Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Balancing shortage risk and equity in healthcare allocation under displacement-driven demand uncertainty in Nigeria A multistage stochastic optimisation study

Chinedu Bright Okechukwu

cbokechukwu@waec.org.ng

West African Examinations Council

Bright Okore Osu

Abia State University

Mark Laisin

Chukwuemeka Odumegwu Ojukwu University

Research Article

Keywords: healthcare resource allocation, displacement, stochastic programming, health equity, conditional value-at-risk, maternal and newborn health, service adequacy, Nigeria

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11580813/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Balancing shortage risk and equity in healthcare allocation under displacement-driven demand uncertainty in Nigeria

A multistage stochastic optimisation study

Chinedu Bright Okechukwu[1,2,*], **Bright Okore Osu**[2], **Mark Laisin**[3]

[1] The West African Examinations Council, Kaduna Zonal Office, P.M.B. 2135, Kaduna, Kaduna State, Nigeria.

[2] Department of Mathematics, Abia State University, Uturu, Abia State, Nigeria.

[3] Department of Mathematics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria.

* Corresponding author: Chinedu Bright Okechukwu. Email: cbokechukwu@waec.org.ng. Telephone: +2348032562420.

Co-author emails: Osu.bright@abiastateuniversity.edu.ng (Bright Okore Osu); laisinmark@gmail.com (Mark Laisin).

ORCID: Chinedu Bright Okechukwu — 0000-0003-4286-8193; Bright Okore Osu — 0000-0003-2463-430X.

Abstract

Background

Displacement can alter healthcare requirements before resources can be procured and deployed. Health planners must balance aggregate service, severe shortages and geographic priorities. We examined whether protecting adverse system losses and extreme-demand conditions also protects primary healthcare and maternal and newborn health provision.

Methods

We formulated a three-stage stochastic linear programme for 177 local government areas and 348 supply hubs in ten Nigerian states. Four displacement-derived archetypes and three severity levels generated twelve planning scenarios. Matched comparisons evaluated expected-loss minimisation, conditional value-at-risk, geographic priority penalties and a shortage cap in a preselected extreme-demand scenario. We

assessed resource-specific coverage, geographic outcomes, shortage feasibility and sensitivity to resource conditions and priority construction. A separate extension imposed minimum maternal and newborn health coverage in every scenario.

Results

At the assumed probabilities, 95% conditional value-at-risk equalled maximum scenario loss. Risk aversion produced small loss reductions. Compared with expected-loss minimisation, the integrated policy reduced the penalty-weighted priority increment by 4.087% and pooled priority-weighted shortage by 0.168%, while expected base loss increased by 0.033%. Primary healthcare coverage remained 37.743%, but maternal and newborn health coverage declined from 70.959% to 70.309%; geographic effects were mixed. Minimum protected shortage was only 20.266 pooled units below the reference cap. Secondary expected-loss minimisation approached this minimum while preserving most non-protected service. The largest feasible common maternal and newborn health coverage floors were 62.716% at baseline capacity and 64.907% at 120% capacity with unchanged budgets. At expanded capacity, feasible positive floors prevented zero maternal and newborn health allocation, at higher expected loss and lower primary healthcare coverage.

Conclusions

Loss protection and geographic priorities do not guarantee resource-specific service adequacy. Planners should specify protected services, test target feasibility and examine geographic and resource-specific trade-offs together. These findings concern a calibrated planning system; they do not demonstrate observed utilisation gains or clinical benefit.

Keywords: healthcare resource allocation; displacement; stochastic programming; health equity; conditional value-at-risk; maternal and newborn health; service adequacy; Nigeria.

Background

Planning problem and study setting

State health authorities and humanitarian health planners must decide how to procure and distribute scarce resources when displacement changes both the location and magnitude of need. Commitments made before demand becomes clear can constrain later responses. The practical question is which services and locations a plan protects when available resources cannot meet all requirements. A low expected system loss may coexist with severe shortages in adverse conditions or insufficient provision in a particular service. We examine these competing objectives for primary healthcare and maternal and newborn health in displacement-affected areas of Nigeria.

Multistage stochastic programming links early commitments to decisions that can adapt as information becomes available. Healthcare applications demonstrate how adaptive allocation can be combined with explicitly defined equity considerations [1]. However, the value of a risk or distributional control depends on the uncertainty representation, the actions permitted and the resources available. Adding a preference weight or an adequacy constraint cannot itself create capacity, stock or funding. Its effects must be assessed against a common physical system and a clearly specified benchmark.

Health-service allocation studies also demonstrate the importance of making competing outcomes explicit. Bowers et al. [2] developed a multicriteria postnatal service allocation tool through stakeholder engagement, supporting comparisons of costs and quality. Wang et al. [3] combined accessibility assessment and mathematical optimisation to examine healthcare equity in an ageing community. These studies connect allocation criteria to service-management questions. Our application examines a different setting: uncertain displacement-related requirements and the consequences of protecting losses, geographic priorities or specific services. It evaluates modelled allocation trade-offs rather than an implemented service redesign.

This study considers 177 local government areas (LGAs) and 348 supply hubs in ten North-Central and North-West Nigerian states. Primary healthcare (PHC) and maternal and newborn health (MNH) have separate service requirements and resource constraints. Four displacement-derived spatial archetypes and three severity levels define twelve terminal demand scenarios. Initial procurement and allocation decisions

are common across scenarios; subsequent decisions respond first to the archetype and then to severity. The three information stages are distinguished from the three physical allocation periods. Demand is balanced at the planning horizon, without intermediate clinical service deadlines.

Loss protection geographic priority and demand adequacy

Three distinct planning concerns motivate the model. Loss-tail protection addresses large realised system losses. Geographic priority changes the marginal penalty attached to shortages using displacement share, the age composition of displaced people and an access deficit. Demand-tail adequacy limits shortage under demand conditions designated as extreme before optimisation. These controls protect different outcomes and need not select the same allocation. A policy can improve its weighted objective while worsening resource-specific coverage or a geographic distributional indicator.

Conditional value-at-risk (CVaR) evaluates the upper tail of a policy-dependent loss distribution [4]. By contrast, the conditional shortage cap considered here uses a protected demand set fixed independently of allocation decisions. High-loss scenarios can change with the policy, whereas the protected demand set remains unchanged. The shortage cap is an adequacy restriction on specified demand conditions, not a risk measure interchangeable with CVaR. The primary protected set contains R18 HIGH, scenario S12; its conditional expected shortage therefore equals shortage in that scenario.

The finite probability distribution also determines what loss-tail protection means. Every terminal scenario has probability at least 0.05, so both 95% and 99% CVaR coincide with maximum scenario loss. These confidence levels describe the assumed planning distribution and do not estimate the occurrence frequency of extreme displacement. Comparisons of policies optimised at different confidence levels therefore use a common evaluation criterion. This prevents a change in the definition of the reported tail statistic from being mistaken for improved policy performance.

Equity likewise requires a declared criterion. The efficiency implications of fairness depend on how fairness is defined [5]. Geographic-priority penalties express priorities, but do not guarantee equal coverage, higher minimum coverage or lower geographic dispersion. The analysis accordingly distinguishes the penalty-weighted priority increment in the objective from the priority-weighted shortage

diagnostic, PHC and MNH system coverage, and LGA-level coverage outcomes. Pooled shortage measures combine unlike resource units and are treated as model indices rather than patient counts.

Research gap and questions

The health-service planning gap addressed here is whether a policy that improves a weighted allocation score also protects the services and locations that planners intend to prioritise. Loss control, geographic priority and extreme-demand safeguards may select different allocations under the same constraints. Their effects must therefore be compared separately, including the scale of any benefit and the service displaced elsewhere. A shortage target also requires a feasibility assessment: a tighter bound may exceed what the system can deliver. Finally, an objective that protects only one adverse demand state can leave other service outcomes unspecified, making the choice of a secondary objective consequential.

