DOI: https://dx.doi.org/10.4314/gjpas.v32i5.9

GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 32, 2026: 783-791
COPYRIGHT© BACHUDO SCIENCE CO. LTD PRINTED IN NIGERIA ISSN 1118 – 0579, e-ISSN: 2992 - 4464
www.globaljournalseries.com.ng, Email: globaljournalseries@gmail.com

A MATRIX-ANALYTIC QUEUEING-INVENTORY MODEL OF COMMUNITY PHARMACY SERVICE: SERVER INTERRUPTION, LOST SALES, AND OPTIMAL SAFETY STOCK IN NIGERIAN PATENT MEDICINE STORES

EKPU NDUKA VICTOR

Corresponding authors email: nduka.ekpu@unidel.edu.ng, ndukavictor23@gmail.com

(Received 28 April 2026; Revision Accepted 20 June 2026)

ABSTRACT

In Nigerian patent medicine stores and community pharmacies, a single pharmacist often doubles as dispenser and stock-keeper: when a drug's availability is uncertain, the pharmacist must step away to check the back store, halting the queue; if the drug is unavailable, the customer leaves without buying (a lost sale), while others may abandon the queue (reneging) or never join it (balking). This paper models the system as a continuous-time Markov chain coupling a finite-capacity M/M/1/N queue with state-dependent balking and reneging to a continuous-review (r,Q) inventory policy with lost sales for a high-turnover drug. Service is represented as a two-phase process (direct dispensing, or a backroom check followed by dispensing) yielding a closed-form effective service rate as a function of the probability a check is required, a proxy for stock-record and layout quality. The joint queue-inventory-service process is formulated as a quasi-birth-death chain and solved via matrix-analytic methods. A numerical study shows that raising this probability from 0.05 to 0.30 cuts the effective service rate by 37.5%, increases mean time in the pharmacy from 14.3 to 21.2 minutes, and more than doubles reneging. Counter-intuitively, the cost-minimizing safety stock falls under poor organization, since congestion, not stockouts, becomes the binding constraint, inventory and service-organization decisions must be optimized jointly, not separately.

KEYWORDS: queueing-inventory systems; matrix-analytic methods; (r,Q) policy; lost sales; balking and reneging; community pharmacy; Nigeria

INTRODUCTION

In a typical Nigerian patent medicine store or community pharmacy, customers arrive without appointment, form a queue in front of a single attendant, and expect to be served in the order they arrive. The attendant, however, performs two distinct roles that are rarely modelled separately in practice: dispensing medication that is known to be on the shelf, and, when the shelf status is uncertain, walking to the back store to verify or retrieve stock before completing the sale. This second activity halts the queue for everyone behind the customer being served. If the check reveals that the drug is genuinely out of stock, the customer leaves without a purchase (a lost sale) and, per the premise motivating this study, often leaves angry, a reputational cost beyond the immediate lost margin. Customers further back in the queue, watching this unfold, may abandon the queue before being served (reneging), and prospective customers who merely observe a long queue from the doorway may never join at all (balking).

This is, at its core, a queueing-inventory system (QIS): a service facility whose ability to complete service depends on the availability of a physical stock item, replenished according to an inventory policy. The research question motivating this paper has two parts. First, how does the pharmacist's effective service rate (and hence queue behaviour) depend on the organization of the back-store inventory, specifically the accuracy of stock counts and the ease of locating items? Second, given that relationship, what reorder point (safety stock level) for a high-turnover drug minimizes the combined cost of holding inventory,

Ekpu Nduka Victor: Department of Mathematics and Statistics, University of Delta, Agbor, Nigeria

ordering, lost sales, and customer impatience? This paper answers both questions by formulating a continuous-time Markov chain (CTMC) that couples a finite-capacity M/M/1/N queue with balking and reneging to a continuous-review (r,Q) inventory policy with lost sales, and by solving the resulting quasi-birth-death (QBD) process using matrix-analytic methods. The contribution is threefold: (i) a two-phase service representation that yields a closed-form expression linking the probability a service requires a backroom check to the pharmacist's effective service rate; (ii) an explicit finite QBD formulation of the joint queue-length/inventory/service-phase process, solved via the matrix-analytic (finite-level) method rather than simulation; and (iii) a numerical case study, calibrated to a representative Nigerian outlet, that jointly optimizes the reorder point against a cost function combining holding, ordering, lost-sale, reneging, and balking costs, and that surfaces a non-obvious interaction between service disorganization and optimal safety stock. The remainder of the paper is organised as follows: Section 2 reviews the queueing-inventory and matrix-analytic literature; Section 3 develops the model, its closed-form service-rate result, and its matrix-analytic solution; Section 4 reports a numerical case study; Section 5 discusses the findings; Section 6 draws out managerial implications; Section 7 notes limitations; and Section 8 concludes.

