04	0.23
40	0.90	0.97	0.05	0.70

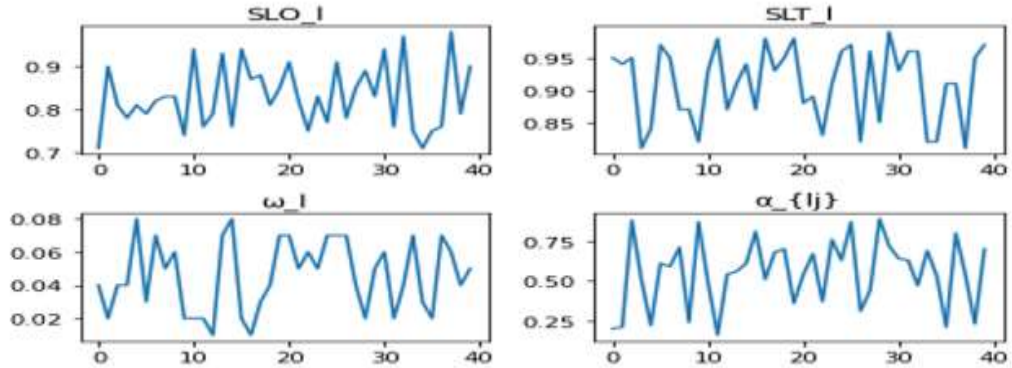

Figure 1. SL_l^0, SL_l^T, ω_l and $\sum \alpha_{lj}$ for the demand location

5. Discussion of COVID-19 Vaccine Distribution Numerical Results

The formulated model was applied in the distribution of the Johnson and Johnson Moderna for 1st doss and Oxford Astrazenica for 2nd dose across the vaccination centers in Delta central senatorial district (eight local government areas in Delta state, Nigeria, as shown in Table 4 with the performance metrics results shown in Figure. The numerical simulation results in Table 4 will be analyzed and discussed based on the performance matrices in Figure 1 as follows:

- Time Based Service Level $\left(\mathrm{SL}_l^{\mathrm{T}}\right)$Performance

The results shows that $\left(\mathrm{SL}_l^{\mathrm{T}}\right)$ values across the 40 locations are consistently high, with most exceeding 0.85 and many reaching 0.95 or higher**.** This indicates that the optimized network is highly effective in meeting the time sensitive vaccine demand. The model successfully ensures that most demand is fulfilled within the critical time window e. g., 6 hours for thawed vaccines which is a core objective of the research.

- Instantaneous Service Level (SL_l^0)Performance

The (SL_l^0) values are more varied, ranging from ~0.71 to 0.98**,** with several locations falling below 0.8. Lower SL_l^0values suggest that local stock alone is insufficient to meet all immediate demand. However, this is compensated for by the use of pipeline stock and lateral transshipments, which together boost the overall.SL_l^{T} This reflects a cost effective strategy in which local inventory is balanced with redistribution mechanisms.

- Role of Pipeline Stockω_l

The values ω_l are generally low to moderate (mostly between 0.01 and 0.08), indicating that pipeline replenishment from central depots contributes modestly to meeting time-sensitive demand. This suggest that pipeline stock is a component of the strategy, it is not the primary mechanism for achieving high service levels. This may be due to the longer lead times or capacity constraints at the central depots.

- The Role of Lateral Transshipments$\sum \alpha_{lj}$

The lateral transshipment contributions are significant and highly variable, ranging from 0.16 to 0.89. This highlights the critical role of lateral transshipments in mitigating local shortages. Facilities with low SL_l^0often have high $\sum \alpha_{lj}$showing that the model effectively redirects surplus stock from neighboring hubs to meet urgent demand. This aligns with the research objective of leveraging supply chain flexibility to enhance responsiveness. The weighted average of SL_l^{T}across all locations is consistently high, demonstrating that the system as a whole meets the time sensitive service level target. The model successfully integrates location, inventory, and transshipment decisions to achieve a resilient and responsive supply chain. The use of a queuing based inventory policy allows for dynamic adjustment to stochastic demand, which is essential during a pandemic.

6. Conclusion

The proposed multi-echelon queuing-theoretic model is validated by numerical simulation as a highly effective tool for designing resilient and responsive medical supply chains during pandemics. The model ensures that time-sensitive medical supplies (like vaccines) are delivered within critical time windows by dynamically balancing local inventory with lateral and pipeline flexibilities, thereby minimizing waste and maximizing public health impact.

7. Recommendation for the Feature Research Directions

The study recommends extending and enhancing the proposed model in several key areas as follows:

1. Model Generalization and Real-World Complexity: Incorporate more real-world factors, such as transportation reliability, road conditions, and climate impacts, into the cold chain.

2. Supply Chain Scope Flexibility: Expand from a single product vaccine to a multi- product supply chain e. g., vaccines, PPE, therapeutics and conduct a deeper cost benefit analysis of different flexibility mechanisms (e.g., lateral transshipment, higher base stock).

3. Integration of Human and Behavioral Factors: Human resource constraints (staffing and training) and workforce scheduling and behavioral factors, such as vaccine hesitancy and community acceptance, are included as endogenous variables affecting demand.
4. Advanced Methodological Approaches: Game theory is applied to model interactions between multiple stakeholders, e.g., governments and private suppliers, and multi-stage stochastic programming is also formulated to handle sequential decision making under uncertainty.
5. The findings of this study provide actionable policy insights for improving the resilience and responsiveness of medical supply chains during public health emergencies to significantly enhance the reliability of vaccine distribution systems.

References

Arora, H., et al. (2021). Modeling and simulation of a large scale COVID 19 vaccination center using queuing theory Healthcare.

Berman, O., Larson, R. C. 2001. A queueing control model for retail services with back room operations and cross trained workers. Computers Operations Research.

Green, L. V. (2006). Queuing analysis in healthcare In Patient flow: Reducing healthcare delivery delay.

Haidari, L. A., et al. (2016). Strategic location of storage facilities to improve vaccine availability in low and middle income countries.

Lee, H. L., & Haijaj, M. (2021). The vaccine supply chain: A call for resilience. Stanford Graduate School of Business.

Moosavi, J., et al., (2022). Supply chain disruption during the COVID-19 pandemic: Recognizing potential disruption management strategies

Nahmias, S. (2011). Perishable inventory systems. Springer.

Paterson, C., et al. (2011). Inventory models with lateral transshipments: a review. European Journal of Operational Research.

Shen, Z., & Qi, L. (2007). Incorporating inventory and routing costs in strategic location models. European Journal of Operational Research.

Taigbenu, E.I., Jarikre, O.E., Ofili, C.C., Dangana, E.A., Nduku, A.M., and Peter, M.B. (2025). Impact of Covid-19 on Routine Immunization and Disease Surveillance in Delta State, Nigeria: A Mixed Methods Assessment of Service Disruptions and Recovery Patterns (2018–2022). International Journal of Tropical Disease and Health, 46(6), 36-46. https://doi.org/10.9734/ijtdh/2025/v46i61665.

Yang, G., Dekker, R., Gabor, A. F., & Axsäter, S. (2013). Controlling Service Part Inventory with Lateral Transshipment and Pipeline Stock Flexibility.