The study asks four questions. First, how much does CVaR-based loss protection change expected base loss, worst loss and resource coverage relative to risk-neutral allocation? Second, how does geographic priority redistribute shortages, and do weighted gains coincide with improvements in resource-specific and geographic outcomes? Third, when does the fixed demand-tail cap add a restriction, how close is it to the attainable shortage frontier, and what allocation results from secondary expected-loss minimisation near that frontier? Fourth, how do budget, inventory and capacity conditions affect service, and what additional protection and trade-offs arise when a minimum MNH coverage requirement is explicitly imposed?

Contributions and scope of the evidence

The contribution is an empirical comparison of established controls within a common three-stage planning system. Risk-neutral, CVaR-only, tail-only and zero-priority hybrid policies distinguish loss protection from demand adequacy. Matched priority-weighted policies identify the additional effect of imposing the cap. A separate optimisation establishes the minimum protected shortage, followed by secondary expected-loss minimisation within a stated numerical tolerance. Resource-specific and LGA diagnostics then test whether improvements in an objective or pooled diagnostic translate into broader service benefits.

The results show small worst-loss reductions from risk aversion and no additional aggregate benefit from the reference cap at zero priority weight. With priority weighting, the integrated policy lowers pooled priority-weighted shortage but reduces MNH coverage and produces mixed geographic effects. The cap

lies close to the attainable shortage frontier; secondary optimisation approaches that frontier while preserving service close to the risk-neutral benchmark. Budget, stock and capacity experiments produce larger resource-specific changes. An explicit MNH floor prevents a zero-service outcome when feasible, while revealing additional cost and PHC trade-offs. These are conditional findings rather than universal superiority of an integrated policy.

The analysis concerns calibrated planning demand, effective capacities, cost indices and assumed scenario probabilities. It does not estimate clinical utilisation, validate displacement forecasts or demonstrate clinically timely treatment. Accessibility is represented by distance and effective-speed proxies. Initial stock is an endowment whose pre-horizon acquisition cost is excluded, so the inventory comparison is not an all-in investment appraisal. The paper builds on the scenario-construction framework of Okechukwu and Osu [6] and the expected-loss allocation component of the wider research programme [7, 8], while its distinct focus is the joint interpretation of loss protection, geographic priority and demand-tail feasibility.

Related research and analytical positioning

The analytical literature informs three health-service decisions: when resources may adapt to new information, which adverse outcomes should be protected, and how shortages should be distributed. These decisions require distinct criteria. We review the relevant approaches to establish what the present comparisons can and cannot show.

Adaptive healthcare allocation under uncertainty

Yin and Büyüktahtakın [1] formulate multistage stochastic epidemic resource allocation with equity considerations, linking decisions to the progressive revelation of uncertainty. Their work provides a foundation for combining adaptive allocation with distributive requirements. Fattahi et al. [9] extend the healthcare planning perspective through integrated resource allocation and sharing, patient transfers and capacity decisions during a pandemic. Their multistage framework considers several patient and resource categories and examines a risk-averse version based on non-accepted patients. These studies show that the operational actions available to the planner are central to the meaning of an allocation model.

The present application uses displacement-derived PHC and MNH requirements as exogenous inputs. It does not model infection transmission, patient transfers, workforce substitution or clinical sequencing. Its

scenario tree represents information about spatial archetypes and severity, while the physical allocation periods govern procurement and deployment. This distinction matters because additional information has value only when feasible actions can respond to it. The three-stage structure supplies a common decision framework for the comparisons; the contribution of this paper is the assessment of risk, priority and adequacy controls within that framework.

Loss tail protection and the meaning of multistage risk

Rockafellar and Uryasev [4] establish the optimisation representation of CVaR for general loss distributions, including distributions with probability atoms. CVaR therefore provides a tractable way to penalise adverse losses in finite scenario programmes. The protected outcome must nevertheless be specified. Fattahi et al. [9] apply risk aversion to non-accepted patients; in the present model, CVaR applies to base scenario loss comprising operating expenditure and shortage penalties. A reduction in this cost index cannot automatically be interpreted as the same proportional reduction in unmet service or clinical harm.

The timing of risk evaluation also deserves attention. In their capacity-planning study, Yu and Shen [10] compare risk-averse multistage and two-stage formulations using expected conditional risk measures and derive bounds on their objective-value differences. Their analysis demonstrates that the value of adaptation depends on the risk formulation and the uncertainty structure. Here, decisions are multistage, but CVaR is evaluated once over total scenario losses. The model does not implement a recursively nested dynamic risk measure, and the presence of non-anticipativity constraints alone is not a claim of dynamic consistency for risk preferences.

Finite probability masses further constrain the interpretation of confidence levels. In this study every terminal scenario has probability at least 0.05; consequently, both 95% and 99% CVaR equal maximum scenario loss, as established in Proposition 1. Identical outcomes at these levels can therefore follow from the distribution itself. This property should be distinguished from a general conclusion that greater risk aversion never matters. The experiments separate the risk weight from the confidence level and compare selected policies using a common 95% diagnostic.

Distributional ambiguity and the boundary of scenario evidence

Recent healthcare research extends risk-sensitive planning beyond a single assumed probability distribution. Heyhat, Rahmani and Shahparvari [11] develop two-stage distributionally robust healthcare logistics models with ambiguity sets for demand probabilities and CVaR-based risk control. Their application includes alternate care sites, heterogeneous patient requirements, several resource categories, and workforce cross-training and mobility. Their out-of-sample evaluation addresses whether policies retain performance when demand conditions differ from the fitted planning distribution.

This is an important boundary for the present study. Its twelve scenarios and their probabilities define a fixed experimental distribution; budget and initial-stock comparisons do not test probability ambiguity or out-of-sample demand performance. The results therefore establish conditional policy trade-offs for the specified calibration. Distributionally robust extensions would require explicit ambiguity-set construction and separate validation. They cannot be inferred merely from the use of empirical archetypes or from the inclusion of CVaR. Likewise, changing preference weights and expanding physical resources answer different questions and are evaluated separately.

Equity vulnerability and resource specific outcomes

Bertsimas, Farias and Trichakis [5] analyse the efficiency implications of defined fairness rules, including proportional and max-min fairness. Their central relevance here is that the price of fairness depends on the criterion imposed. A priority coefficient attached to shortage expresses a distributive preference, whereas a minimum-coverage requirement protects a service floor. These devices can select different allocations even when applied to the same feasible set. Their outcomes should not be described by a common equity label without identifying the underlying rule.

Huang et al. [12] make this distinction concrete in a two-stage stochastic emergency-supplies allocation model. They combine operating cost with a Gini-inspired distributional index and use an epsilon-constraint approach to examine the objectives, with an application to mask distribution in Shanghai. Their formulation directly represents distributional disparity. The present model instead modifies shortage penalties using calibrated geographic priorities. It does not minimise a Gini index or impose max-min

service. Its geographic diagnostics assess the consequences of that priority rule without claiming that the rule itself guarantees lower inequality.

Dik, Lo and Lai [13] review fairness, equity and equality in humanitarian logistics through a citation-network analysis of 275 articles published between 2007 and 2025. They identify a growing connection between deprivation-cost and distributional-preference approaches. This broader literature reinforces the need to state whose outcomes are compared and what notion of disadvantage is protected. In the present application, LGA coverage dispersion, minimum coverage and the number of low-coverage LGAs are reported alongside demand-weighted system coverage. PHC and MNH are assessed separately because pooled improvements can conceal deterioration in one resource.

The distinction also applies within model accounting. The objective priority increment includes resource-specific shortage penalties, while the priority-weighted shortage diagnostic uses the full geographic weight without those penalties. Neither quantity is an inequality index. Their separate definitions prevent an improvement in one weighted total from being reported as a demonstrated gain in every distributive dimension. The coefficients express normative planning priorities rather than estimated social welfare weights or empirically elicited public preferences.

Demand tail adequacy and multicriteria safeguards

Alturki and Lee [14] survey multicriteria humanitarian logistics models published from 2015 to mid-2023 and identify persistent gaps between modelling research and operational challenges. Together with the explicit efficiency-equity formulation of Huang et al. [12], this work supports examining competing criteria transparently instead of relying on one aggregate score. However, the meaning of a constraint depends on the quantity bounded and the scenarios over which it is evaluated.