LITERATURE REVIEW

Queueing-inventory systems integrate two historically separate strands of applied probability (queueing theory and inventory theory) into a single Markov process in which service completion is contingent on physical stock availability. Salini, Arya, and Rangaswamy (2023) provide a comprehensive survey of the field from 2016 to 2023, cataloguing product-form solutions, perishable-inventory variants, server vacations, retrial behaviour, and multi-commodity extensions, and noting that the matrix-geometric method remains the dominant analytical technique for obtaining the stationary joint distribution of queue length and on-hand inventory.

Chakravarthy's (2022a, 2022b) two-volume treatment of matrix-analytic methods in queues provides the modern methodological foundation for both the infinite-level matrix-geometric recursion and its finite-level (truncated) analogue used in this paper, while Dudin, Klimenok, and Vishnevsky (2020) develop the general theory of matrix-analytic methods for queueing systems with correlated arrival processes.

Several recent queueing-inventory studies are methodologically close to the present model. Melikov, Shahmaliyev, and Nair (2021) analyse a single-server queueing system with perishable inventory using the matrix-geometric method, deriving the joint stationary distribution of queue length and stock level under a continuous-review replenishment policy. Melikov, Poladova, Edayapurath, and Sztrik (2023) extend matrix-geometric analysis to a single-server queueing-inventory system subject to catastrophes in the warehouse, formulating the system as a multi-dimensional CTMC and establishing ergodicity conditions before solving for the stationary distribution numerically.

Zhang, Yue, and Yue (2022) study a queueing-inventory system with a random order-size replenishment policy and lost sales, deriving the stationary joint distribution of queue length, on-hand inventory, and server status in explicit product form. Zhang, Yue, Sun, Zuo, and Yang (2022) analyse a queueing-inventory system with impatient customers directly, obtaining the matrix-geometric solution via a truncated-approximation method (the same finite-truncation logic this paper's numerical solution relies on) and using it to compute an optimal replenishment policy under a cost criterion, closely paralleling the safety-stock optimization performed in Section 4.

Otten and Daduna (2022) examine the stability conditions of networked queueing-inventory systems under different service priorities, underscoring that stability and performance in these systems cannot generally be assessed by treating the queueing and inventory subsystems independently, a finding this paper's numerical results reinforce at the level of a single outlet.

On the queueing side specifically, Kuaban, Kumar, Soodan, and Czekalski (2020) develop a finite-capacity multi-server queueing model with balking and correlated reneging explicitly motivated by health-care facility management, demonstrating that customer impatience behaviour materially affects the rate at which a health-facility queue loses patients, the same phenomenon this paper models for a single-server pharmacy queue via independent balking and reneging.

Despite this active literature, the authors are not aware of a published queueing-inventory study that addresses this specific configuration, a Nigerian patent medicine store or community pharmacy in which the server's own search behaviour, rather than a scheduled vacation or an external catastrophe, is the mechanism coupling queue and inventory dynamics. This gap is practically significant: recent Nigerian health-systems research documents that stock availability at patent and proprietary medicine vendors (PPMVs) is a persistent operational concern. Akuiyibo, Anyanti, Amoo, Aizobu, and Idogho (2022), surveying PPMVs across two Nigerian states, found that a majority of vendors experiencing a stockout in the preceding month reported it lasting less than a week, indicating that stockouts are a frequent, recurring feature of this retail channel rather than a rare event. Adepoju, Adelekan, and Oladimeji (2023) similarly document uneven stocking practices for essential medications among community pharmacists and PPMVs across two Nigerian states, with stocking rates varying substantially by drug and by respondent training background. Neither study models the resulting customer-facing queueing consequences; this paper supplies that missing analytical link.