The demand-tail safeguard in this paper bounds conditional expected shortage on a demand-defined protected set selected before allocation. CVaR, by comparison, evaluates the upper tail of the loss distribution induced by a policy. High demand and high loss need not identify the same scenarios. The safeguard is therefore a service-adequacy constraint, not a replacement definition of CVaR, a new coherent risk measure or a chance constraint guaranteeing service with a stated probability. With the singleton protected set S12, the conditional expectation reduces to that scenario's pooled shortage.

This construction permits a direct test of whether the safeguard adds an effective restriction. Risk-neutral, CVaR-only, cap-only and zero-priority hybrid comparisons hold physical inputs fixed. Matched joint-priority and integrated policies isolate the cap effect when priority weighting is active. An inactive cap can express a requirement that is already met, but its inclusion alone does not demonstrate additional protection. Because the cap pools unlike PHC and MNH units, it remains a model adequacy index. The resource-specific coverage constraint examined separately in this paper protects a different outcome and has its own feasibility assessment.

Feasibility frontiers and secondary allocation

A shortage target must be assessed against what the resource system can deliver. Before interpreting tighter caps as stronger protection, the present study minimises protected shortage over the same physical feasible set. Any cap below that optimum is infeasible under the specified assumptions. This feasibility calculation separates a preference for lower shortage from the capacity, stock and budget conditions needed to achieve it. It also prevents an infeasible target from being reported as an attainable policy improvement.

The frontier calculation creates a second interpretative issue: an objective concerned only with protected shortage may leave outcomes elsewhere unresolved. A solver can then return an allocation with poor non-tail service even though another allocation achieves essentially the same protected shortage with lower expected loss. The secondary optimisation minimises expected base loss subject to remaining within the stated numerical tolerance of the minimum protected shortage. This is a numerical lexicographic approximation, not a claim of a new optimisation algorithm or a complete Pareto frontier. It tests whether poor aggregate coverage is necessary for tail protection or is an artefact of an objective that leaves secondary outcomes unspecified.

Latunde et al. [15] examine parameter sensitivity in a linear-programming transportation application. Their study provides a related Nigerian allocation setting; the present analysis extends the comparison to staged uncertainty, loss-tail protection and resource-specific healthcare safeguards.

Research gap and analytical position

The reviewed studies already establish adaptive healthcare allocation, CVaR optimisation, distributional objectives and multicriteria trade-offs. The present contribution is consequently not the introduction of

these ingredients. It is a controlled empirical assessment of how loss protection, geographic priority and a fixed demand-tail adequacy requirement interact in a displacement-derived healthcare system. The analysis connects these controls to a feasibility frontier and to resource-specific and geographic diagnostics, so that changes in a weighted objective can be distinguished from changes in service and distribution.

This positioning follows the four questions in the Background section. Risk experiments address loss protection; matched priority comparisons examine redistribution; cap and frontier experiments establish incremental adequacy and secondary allocation; resource experiments and a separately formulated MNH floor assess service protection. The framework allows small risk benefits, inactive safeguards, mixed equity outcomes and infeasible targets to be reported as substantive findings. Conclusions remain conditional on the twelve-scenario calibration, assumed accessibility network and available actions, without claims of distributional robustness or clinical effectiveness.

Methods

Study setting and input calibration

This is a computational health-service planning study using secondary aggregate displacement, population and facility inputs. It compares allocation policies within a fixed calibrated network; it is not a clinical trial or an evaluation of an implemented programme. The intended decision users are state health planners and humanitarian coordinators deciding how to allocate limited procurement and deployment resources across areas and service classes. The study system separates demand, available resources, information revelation and normative priorities so that matched policy comparisons can isolate the effect of each control.

Study area and healthcare resources

The demand system comprises 177 local government areas (LGAs) in ten North-Central and North-West Nigerian states, linked to 348 supply hubs. An LGA is the geographic demand unit, while a hub is a supply node in the calibrated network. Primary healthcare (PHC) and maternal and newborn health (MNH) are represented as separate resource classes, with their own demand, effective capacity, procurement cost and shortage-penalty inputs. The model allocates service resources to geographic demand; it does not track individual patients or clinical pathways.

Demand quantities represent calibrated planning requirements. They should not be interpreted as directly observed attendance, distinct patient counts or validated forecasts of clinical utilisation. PHC and MNH quantities retain their respective model units. Adding them produces a pooled planning index rather than a clinical equivalence between the two services. Accordingly, system coverage and geographic outcomes are reported separately for PHC and MNH, alongside the pooled measure used by the demand-tail safeguard.

The ten states are Benue, Kaduna, Kano, Katsina, Kogi, Nasarawa, Niger, Plateau, Sokoto and Zamfara. International Organization for Migration Displacement Tracking Matrix (DTM) rounds R15–R18 provide the displacement archetypes [16]. Fixed population and the facility network draw on the Geo-Referenced Infrastructure and Demographic Data for Development programme (GRID3); facility presence is not independently validated effective capacity [17, 18].

For each LGA and archetype, primary PHC demand equals the quarterly planning coefficient multiplied by the sum of fixed population and 0.50 times the recorded displaced population. MNH demand uses quarterly expected births and the state facility-delivery proportion. The premium represents service intensity, not an additional head count. The primary demand formulas are:

$$D_{i,\mathrm{PHC},s} = m_s \frac{45\,000\,000}{237\,527\,782}\left(P_i + 0.50 I_{i,k_s}\right) \quad (1)$$

$$D_{i,\mathrm{MNH},s} = m_s \frac{32.493}{4\,000}\,\mathrm{FBR}_{\mathrm{state}(i)}\left(P_i + 0.50 I_{i,k_s}\right) \quad (2)$$

In equations (1)–(2), fixed population and recorded internally displaced persons (IDPs) are indexed by LGA and archetype, respectively; the scenario severity multiplier is 0.90, 1.00 or 1.15, and FBR denotes the state facility-delivery proportion. The PHC numerator is the reported second-quarter 2025 Basic Health Care Provision Fund facility visits; the denominator is a national population anchor [19, 20]. Their different coverage bases make this a planning proxy rather than an observed national utilisation rate. The crude birth rate is 32.493 per 1,000 annually [21]; division by 4,000 converts it to a quarterly per-person coefficient. Facility-delivery proportions use the live-birth rows of NDHS Table 9.8, printed page 227 [22]: Benue 0.590, Kaduna 0.259, Kano 0.327, Katsina 0.158, Kogi 0.622, Nasarawa 0.557, Niger 0.302, Plateau 0.457, Sokoto 0.125 and Zamfara 0.153. Vulnerability multipliers do not enter primary demand.

The coefficient 0.50 represents an assumed additional planning requirement associated with displacement, applied consistently in both resource formulas; it is not an estimated causal effect of displacement on

utilisation. Fixed population provides the population anchor, and the displacement term supplies an intensity premium. The analysis therefore does not identify how many additional consultations or deliveries displacement causes. The facility-delivery proportion is a utilisation-based calibration input rather than a normative entitlement to maternity care. These choices define the demand quantities against which modelled coverage is measured.

Supply connections and physical allocation periods

Admissible hub-to-LGA connections are defined by the specified distance and effective-speed accessibility estimates. These connections determine which allocations the programme permits. They are fixed across the reported comparisons and are not generated from raster-based travel-time calculations. The network therefore represents a planning approximation to access, without establishing actual journey times, route availability during disruption or clinically timely arrival.

The physical system contains three allocation periods. Hub capacity limits allocations within each period, while cumulative availability is determined by initial inventory and procurement up to that point. A separate expenditure bound applies at each decision node. These restrictions have different roles: capacity controls the volume that can be deployed, inventory and procurement control availability, and the budget controls affordability. Increasing a preference weight changes the ranking of feasible allocations but does not relax any of these resource restrictions.

Demand balance is imposed at the planning horizon. The formulation does not require intermediate clinical service deadlines or represent deterioration caused by waiting. Consequently, a feasible end-of-horizon allocation establishes modelled service availability, not timely treatment. This temporal boundary is retained when interpreting both aggregate coverage and the benefit of initial stock.