METHODOLOGY

System Description

Consider a single-attendant patent medicine store or community pharmacy stocking one high-turnover drug of particular interest (for concreteness, an analgesic or antimalarial in frequent demand). Customers arrive singly; if the system is not full, an arriving customer may still balk (leave without joining) with a probability that increases in the number already present, reflecting a visual assessment of queue length. Once a customer begins service, the pharmacist either dispenses the drug directly (if its availability is not in doubt) or must first walk to the back store to check, before either dispensing (if stock is found) or turning the customer away (if stock is genuinely exhausted, a lost sale). Customers waiting in queue (but not the one currently in service) renege (abandon the queue) at a constant per-customer rate reflecting impatience. The drug's on-hand inventory is managed under a continuous-review (r,Q) policy: whenever a sale reduces on-hand stock to the reorder point r and no order is already outstanding, a fixed quantity Q is ordered; the order arrives after an exponentially distributed lead time and is added in full to whatever stock remains at that moment.

MODEL ASSUMPTIONS

(1) Potential customer arrivals form a Poisson process; balking is state-dependent, with the effective arrival rate at queue length n given by a non-increasing function of n that reaches zero at the system's finite capacity N, so N is enforced endogenously through balking rather than by a separate hard blocking rule.
(2) Each waiting customer (excluding the one in service) reneges independently at a constant rate; the total reneging rate at queue length n is therefore proportional to (n − 1).
(3) Service is single-phase (direct dispensing) with probability (1 − a), or two-phase (a backroom check followed by dispensing) with probability a, where a is interpreted as the probability that the pharmacist's confidence in the shelf record is insufficient to dispense without verification, a proxy for inventory-record and back-store layout quality. If the shelf is genuinely empty (on-hand stock is zero), a check is always required regardless of a.
(4) Dispensing times, check times, reneging times, and replenishment lead times are independent and exponentially distributed.
(5) A single high-turnover drug is tracked explicitly under a fixed-quantity (r,Q) policy with at most one order outstanding at a time; at most one unit of the drug is required per successfully dispensed customer.
(6) A check that reveals zero on-hand stock ends that customer's service as a lost sale; the customer departs and the server becomes free (or immediately begins the next customer's service, if any remain).

STATE SPACE AND NOTATION

The system state at any time is the quadruple (n, s, q, j), where:

- $n \in \{0, \dots, N\}$ - the number of customers in the system
- $s \in \{0, \dots, S_{cap}\}$ - the on-hand inventory of the focal drug
- $q \in \{0, 1\}$ - indicates whether a replenishment order is currently outstanding
- j - indexes the server's phase: $j = 0$ when the system is empty (server idle), $j = 1$ when the server is dispensing directly, and $j = 2$ when the server is performing a backroom check

The state $s = 0$with $j = 1$ is excluded from the state space, since a genuinely empty shelf always forces a check (Assumption 3).
The parameters governing transitions are:

Symbol	Meaning
λ	potential arrival rate
$\lambda(n)$	balking-adjusted arrival rate at level nn
μ	dispensing rate
ξ	check rate
ν	per-customer reneging rate
γ	replenishment rate (reciprocal of the mean lead time)
r	reorder point
Q	fixed order quantity

TRANSITION STRUCTURE

Grouping states by the queue-length level nn produces a level-dependent quasi-birth-death (QBD) structure: transitions from level nn go only to levels $n-1$, n, or $n+1$, with the transition rates within and between levels depending on the phase variables (s,q,j) but not on nn itself beyond the level-dependent balking rate $\lambda(n)$. Concretely, from a state (n,s,q,j):

- an **arrival** (rate $\lambda(n)$, for $n<N$) moves the system to level $n+1$, preserving (s,q,j) if $n\geq 1$, or initiating service (and hence assigning a fresh phase j'j' according to Assumption 3) if $n=0$;
- a **reneging event** (total rate $(n-1)\nu$, for $n\geq 2$) moves the system to level $n-1$, preserving (s,q,j);
- a **dispensing completion** (rate μ, from phase $j=1$) moves the system to level $n-1$, decrements ss by one unit, updates qq if the reorder point is reached, and assigns the next customer's phase (if $n-1\geq 1$) according to the new stock level;
- a **check completion** (rate ξ, from phase $j=2$) either transitions to phase $j=1$ at the same level (if stock is found, $s>0$) or moves to level $n-1$ as a lost sale with ss unchanged (if $s=0$);
- a **replenishment arrival** (rate γ, whenever $q=1$) adds Q units to s, capped at the state-space ceiling, and resets q to 0, independently of n and j.

This transition structure is summarised in block form as a generator matrix with block-tridiagonal structure across levels $n=0,\dots,N$:

$$Q_{gen}=\begin{bmatrix} B_0 & A_0 & 0 & \cdots & 0 \\ A_2 & A_1 & A_0 & \cdots & 0 \\ 0 & A_2 & A_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & A_0 \\ 0 & 0 & 0 & A_2 & B_N \end{bmatrix}$$

where each block A_0, A_1, A_2 (and the boundary blocks B_0, B_N) is a square matrix over the finite phase space $\{(s,q,j)\}$, with A_0 collecting arrival transitions (level $n\rightarrow n+1$), A_2 collecting reneging, dispensing-completion, and lost-sale transitions (level $n\rightarrow n-1$ and A_1 collecting within-level transitions (check-completion-to-dispensing, replenishment arrivals) together with the diagonal holding-time rates. This is precisely the block structure to which matrix-analytic methods apply (Chakravarthy, 2022a; Dudin et al., 2020; Salini et al., 2023).

CLOSED-FORM EFFECTIVE SERVICE RATE

Proposition 1 (Effective service rate under inventory disorganization). Consider a single service in isolation, under the idealized condition that the item is in stock at the moment service begins ($s>0$). By Assumption 3, the service time T is a phase-type random variable: with probability $(1-a)$, $T\sim$ Exponential(μ)); with probability a, T is the sum of two independent exponential phases, Exponential(ξ) followed by Exponential(μ). The effective service rate, defined as the reciprocal of the mean service time, is

$$\mu_{eff}(a)=\frac{1}{E[T]}=\frac{\mu\xi}{\xi+a\mu} \qquad (1)$$

Proof. By the law of total expectation,

$$E[T]=(1-a)\cdot\frac{1}{\mu}+a\cdot\left(\frac{1}{\xi}+\frac{1}{\mu}\right)=\frac{1}{\mu}+\frac{a}{\xi}=\frac{\xi+a\mu}{\mu\xi}$$

Inverting gives (1).

Equation (1) is strictly decreasing in a

$$\left(\frac{\partial\mu_{eff}}{\partial a}=\frac{-\mu^2\xi}{(\xi+a\mu)^2}<0\right)$$

and strictly increasing in ξ

$$\left(\frac{\partial\mu_{eff}}{\partial\xi}=\frac{\mu^2 a}{(\xi+a\mu)^2}>0 \text{ for } a>0\right)$$

confirming the intuitive relationship the research question anticipates: service degrades as the probability of needing a stock check rises, and this degradation is mitigated by a faster check process (better back-store layout, clearer labelling, a more searchable stockroom).

Because Proposition 1 is derived under the idealized condition $s>0$, it characterises the service-rate cost of inventory disorganization holding stockouts aside; the full CTMC of Section 3.4 additionally captures the compounding effect of genuine stockouts ($s=0$), which force a check regardless of aa and can end service in a lost sale rather than a completed dispensation.

Matrix-Analytic Steady-State Solution

Because the system has finite capacity NN and finite inventory ceiling S_{cap}Scap, the joint process (n, s, q, j) is a finite, irreducible CTMC (for any parameter values admitting a viable (r, Q) policy, see the remark below), and its stationary distribution ππ exists uniquely and satisfies the global balance equations

$$\pi Q_{gen} = 0,\ \pi\mathbb{1} = 1$$

For the block-tridiagonal structure of Section 3.4, this system is solved by the finite-level matrix-analytic (folding / linear level-reduction) method: writing $\pi = (\pi_0, \pi_1, \dots, \pi_N)$ with each π_n a row vector over the phase space at level nn, the balance equations reduce to a linear recursion between adjacent levels,

$$\pi_{n+1} = \pi_n R_n$$

where the level-dependent matrices R_n are obtained by backward elimination from the boundary level $n = N$ down to $n = 0$ (Chakravarthy, 2022a, 2022b). This is the finite-N analogue of the classical matrix-geometric solution $\pi_n = \pi_0 R^n$ that applies when the queue has unbounded capacity and the QBD is level-independent (Neuts-type); the finite case requires no separate stability condition on the spectral radius of RR, since a finite irreducible chain is automatically positive recurrent, but it sacrifices the homogeneous geometric form in exchange for exact level-by-level matrices. Numerically, this recursion is equivalent to direct sparse solution of the full balance-equation system, the approach adopted in Section 4, and is consistent with the truncated-approximation matrix-geometric methodology used for related queueing-inventory models with impatient customers (Zhang, Yue, Sun, Zuo, & Yang, 2022).