The fixed network contains 885 PHC and 1,416 MNH admissible links. Effective capacity is constructed from facility counts, nominal throughput and readiness assumptions, with period shares 0.40, 0.30 and 0.30. Baseline budgets are 2,048,311.829627, 1,689,857.259442 and 1,920,292.340275 cost-index units. These physical and financial inputs are held fixed when preference weights or demand protection change.

Capacity represents the ability to deploy the modelled service resources; it does not establish staffing, opening hours, drug availability or readiness at every facility. The national calibration anchors are

combined with historical displacement archetypes to construct planning conditions, rather than a contemporaneous observational panel. Consequently, the comparisons answer how the specified system allocates resources under those conditions, not how well it predicts a particular future quarter.

Demand archetypes and information revelation

Four spatial demand archetypes correspond to Displacement Tracking Matrix (DTM) rounds R15–R18. Each archetype is combined with LOW, BASE and HIGH severity to form twelve terminal scenarios. The severity multipliers are 0.90, 1.00 and 1.15, respectively. These branches scale the specified archetype requirements; they provide magnitude perturbations within the scenario construction rather than twelve independently observed demand fields.

The three information stages specify when decisions may differ across scenarios. Initial decisions precede revelation of the archetype and are shared by all paths. Second-stage decisions depend on the archetype and are shared by its severity branches. Third-stage decisions depend on the terminal scenario. The historical rounds label alternative archetypes; they are not treated as successive future periods through which every scenario path must pass. Shared-node variables enforce this information restriction in the Allocation model and outcome measures section.

Information stages and physical periods therefore serve distinct purposes, even though both have three levels in this implementation. The first determines what is known when a decision is chosen; the second determines when procurement and allocation enter the resource balances. No allocation decision may exploit severity information before the corresponding branch is revealed.

Planning probabilities and the protected demand set

Each archetype receives probability 0.25. Conditional probabilities for LOW, BASE and HIGH are 0.20, 0.60 and 0.20. Multiplying by the archetype probability gives terminal probabilities of 0.05, 0.15 and 0.05 within each archetype. These sum to 0.25 per archetype and to one across all twelve scenarios. The weights define the experimental planning distribution; they are not estimated frequencies of future displacement events.

The primary protected state is S12, R18 HIGH, selected by an equal-weight score of the two resource totals after each is divided by its mean across the four BASE archetypes. This unit normalisation prevents

the numerical scale of PHC from determining the ranking alone. The selected set is fixed before optimisation and is not reselected using a policy's losses. Because rankings among high-severity archetypes are close, a separate sensitivity protects S06, S09 and S12, whose scores are within 0.1% of the largest score. This near-maximum set has total probability 0.15, not 0.05.

To make selection reproducible, let $d_{rs} = \sum_{i \in I} D_{irs}$ be total demand for resource r in scenario s, and let $s(k, \text{BASE})$ denote the BASE scenario for archetype k. Define

$$\bar{d}_r = \frac{1}{4}\sum_{k=1}^{4} d_{r,s(k,\text{BASE})}, \qquad z_s = \frac{1}{2}\sum_{r \in R} \frac{d_{rs}}{\bar{d}_r} \quad (3)$$

Both denominators are strictly positive in this application. The primary set contains a maximiser of z_s, selected before any allocation solve; the selected catalogue entry is S12. Exact ties require a declared deterministic selection rule or protection of all maximisers. The sensitivity instead uses

$$T_{0.001} = \{s \in S : z_s \geq (1 - 0.001)\max_{v \in S} z_v\} \quad (4)$$

Rescaling all quantities of one resource by a positive constant leaves this score unchanged because its numerator and denominator scale together. This unit invariance applies to scenario selection; it does not apply to the pooled shortage cap, which retains the original PHC and MNH units.

S12 has probability 0.05 and is the sole member of the primary protected set. Conditional expected shortage on that set equals its pooled shortage. The reference cap is 9,147,367.899574 pooled model units, defined by the E00 allocation under the same physical inputs. It is a benchmark-relative planning bound, not an externally established service standard. The near-maximum sensitivity uses its own E00 conditional-shortage reference, 9,143,018.898737 units. Attainability is assessed by frontier optimisation, not assumed from a percentile label. The scenario probabilities also imply that 95% and 99% CVaR equal maximum scenario loss.

Cost budget and inventory calibration

Shortage penalties are 10 cost-index units per unmet PHC unit and 25 per unmet MNH unit. Procurement and allocation costs enter operating expenditure, while unmet demand contributes to the base loss through the resource-specific penalties. These quantities are normalised cost indices, not currency expenditure. The

difference between penalty coefficients affects substitution in the optimisation and should not be read as an empirically measured ratio of clinical harm.

The baseline stage budget is 90% of the supplied full-deployment procurement benchmark. The LOW and HIGH settings use 80% and 100% of that same benchmark. Thus, the LOW and HIGH percentages refer to the benchmark, not to multipliers of the baseline budget. Effective capacities, demand, admissible links and scenario probabilities remain fixed in these budget experiments. Their comparator retains the same risk and priority weights, allowing the budget effect to be assessed without a simultaneous preference change.

Initial inventory is zero except in the designated inventory experiment. That experiment uses the capacity-derived stock vector equal to 10% of the archived capacity benchmark. It adds starting stock without expanding any period capacity constraint. Initial stock is an endowment available at the beginning of the horizon; pre-horizon acquisition expenditure is excluded. The comparison concerns deployment under alternative starting stock rather than the all-in economic value of purchasing inventory.

Geographic priorities and their interpretation

The supplied priority weights combine displaced-population share, older-IDP composition and access deficit with coefficients 0.40, 0.35 and 0.25, respectively. Denoting these components by a_{ik}, b_{ik} and c_{ik}, the input rule is

$$w_{ik} = 1 + 0.40a_{ik} + 0.35b_{ik} + 0.25c_{ik} \quad (5)$$

The first component is the recorded IDP count divided by fixed LGA population. The second is the share of IDPs aged 60 years or above, obtained by adding the explicitly recorded male and female age-60-plus counts. It is defined as zero for the four LGA-round observations with both zero IDPs and zero age-60-plus counts. These age counts are available for all 708 observations. The third component is one minus the health-access share. Access is observed in R15 and R17; the LGA-specific mean of these observed values supplies R16 and R18, covering 354 imputed observations. The age component measures older-person composition, not comprehensive vulnerability or MNH clinical need. Missing R16 maternal subgroup indicators are not replaced by zero or used in this priority rule.

$$a_{ik} = \frac{I_{ik}}{P_i}, \quad b_{ik} = \frac{E60_{ik}}{I_{ik}} \ (I_{ik} > 0), \quad c_{ik} = 1 - A_{ik} \quad (6)$$

Here $E60$ denotes the recorded number of IDPs aged at least 60 and A the observed or imputed health-access share; the zero-IDP convention stated above completes the age-share definition.

The coefficients 0.40, 0.35 and 0.25 express normative priorities; they are not fitted welfare parameters. The same LGA-round multiplier applies to both resources. Increasing the objective priority weight changes the ranking of feasible allocations but does not impose equal coverage or a service floor. Sensitivities omit the age or access component without renormalising the other coefficients, and set imputed access deficits to zero or one. All variants are evaluated with the same primary priority diagnostics, preventing a changed scoring rule from being mistaken for an outcome improvement.

Input consistency and experimental scope

The computational input structure separates the scenario catalogue and resource demand from capacity, cost, budget, initial inventory, shortage penalties and priority weights. This separation is needed to preserve a common feasible system when comparing policy controls and to identify the single resource change in each sensitivity case. The verification scope and numerical residuals are reported in the Policy comparisons and verification section; arithmetic agreement does not by itself validate the underlying clinical coefficients or access assumptions.