Remark (viability of the (r, Q) policy). The recursion above presumes the chain is irreducible. This requires $Q > r$: if the order quantity does not exceed the reorder point, a replenishment arriving after the on-hand stock has been driven to zero by continued demand during the lead time can land at or below rr again, and the reorder trigger (defined on the downward crossing of r) is never re-armed, trapping the system at zero stock permanently. This is not a numerical artefact but a structural property of the (r, Q) policy: the reorder point must leave enough of a margin below the replenishment target that a full order quantity is guaranteed to lift stock back above rr even after further depletion during the lead time. All instances solved in Section 4 satisfy $Q > r$.

3.7 Performance Measures

Given the stationary distribution π, the mean number of customers in the system is

$$L = \sum n \cdot \pi(n, s, q, j) \text{ (summed over all states)}$$

and by Little's law the mean time a customer spends in the system is

$$W = \frac{L}{\lambda_{eff}}$$

where $\lambda_{eff} = \lambda -$ (balking rate) is the effective admitted arrival rate.

- **Lost-sale rate**: $\xi \cdot P(j, =, 2\ s = 0\ n \geq 1)$
- **Reneging rate**: $\sum (n-1)\nu \cdot \pi(n, s, q, j)$
- **Balking rate**: $\sum_n [\lambda - \lambda(n)] \cdot P(n)$
- **Replenishment (order) rate**: $\gamma \cdot P(q = 1)$, which by flow balance on the inventory process equals (sales rate)$/Q$(sales rate)/Q in steady state.

Cost Model and the Optimal Safety Stock Problem

Combining holding, ordering, and impatience-related costs, the long-run average cost rate as a function of the reorder point rr (for fixed Q) is

$$C(r) = h \cdot E[s] + K \cdot \text{(order rate)} + c_{LS} \cdot \text{(lost-sale rate)} + c_{ren} \cdot \text{(reneging rate)} + c_{bal} \cdot \text{(balking rate)} \qquad (2)$$

where hh is the holding cost per unit of on-hand inventory per unit time, K is the fixed cost per replenishment order, and c_{LS}, c_{ren} , c_{bal} are the per-event costs of a lost sale, a reneging event, and a balking event, respectively. As in related queueing-inventory cost studies (Zhang, Yue, Sun, Zuo, & Yang, 2022), no closed-form minimizer of (2) is available because $E[s]$ and the event rates are outputs of the matrix-analytic solution rather than elementary functions of r; the optimal reorder point $r^* = \arg\min_r C(r)$ is therefore obtained by direct numerical search over the feasible range r < Q, reported in Section 4.

Numerical Case Study: A High-Turnover Drug at a Nigerian Patent Medicine Store

Parameterization

The case study is a numerical illustration calibrated to plausible operating conditions for a busy Nigerian patent medicine store, rather than a fitted empirical estimate; Table 1 lists the parameter values used throughout this section. Two scenarios are compared, differing only in the probability aa that a service requires a backroom stock check: $a = 0.05$ represents a well-organized store with an accurate, computerized or carefully maintained stock count, and $a = 0.30$ represents a poorly organized store relying on memory or an informal count that frequently requires physical verification.

Table 1: Model Parameters for the Numerical Case Study

Parameter	Symbol	Value	Interpretation
Potential arrival rate	λ	10.0 / hr	Customers approaching the store per hour
Finite system capacity	N	8	Max customers in system before balking reaches 1
Base dispensing rate	μ	12.0 / hr	Rate of direct dispensing when item is at hand
Backroom check rate	ξ	6.0 / hr	Rate of completing a stock check in the back store
Reneging rate (per waiting customer)	ν	0.5 / hr	Impatience rate while waiting, not in service
Replenishment rate	γ	0.5 / hr	1 / mean lead time (mean lead time = 2 hours)
Fixed order quantity	Q	20 units	
Disorganization probability (well-organized)	a	0.05	Scenario 1: accurate, computerized stock count
Disorganization probability (poorly organized)	a	0.30	Scenario 2: manual, error-prone stock count

Effect of Inventory Organization on Effective Service Rate

Applying Proposition 1 directly, Table 2 reports the closed-form effective service rate under each scenario. Moving from a well-organized store ($a = 0.05$) to a poorly organized one ($a = 0.30$) reduces the pharmacist's effective service rate by 37.5%, from 10.91 to 7.50 customers per hour, a substantial degradation driven entirely by the increased frequency of backroom checks, holding the check rate ξ itself unchanged. This isolates and quantifies the first part of the research question: inventory organization affects queue performance not through inventory availability alone, but directly through the time cost imposed on every service, whether or not that particular service ultimately succeeds.

Table 2: Closed-Form Effective Service Rate by Inventory-Organization Scenario

Quantity	Formula	$a = 0.05$	$a = 0.30$
Effective service rate μ_{eff}	$\frac{\mu\xi}{\xi + a\mu}$	10.91 / hr	7.50 / hr
Reduction relative to base $\mu = 12.0$	$1 - \frac{\mu_{eff}}{\mu}$	9.1%	37.5%

Optimal Safety Stock: Cost-Minimization Results

Solving the full CTMC of Sections 3.4–3.6 for each integer reorder point r from 0 to $Q - 1 = 19$ and evaluating the cost function (2) with $h = ₦95/$ unit/hour, $K = ₦2{,}000/$order, $c_{LS} = ₦1{,}200/$lost sale, $c_{ren} = ₦700/$reneging event, and $c_{bal} = ₦250/$ balking event, produces a genuine interior cost minimum in both scenarios. Table 3a and Table 3b report the region of the search around each optimum; the full twenty-point sweep is available from the authors.

Table 3a: Reorder-Point Sweep, a = 0.05 (Well-Organized Store) Optimum Marked *

r	C(r) ₦/hr	E[s]	Lost sale/hr	Reneging/hr	Balking/hr	Order rate/hr
10	4293.14	9.50	1.5196	0.5894	2.4397	0.2726
12	4233.01	10.58	1.3860	0.5740	2.3936	0.2823
14	4193.78	11.77	1.2612	0.5596	2.3505	0.2914
16	4174.61	13.05	1.1461	0.5464	2.3108	0.2998
17*	4172.09	13.71	1.0922	0.5402	2.2922	0.3038
18	4174.00	14.39	1.0407	0.5343	2.2744	0.3075
19	4180.17	15.10	0.9913	0.5286	2.2574	0.3111

Table 3b: Reorder-Point Sweep, a = 0.30 (Poorly Organized Store), Optimum Marked *

r	C(r) ₦/hr	E[s]	Lost sale/hr	Reneging/hr	Balking/hr	Order rate/hr
10	4294.69	10.65	1.2106	0.8095	3.1033	0.2438
12	4257.02	11.92	1.0735	0.8023	3.0827	0.2521
14*	4244.73	13.31	0.9486	0.7957	3.0638	0.2596
15	4247.57	14.04	0.8908	0.7927	3.0551	0.2631
16	4256.06	14.80	0.8359	0.7898	3.0468	0.2664
18	4288.92	16.37	0.7349	0.7846	3.0316	0.2724
19	4312.78	17.18	0.6885	0.7821	3.0246	0.2752

The cost-minimizing reorder point is r = *17 under good organization and r* = 14 under poor organization, safety stock is lower, not higher, when the store is poorly organized. Table 4 compares the two scenarios at their respective optima.

Table 4: System Performance at the Cost-Minimizing Reorder Point, by Scenario

Measure at r*	a = 0.05 (r*=17)	a = 0.30 (r*=14)
Minimum total cost C(r*), ₦/hr	4172.09	4244.73
Mean on-hand inventory E[s]	13.71	13.31
Mean number in system L	1.834	2.451
Mean time in system W, hours (minutes)	0.238 (14.3 min)	0.353 (21.2 min)
Lost-sale rate, events/hr	1.092	0.949
Reneging rate, events/hr	0.540	0.796
Balking rate, events/hr	2.292	3.064

MANAGERIAL INTERPRETATION

The result in Table 4 is initially counter-intuitive: a poorly organized store, which loses more effective service capacity, ends up with a lower optimal safety stock than a well-organized one. The mechanism is visible in the rest of the table.