The calibration supports controlled comparisons of loss protection, redistribution, demand adequacy and resource conditions. The primary demand excludes vulnerability multipliers; subgroup composition enters only through the separately declared priority mechanism. Neither arithmetic consistency nor these comparisons establish predictive accuracy, distributional robustness or operational access. Those questions require additional demand observations, probability validation, access evidence and resource calibration.

Allocation model and outcome measures

The formulation separates physical feasibility from expected base loss and three policy controls: loss-tail aversion, geographic priority and an optional demand-tail shortage cap. All comparisons use the same node-based procurement and allocation structure unless a resource input is explicitly varied. The model is a finite-scenario stochastic linear programme with continuous, non-negative resource quantities. The following definitions make the information restrictions, accounting conventions and outcome measures explicit.

Sets and decisions

Indices distinguish demand locations, supply hubs, resource classes, physical periods and terminal scenarios. A node belongs to one information stage and one allocation period; each terminal scenario identifies a unique path from the root to its terminal node. Only admissible hub-to-LGA links carry allocation variables. Quantities on excluded links are implicitly zero, so accessibility is a feasibility restriction rather than an additional loss penalty.

Let I be the LGA set, J the hubs, $R = \{\mathrm{PHC}, \mathrm{MNH}\}$ the resources and S the terminal scenarios. For each resource, $A_r \subseteq I \times J$ is the admissible allocation network. Let p_s be scenario probability, $k(s)$ its archetype and D_{irs} terminal demand. A scenario path contains nodes $n_t(s)$ for periods $t = 1,2,3$. All paths share the root node; paths with the same archetype share the second node.

At node n, let $x_{jrn} \geq 0$ denote procurement and $y_{ijrn} \geq 0$ allocation. Terminal shortage is $u_{irs} \geq 0$. The full policy is denoted by π. Node-indexed variables impose non-anticipativity directly: scenarios sharing a node use the same decisions. Capacity K_{jrt}, initial inventory I^0_{jr}, budget B_t, procurement cost c^P_{jrt} and allocation cost c^A_{ijrt} are fixed within a given experiment.

Non-anticipativity is imposed through shared variables, not through scenario-specific choices made with advance knowledge. Two paths sharing a node use the same procurement and allocation decisions at that node. Their later variables can differ only after their information paths separate. Parameters known at the outset, including the physical capacities and cost coefficients, remain fixed within an experiment; terminal demand and archetype-dependent priorities distinguish the scenarios.

Physical and budget constraints

Equations (7) and (8) hold for every hub, resource, terminal scenario and period on its path. Capacity bounds outgoing allocation in a period. The cumulative availability inequality allows unused stock to carry forward and prevents allocation before procurement or initial stock makes it available. The formulation assumes no stock loss or expiry over the planning horizon.

For every scenario path, resource and hub, allocation in a period is bounded by capacity:

$$\sum_{i:(i,j)\in A_r} y_{ijr,n_t(s)} \leq K_{jrt}, \qquad t = 1,2,3 \quad (7)$$

Cumulative allocation cannot exceed initial inventory plus procurement to that point:

$$\sum_{\tau=1}^{t} \sum_{i:(i,j)\in A_r} y_{ijr,n_\tau(s)} \le I_{jr}^{0} + \sum_{\tau=1}^{t} x_{jr,n_\tau(s)} \quad (8)$$

Define node expenditure, for a node at period t, as

$$C_n = \sum_{j\in J} \sum_{r\in R} c_{jrt}^{P}\, x_{jrn} + \sum_{r\in R} \sum_{(i,j)\in A_r} c_{ijrt}^{A}\, y_{ijrn} \quad (9)$$

The budget applies separately at each decision node:

$$C_n \le B_t \quad (10)$$

Equation (9) counts procurement and allocation expenditure at the node's period. Equation (10) is imposed at every node separately: the stage budget is available on the realised path, not pooled across mutually exclusive branches. A shared-node constraint need only be represented once. Shortage penalties influence the objective but do not consume this operating budget. The cumulative stock restriction and node budget therefore control different aspects of feasibility.

Terminal service and shortage satisfy demand balance:

$$\sum_{t=1}^{3} \sum_{j:(i,j)\in A_r} y_{ijr,n_t(s)} + u_{irs} = D_{irs}, \quad i \in I,\ r \in R,\ s \in S \quad (11)$$

Equations (7)–(11), non-negativity and the shared-node information structure define the physical feasible set $\mathcal{F}$. Repeated path constraints at a shared node are implemented once per node. No hard minimum LGA coverage floor is included in these experiments.

Assume finite non-negative demand, capacities, initial inventories, budgets and cost coefficients, strictly positive scenario probabilities summing to one, positive shortage penalties, and finite geographic weights at least one. The zero-procurement, zero-allocation policy with $u_{irs} = D_{irs}$ belongs to $\mathcal{F}$. Demand balance bounds every shortage between zero and its demand. Thus $\mathcal{F}$ is a non-empty polyhedron, and each linear objective below that is bounded below attains its optimum; boundedness of the entire procurement feasible set is not required.

A useful consequence of non-anticipativity and non-negative remaining service is the early-allocation bound. If $S(n)$ contains the terminal descendants of a node n in period t, then, for each LGA and resource,

$$\sum_{\tau=1}^{t} \sum_{j:(i,j)\in A_r} y_{ijr,n_\tau(s)} \le \min_{v\in S(n)} D_{irv}, \qquad s \in S(n) \quad (12)$$

The left-hand side is common to all descendants because their histories coincide through n; applying terminal demand balance to each descendant proves the bound. Consequently, unrestricted shortage ensures feasibility of the initial model but cannot repair an early allocation exceeding a possible descendant's demand.

Equation (11) equates cumulative service plus unmet demand to terminal demand for each LGA, resource and scenario. Non-negative shortage prevents over-service relative to that terminal requirement. There are no transfers between PHC and MNH units and no constraint that makes one service a clinical substitute for the other. Coupling arises through shared financial resources and the objective. Because demand is balanced only at the horizon, these constraints do not guarantee intermediate clinical timeliness.

Base loss and priority increment

Base loss is the common accounting measure used to compare policies. It adds expenditure along the realised path and the terminal shortage penalties. When expected loss is formed, each terminal loss is multiplied by its scenario probability; shared-node expenditure is thereby weighted by the probability of reaching that node. It is not charged as if all mutually exclusive branches occurred together.

For shortage penalty h_{ir}, define base scenario loss by

$$L_s(\pi) = \sum_{t=1}^{3} C_{n_t(s)} + \sum_{i\in I} \sum_{r\in R} h_{ir}\, u_{irs} \quad (13)$$

The geographic-priority increment added to expected loss is

$$V(\pi) = \sum_{s\in S} p_s \sum_{i\in I} \sum_{r\in R} h_{ir} \left(w_{i,k_s} - 1\right) u_{irs} \quad (14)$$

Thus, with priority weight η, the shortage coefficient in the expected-loss and priority terms is $p_s h_{ir}\left[1 + \eta\left(w_{i,k(s)} - 1\right)\right]$. The separately reported priority-weighted shortage is

$$W(\pi) = \sum_{s\in S} p_s \sum_{i\in I} \sum_{r\in R} w_{i,k_s}\, u_{irs} \quad (15)$$

The diagnostic W omits the resource penalty and uses the full priority weight; it must not replace V in the objective or be interpreted as an inequality index. Both pooled shortage diagnostics retain unlike service units. Resource-specific values and coverage are reported alongside them.

Subtracting one from the full priority weight in Eq. (14) makes this term an increment above the base shortage penalty. A zero priority weight therefore recovers the base-loss model. The displayed coefficient following Eq. (14) describes only the expected-loss and priority terms; CVaR can add further sensitivity to shortage through scenario loss. Objective values from different preference weights should not be interpreted as directly comparable welfare scores. The base-loss and service diagnostics retain their definitions across those comparisons.

CVaR and discrete tail interpretation

The confidence level lies strictly between zero and one. The threshold variable identifies a loss cutoff, and each non-negative excess variable bounds the amount by which scenario loss exceeds that cutoff. The CVaR minimisation is jointly over the threshold and the excess variables under the inequalities below Eq. (16). For a fixed threshold and loss vector, minimising excess gives the positive part of scenario loss minus the threshold. This yields the standard convex piecewise-linear tail-loss representation.