Poor organization (a = 0.30) lowers the effective service rate (Proposition 1), which raises the mean number in the system from 1.83 to 2.45 and mean time in the pharmacy from 14.3 to 21.2 minutes; this congestion, not inventory scarcity, becomes the dominant source of customer loss, reflected in the reneging rate nearly doubling (0.540 to 0.796 events per hour) and the balking rate rising by a third (2.292 to 3.064 events per hour). Because additional safety stock does nothing to relieve a bottleneck whose root cause is service time rather than stock availability, the marginal holding cost of extra inventory in the poorly organized scenario is no longer justified by a commensurate reduction in lost sales, the lost-sale rate at the poorly organized optimum (0.949 per hour) is in fact already lower than at the well-organized optimum (1.092 per hour), precisely because fewer customers are surviving the queue long enough to reach the point of a stock check at all. The substantive managerial lesson is that safety stock and service organization are not independent levers: a manager who observes frequent lost sales and angry customers cannot assume more inventory is the correct response without first establishing whether the underlying bottleneck is stock availability or service-time disorganization, since the two call for different, and in this case partially offsetting, remedies.

RESULTS AND DISCUSSION

Three findings emerge from the coupled model. First, the closed-form result of Proposition 1 demonstrates that inventory-record and layout quality has a direct, quantifiable effect on service capacity independent of any stockout event, a 25-percentage-point increase in the probability of needing a backroom check (from $a = 0.05$ to $a = 0.30$) costs the store more than a third of its effective service rate. Second, the matrix-analytic solution of the full joint queue-inventory chain shows that this service-rate degradation propagates into materially worse queueing outcomes: longer waits, more reneging, more balking. Third, and most substantively, the joint optimization of the reorder point against a cost function spanning both inventory and queueing costs reveals that the two subsystems interact in ways a queueing-only or inventory-only analysis would miss entirely, the optimal safety stock response to service disorganization is a reduction, not an increase, because the source of customer loss has shifted from the inventory side to the queueing side.

These findings are consistent with the broader queueing-inventory literature's emphasis on joint rather than decomposed analysis: Otten and Daduna (2022) reach an analogous conclusion for network stability in queueing-inventory systems with service priorities, and the cost-minimization approach and its reliance on numerical rather than closed-form optimization mirrors Zhang, Yue, Sun, Zuo, and Yang (2022). The specific direction of the safety-stock result in Section 4.4, however, appears not to have been previously reported for a lost-sales (r,Q) system coupled to a server whose own search behaviour (rather than an exogenous vacation or catastrophe process) drives the queue-inventory coupling.

IMPLICATIONS FOR PRACTICE AND POLICY

For individual patent medicine store and community pharmacy operators, the model offers a concrete decision-support tool that goes beyond the common assumption that customer dissatisfaction is primarily an inventory problem to be solved by carrying more stock. Where stockout-driven reports of stocking gaps are already documented among Nigerian PPMVs and community pharmacists (Akuiyibo et al., 2022; Adepoju et al., 2023), this paper's results suggest that back-store organization (accurate, quickly searchable stock records) may deliver a larger service-quality return than additional safety stock alone, particularly once queueing congestion, rather than inventory scarcity, becomes the binding constraint. For bodies such as the Pharmacists Council of Nigeria (PCN) and the National Association of Patent and Proprietary Medicine Dealers (NAPPMED), the model's explicit separation of the service-organization effect (Proposition 1) from the inventory-policy effect (Section 4.3) offers a template for structuring training or accreditation guidance around both levers rather than treating stock availability as the sole determinant of service quality.

LIMITATIONS AND FUTURE RESEARCH

The numerical case study is an illustration calibrated to plausible operating parameters rather than a fitted estimate from field data collected at a specific outlet; a natural extension, following the primary-data collection approach used elsewhere in this line of work, would estimate λ, μ, ξ, ν, and the lead-time distribution directly from time-and-motion observation at one or more Nigerian outlets. The model assumes exponential service, check, reneging, and lead-time distributions throughout; while this is standard in the matrix-analytic queueing-inventory literature and yields the tractable CTMC structure used here, service and lead times in practice may be better represented by phase-type or more general distributions, which the matrix-analytic framework (Chakravarthy, 2022a, 2022b) can in principle accommodate at the cost of a larger phase space. The model tracks a single high-turnover drug; a genuinely multi-item pharmacy faces correlated stockout risk across many drugs sharing the same server and queue, which would require a substantially larger joint state space. Finally, the cost parameters in Section 4.3 are illustrative; a decision-ready application would require estimates of the actual holding, ordering, and lost-sale costs specific to a given outlet, and the qualitative safety-stock finding of Section 4.4 (that poor service organization can lower rather than raise the optimal reorder point) should be re-examined once real cost data are available, since the result is sensitive to the relative magnitudes of holding versus impatience-related costs, as the sensitivity analysis in Section 4.3 (varying h) demonstrated during model development.