For confidence level α, the CVaR representation is

$$\mathrm{CVaR}_{\alpha}(L) = \min_{\zeta \in \mathbb{R},\, \xi \geq 0} \left\{ \zeta + \frac{1}{1-\alpha} \sum_{s \in S} p_s \, \xi_s \right\} \quad (16)$$

subject to $\xi_s \geq L_s - \zeta$ and $\xi_s \geq 0$. In the present non-negative loss model, restricting $\zeta \geq 0$ does not change the optimal value. The epigraph variables are omitted when the risk weight is zero; CVaR is still calculated as an outcome diagnostic.

In a discrete distribution, the upper-tail probability mass may include only part of a scenario's probability. CVaR is therefore not generally obtained by averaging all losses strictly above an empirical quantile. Fractional boundary mass must be retained. The following result identifies the special case relevant to this calibration; it holds for any feasible policy and therefore also for policies selected under different priority weights.

Proposition 1. For a fixed policy, CVaR at confidence level α equals maximum loss if and only if the total probability of the maximum-loss scenarios is at least 1 − α. In particular, a minimum scenario probability of at least 1 − α is sufficient for this identity to hold for every feasible policy.

The proof is provided in Additional file 1. The identity concerns the specified finite planning distribution; it does not estimate the frequency or severity of unobserved displacement events.

The proposition applies at both 95% and 99% here. For other confidence levels, CVaR is computed by sorting losses in descending order and taking exactly $1 - \alpha$ probability mass, including a fractional boundary scenario where necessary.

At the two confidence levels satisfying Proposition 1, the optimisation can equivalently use one worst-loss epigraph variable m:

$$m \geq L_s(\pi)\ (s \in S), \qquad \min_{\pi \in \mathcal{F},\, m} E(\pi) + \lambda m + \eta V(\pi) \quad (17)$$

with the chosen cap imposed separately and $E(\pi) = \sum_s p_s\, L_s(\pi)$. For $\lambda > 0$, minimisation sets m to maximum loss. This is an equivalent finite-scenario representation, not robustness against demand states outside the catalogue.

For positive risk weight, the CVaR epigraph can be included directly in the linear objective together with its excess constraints. For zero risk weight, the epigraph variables do not determine the allocation and their arbitrary feasible values must not be used to report CVaR. The diagnostic is instead evaluated from the resulting scenario losses. This preserves a common assessment of risk-neutral and risk-averse policies.

Demand tail adequacy and integrated objective

The protected set is non-empty and has strictly positive total probability. Its membership is fixed from demand information before optimisation. Conditional weights are the original scenario probabilities divided by the total probability of this set; they sum to one over its members. Equation (18) averages pooled shortage using these conditional weights, while Eq. (19) imposes a prescribed non-negative adequacy bound.

For the fixed protected set $\mathcal{T}$, write $P_{\mathcal{T}} = \sum_{s \in \mathcal{T}} p_s$ and define

$$H_T(\pi) = \frac{1}{P_T} \sum_{s \in T} p_s \sum_{i \in I} \sum_{r \in R} u_{irs} \quad (18)$$

The safeguard is

$$H_T(\pi) \le \Gamma \quad (19)$$

For $\mathcal{T} = \{\mathrm{S12}\}$, Eq. (18) is total pooled shortage in S12. The integrated model is

$$\min_{\pi \in \mathcal{F}} \sum_{s \in S} p_s \, L_s(\pi) + \lambda \mathrm{CVaR}_\alpha\big(L(\pi)\big) + \eta V(\pi) \quad (20)$$

with Eq. (19) included only when a cap is specified. Both preference weights are non-negative. CVaR is applied to base loss, excluding the extra priority increment. This prevents an unannounced change of the loss definition across priority experiments.

Equation (20), the physical constraints, and the CVaR excess inequalities form a deterministic-equivalent linear programme because losses, priority increments and protected shortage are linear in the allocation and shortage variables. The risk-neutral model sets both preference weights to zero and omits the cap. CVaR-only allocation uses a positive risk weight and zero priority weight without the cap; cap-only allocation adds the cap to expected-loss minimisation. The zero-priority hybrid combines CVaR and the cap. Matched joint-priority and integrated policies have the same positive preference weights and differ only in whether the cap is imposed. These definitions correspond to the comparisons in Table 1.

For general confidence levels, the explicit linear epigraph form of Eq. (20) is

$$\min_{\pi, \zeta, \xi} \sum_{s \in S} p_s \, L_s(\pi) + \lambda \zeta + \frac{\lambda}{1-\alpha} \sum_{s \in S} p_s \, \xi_s + \eta V(\pi) \quad (21)$$

subject to $\pi \in \mathcal{F}$, $\xi_s \ge L_s(\pi) - \zeta$, $\xi_s \ge 0$ for every s, and $H_T(\pi) \le \Gamma$ when enabled. The threshold ζ is a real variable; its non-negative restriction is optional here because base losses are non-negative. When $\lambda = 0$, the threshold and excess variables can be omitted. A shared node has probability $q_n = \sum_{s : n \in \mathrm{path}(s)} p_s$, so expected operating expenditure is equivalently $\sum_n q_n \, C_n$. This verifies that each root expenditure is counted once in expectation.

Adding the cap restricts the feasible set. At fixed preference weights and physical inputs, the optimal value of the complete minimisation objective therefore cannot improve when the cap is added. Individual diagnostics can nevertheless improve because the constrained optimum redistributes losses and shortages. This distinction is essential when interpreting lower base loss or CVaR alongside a higher complete weighted objective.

Feasibility and secondary optimisation

The frontier calculation uses the physical feasible set without imposing the candidate cap. With non-negative demand and inputs, the zero-procurement, zero-allocation policy with all demand unmet provides a baseline feasible point before a cap is added. Protected shortage is bounded below by zero, so Eq. (22) identifies the smallest attainable value whenever an optimum is attained. The subsequent comparisons use the same physical inputs and the same protected set.

The minimum attainable protected shortage is

$$\Gamma^* = \min_{\pi \in \mathcal{F}} H_T(\pi) \quad (22)$$

Any cap below Γ^* is infeasible under the same physical inputs. A tail-minimising policy can have arbitrary outcomes outside the protected set when these outcomes do not affect Eq. (22). To obtain an interpretable secondary allocation, solve

$$\min_{\pi \in \mathcal{F}} \sum_{s \in S} p_s \, L_s(\pi) \quad \text{subject to} \quad H_T(\pi) \leq \Gamma^* + \varepsilon \quad (23)$$

The numerical tolerance is $\varepsilon = 10^{-5}$ pooled model units. Equation (23) is a numerical lexicographic approximation. It does not claim exact equality to an analytically known frontier. For a protected set containing several scenarios, both Eq. (18) and Eq. (22) require conditional probability weights, even if the probabilities differ.

Proposition 2. Cap feasibility and the value of relaxation. For the stated finite non-negative inputs and any non-empty protected set of positive probability, the minimum in Eq. (22) is attained. A cap is feasible exactly when it is no smaller than that minimum. For fixed risk and priority weights, the optimal complete objective is a non-increasing, convex, piecewise-linear function of the cap on its feasible domain.

The proof is provided in Additional file 1. For planners, the minimum protected shortage is a limit imposed by the specified resource system. A target below that value cannot be delivered merely by changing the priority given to shortage.

A binding cap does not necessarily increase the optimal objective: an unconstrained optimum may already lie on the boundary. Conversely, a slack cap at an attained optimum has zero associated dual multiplier by complementary slackness. Where the value function is differentiable, its slope is the negative of an optimal

non-negative cap multiplier. At a kink, optimal multipliers describe subgradients rather than a single marginal value. No cap shadow prices are estimated in the reported experiments.

The frontier comparison in Eq. (23) is an ε-constraint selection: it minimises expected base loss among policies within ε of the primary shortage optimum. It does not guarantee unique geographic allocations. The stated 10^{-5} tolerance must be interpreted together with primal residuals, rather than rounded to an exact zero.

Service and geographic diagnostics

Coverage is evaluated from expected service and demand under the same scenario probabilities. The ratios in Eq. (24) and Eq. (25) require strictly positive expected demand in the relevant denominator. If a resource or LGA-resource pair has zero expected demand, its coverage ratio is undefined and should be reported as not applicable rather than assigned an artificial value. Equation (24) is a ratio of expected totals, not an unweighted average of scenario coverage percentages.

Resource-specific expected coverage is

$$Q_r = 1 - \frac{\sum_{s\in S} p_s \sum_{i\in I} u_{irs}}{\sum_{s\in S} p_s \sum_{i\in I} D_{irs}} \quad (24)$$

For an LGA, define

$$q_{ir} = 1 - \frac{\sum_{s\in S} p_s\, u_{irs}}{\sum_{s\in S} p_s\, D_{irs}}, \qquad \overline{q}_r = \frac{1}{|I|}\sum_{i\in I} q_{ir} \quad (25)$$

The reported dispersion and low-coverage count are

$$\sigma_r = \left[\frac{1}{|I|}\sum_{i\in I}\left(q_{ir} - \overline{q}_r\right)^2\right]^{1/2}, \quad N_r = \sum_{i\in I} \mathbf{1}\{q_{ir} < 0.20\} \quad (26)$$

Minimum LGA coverage is reported separately. These diagnostics weight LGAs equally, whereas Eq. (24) is demand-weighted. The pooled coverage index replaces the resource-specific sums in Eq. (24) by sums over both resources; it is retained for comparison with the pooled tail measure and is not a clinical equivalence between PHC and MNH.

The standard deviation in Eq. (26) uses the population divisor because the reported LGAs constitute the modelled study set. When coverage is expressed as percentages, its standard deviation is reported in

percentage points. The indicator count uses a strict threshold below 20%, and minimum coverage refers to the smallest defined LGA ratio. These spatial diagnostics differ from minimum scenario coverage: one compares locations, whereas the other compares terminal demand states. All geographic comparisons retain the same demand denominators and resource definitions.

Explicit MNH coverage extension

The main model has no resource-specific coverage floor. A separate sensitivity requires at least a common fraction of total MNH demand to be served in every terminal scenario:

$$\sum_{i\in I} u_{i,\mathrm{MNH},s} \le (1-\rho)\sum_{i\in I} D_{i,\mathrm{MNH},s}\,, \quad s\in S, \quad 0\le\rho\le 1 \quad (27)$$

For fixed ρ, equation (27) is linear. It protects aggregate MNH coverage in each scenario, not every LGA or individual patient, and it is stronger than imposing the same fraction only on probability-weighted demand. The extension retains the F01 objective, S12 cap and all other inputs except the expressly varied capacity multiplier. The largest feasible common floor is found separately:

$$\rho^* = \max_{\pi,\rho} \rho \quad \text{subject to} \quad \pi\in\mathcal{F},\ H_T(\pi)\le\Gamma,\ (27),\ 0\le\rho\le 1 \quad (28)$$

Demand is fixed, so maximising ρ remains a linear programme. Its purpose is to establish a feasibility ceiling; its other allocation decisions do not minimise loss and are not reported as efficient policies. All floors at or below the attained maximum are feasible, while larger floors are infeasible under the same inputs. The tested fractions are illustrative planning choices rather than clinically established standards.

Proposition 3. Feasible MNH protection. Suppose the physical and cap constraints are feasible and total MNH demand is positive in every scenario. For policy π, define its worst-scenario MNH coverage by

$$r_{\min}(\pi) = \min_{s\in S}\left(1-\frac{\sum_{i\in I} u_{i,\mathrm{MNH},s}}{\sum_{i\in I} D_{i,\mathrm{MNH},s}}\right) \quad (29)$$

Then Eq. (28) equals the attained maximum of $r_{\min}(\pi)$ over cap-feasible policies. The feasible common-floor interval is exactly $[0,\rho^*]$. With all other inputs and the numerical cap fixed, increasing capacity, budget or initial stock cannot decrease ρ^*.

The proof is provided in Additional file 1. This result concerns the feasibility of a service requirement. It does not guarantee that an allocation chosen solely by cost minimisation will provide more MNH service when capacity increases.

For fixed preference weights, the minimum complete objective is non-decreasing and convex piecewise-linear in a feasible prescribed floor, by the same argument as Proposition 2 with the constraint direction reversed. Expected base loss alone need not be monotone when other objective components are present. Moreover, the scenario-wise floor implies at least the same expected aggregate MNH coverage because the latter is a demand-weighted mean of scenario coverages; the converse does not hold.

Policy comparisons and verification

The experimental design holds the physical system and scenario distribution fixed when comparing risk, priority and demand protection. Separate experiments vary budget, starting inventory or capacity. This distinction separates preferences about who should receive service from changes in what the system can afford or deploy. Comparisons use common definitions of loss, shortage and coverage. Additional file 1 establishes the mathematical ordering properties, while the computational records support the reported numerical checks.

Matched policy comparisons

Table 1 identifies the principal policies. E00 minimises expected base loss without a demand-tail cap. C02 adds CVaR at risk weight 0.50, while T01 adds only the baseline cap to expected-loss minimisation. HYBRID 0 combines that risk weight and the cap with zero priority weight. These four cases form a matched comparison: E00 against C02 isolates risk aversion without the cap; E00 against T01 isolates the cap without risk aversion; and C02 against HYBRID 0 tests the incremental cap effect in a risk-sensitive model. Fig. 1 shows these relationships.

The reference cap is 9,147,367.899574 pooled model units and protects S12 throughout the principal comparisons. J01 and F01 use risk weight 0.50 and priority weight 0.25; only F01 includes the cap. Their difference identifies the cap effect at fixed weights. F01 against E00 describes the combined policy change and cannot attribute it to one control. GSTAR and LEX TE answer the separate frontier and secondary-allocation questions in the Tail feasibility and secondary allocation section.

Table 1. Principal model comparisons

Case	Risk weight λ	Priority weight	Demand-tail cap	Purpose
E00	0	0	None	Expected-loss benchmark
C02	0.50	0	None	CVaR only
T01	0	0	Γ	Tail cap only
HYBRID 0	0.50	0	Γ	CVaR and tail cap
J01	0.50	0.25	None	Joint risk and priority
F01	0.50	0.25	Γ	Integrated policy
GSTAR	0	0	None	Minimise (18) only
LEX TE	0	0	$\Gamma^* + 10^{-5}$	Minimise expected loss at frontier

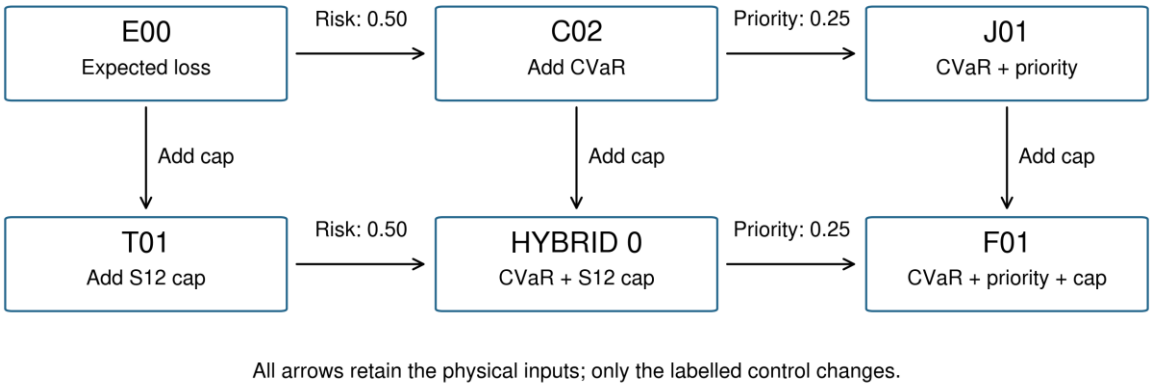


Fig. 1. Controlled policy comparisons. Each arrow changes one policy component while retaining the physical system and the other stated settings. Risk and priority values are objective weights; each cap is the same baseline S12 bound.