CONCLUSION

This paper formulated a continuous-time Markov chain coupling a finite-capacity M/M/1/N queue with balking and reneging to a continuous-review (r,Q) inventory policy with lost sales, representing a single-attendant Nigerian patent medicine store or community pharmacy in which the server's own stock-checking behaviour couples queue and inventory dynamics. A closed-form result showed that the pharmacist's effective service rate declines sharply as the probability of needing a backroom stock check rises, independent of any stockout event, while the full joint chain (solved via the matrix-analytic method for finite quasi-birth-death processes) showed that this decline propagates into materially longer waits and higher reneging and balking rates. Jointly optimizing the reorder point against a cost function spanning holding, ordering, lost-sale, reneging, and balking costs revealed that the two research questions posed at the outset are not separable: the optimal response to poor inventory organization is not simply more safety stock, since the resulting congestion, rather than stock scarcity, becomes the binding constraint on customer retention. The results support treating inventory organization and safety-stock policy as jointly optimized decisions in queueing-inventory systems generally, and offer a quantitative starting point for improving service delivery in Nigeria's patent medicine and community pharmacy sector specifically.

REFERENCES

Adepoju, V. A., Adelekan, A., and Oladimeji, O., 2023. Stocking practices of anti-tuberculosis medications among community pharmacists and patent proprietary medicine vendors in two states in Nigeria. Healthcare, 11(4), 584. https://doi.org/10.3390/healthcare11040584

Akuiyibo, S. M., Anyanti, J., Amoo, B. A., Aizobu, D., and Idogho, O., 2022. Common childhood illnesses: A cross-sectional study of commodity stocking patterns and management knowledge among patent and proprietary medicine vendors in Nigeria. Archives of Public Health, 80(1), 88. https://doi.org/10.1186/s13690-022-00846-x

Chakravarthy, S. R., 2022a. Introduction to matrix-analytic methods in queues 1: Analytical and simulation approach—basics. John Wiley and Sons.

Chakravarthy, S. R., 2022b. Introduction to matrix-analytic methods in queues 2: Analytical and simulation approach—queues and simulation. John Wiley and Sons.

Dudin, A. N., Klimenok, V. I., and Vishnevsky, V. M., 2020. The theory of queueing systems with correlated flows. Springer Nature Switzerland.

Kuaban, G. S., Kumar, R., Soodan, B. S., and Czekalski, P., 2020. A multi-server queuing model with balking and correlated reneging with application in health care management. IEEE Access, 8, 169623–169639. https://doi.org/10.1109/ACCESS.2020.3021918

Melikov, A., Poladova, L., Edayapurath, S., and Sztrik, J., 2023. Single-server queueing-inventory systems with negative customers and catastrophes in the warehouse. Mathematics, 11, 2380.

Melikov, A. Z., Shahmaliyev, M. O., and Nair, S. S., 2021. Matrix-geometric method for the analysis of a queuing system with perishable inventory. Automation and Remote Control, 82(12), 2169–2182. https://doi.org/10.1134/S0005117921120080

Otten, S., and Daduna, H., 2022. Stability of queueing-inventory systems with different priorities. Annals of Operations Research. Advance online publication. https://arxiv.org/abs/2209.08957

Salini, K., Arya, P. S., and Rangaswamy, M., 2023. Queueing-inventory systems: A survey (arXiv:2308.06518). arXiv. https://arxiv.org/abs/2308.06518

Zhang, Y., Yue, D., and Yue, W., 2022. A queueing-inventory system with random order size policy and server vacations. Annals of Operations Research, 310(2), 595–620. https://doi.org/10.1007/s10479-020-03859-3

Zhang, Y., Yue, D., Sun, L., Zuo, J., and Yang, B., 2022. Analysis of the queueing-inventory system with impatient customers and mixed sales. Discrete Dynamics in Nature and Society, 2022, Article 2333965. https://doi.org/10.1155/2022/2333965