Demand, probabilities, capacities, admissible links, cost coefficients, priority inputs and baseline inventory are identical within these matched policy contrasts. The controls may yield the same aggregate outcome even when their formulations differ. Equality of reported totals does not establish equality of every hub allocation or LGA service level. Conversely, adding a cap at fixed objective weights cannot improve the optimum of the complete minimisation objective, although selected base-loss or service diagnostics may improve, as explained in the Demand tail adequacy and integrated objective section.

Preference and resource sensitivity

The CVaR-only sequence uses risk weights 0.25, 0.50, 1 and 2, corresponding to C01–C04, with zero priority weight and no demand-tail cap. The priority-only sequence Q01–Q03 uses weights 0.10, 0.25 and 0.50, respectively, with zero risk weight and no cap. These sequences test changes in the ranking of feasible allocations. They do not increase available resources or impose a geographic coverage floor. Their outcomes are assessed using common base-loss, priority-weighted shortage and resource-coverage diagnostics rather than by comparing differently weighted objective values as welfare scores.

Budget cases B01 and B02 and initial-inventory case I01 retain the J01 preference weights and use J01 as their matched comparator. B01 sets budgets at 80% of the supplied full-deployment procurement benchmark, B02 at 100%, and J01 at 90%. These are percentages of the same benchmark, not reductions or increases of 20% relative to J01. Capacity, demand and probabilities remain fixed. I01 retains the baseline budget and uses the supplied stock vector labelled 10%; it does not imply a uniform increase in hub capacity. Pre-horizon stock acquisition expenditure is excluded from base loss, as specified in the Cost budget and inventory calibration section.

The resource experiments assess the consequences of the designated input change under the same preference settings. They do not identify a unique bottleneck, estimate returns on investment or establish how coverage would respond to arbitrary resource expansions. In particular, comparisons with E00 would combine a resource change with different objective weights. The matched J01 comparisons avoid that confounding and support the resource-specific interpretation in the Preference and resource sensitivity section.

Common evaluation of losses and service

Confidence-level cases A90 and A99 use 90% and 99% CVaR in policy optimisation while retaining the J01 risk and priority weights. Their realised scenario losses are also evaluated at a common 95% confidence level. This separates the choice of policy from the definition of the reported loss-tail statistic. A lower 90% CVaR than 95% CVaR is not, by itself, evidence of a better policy because the two averages use different upper-tail probability masses.

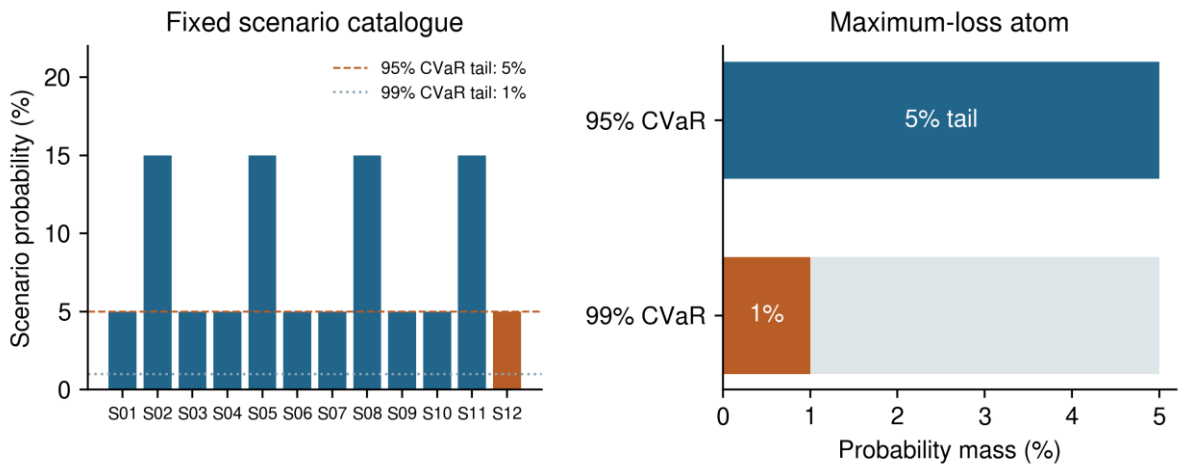


Fig. 2. Planning probability masses and the loss-tail sizes. Panel (a) shows catalogue probabilities, with S12 highlighted as the fixed demand-protection state. Panel (b) illustrates how a maximum-loss atom of probability 5% can supply either the entire 5% upper tail or the entire 1% upper tail. Its identity need not be S12; larger maximum-loss atoms also satisfy Proposition 1. In panel (a), the horizontal dashed line marks 5% probability mass (the upper tail for 95% CVaR), and the dotted line marks 1% (the upper tail for 99% CVaR). They are reference masses, not additional scenario probabilities.

For each policy, CVaR is calculated from its base scenario losses using descending loss order and exactly the required upper-tail mass, including fractional boundary mass when necessary. Fig. 2 explains the special case in Proposition 1: the 5% and 1% upper tails can lie entirely within a maximum-loss atom, making both 95% and 99% CVaR equal maximum loss. The identity of that atom need not be S12. The demand-tail diagnostic instead evaluates pooled shortage in the predetermined S12 state and retains its distinct interpretation.

Expected base loss uses the original scenario probabilities. The priority increment in equation (14) enters the optimisation objective, whereas equation (15) supplies the separately reported priority-weighted shortage diagnostic. These quantities are reconstructed separately. PHC and MNH coverage use the ratio of expected service to expected demand in equation (24); they are not unweighted means of scenario percentages. LGA dispersion, minimum coverage and counts strictly below 20% use equations (25)–(26) and are evaluated separately for each resource. A pooled coverage change is therefore not treated as sufficient evidence of improved geographic equity or improved service in both resources.

Tail feasibility and secondary allocation

The frontier experiment first solves GSTAR, equation (22), over the common physical feasible set without a candidate cap. Its objective is protected S12 shortage alone. This calculation establishes the minimum attainable bound for the specified inputs; comparing a proposed cap with that minimum distinguishes a feasible adequacy requirement from a requirement that the system cannot meet. The frontier is specific to the baseline budgets, capacities, inventory and admissible links. It must not be transferred unchanged to a different resource system.

The second calculation solves LEX TE, equation (23), minimising expected base loss while permitting protected shortage no more than 0.00001 pooled model units above the computed minimum. The tolerance implements a numerical lexicographic approximation and is considered alongside solver residuals. GSTAR and LEX TE are consequently compared on both protected shortage and secondary service outcomes. Poor non-tail service in one GSTAR allocation is not interpreted as a necessary cost of reaching the frontier when a secondary allocation achieves essentially the same protected shortage with lower expected loss.

This procedure identifies one secondary-efficient allocation near the protected-shortage minimum; it does not trace a complete Pareto frontier or establish uniqueness. T02 and T03 impose 90% and 80% of the reference cap with zero risk and priority weights. JT1 imposes the 90% cap with risk weight 0.50 and priority weight 0.25. These explicit caps are evaluated by direct solves and against the computed feasibility bound.

Linear programme implementation and evidence coverage

Sparse deterministic-equivalent linear programmes are solved through SciPy's linear-programming interface with HiGHS and presolve enabled. The verification environment uses Python 3.12.14, SciPy 1.17.0 and NumPy 2.3.5. Huangfu and Hall [23] provide the solver-method reference recommended by HiGHS. Shared node variables implement the information structure directly: common decisions cannot differ across scenarios before the corresponding archetype or severity revelation. The model uses continuous quantities, so the reported solutions concern a linear allocation programme rather than indivisible clinical units.

The computational evidence covers the 18 core cases E00, C01–C04, Q01–Q03, J01, T01–T03, JT1, B01–B02, I01, A90 and A99, together with integrated, frontier, priority-construction, protected-set and MNH-floor experiments. Each reported policy uses the common calibrated demand and declared settings. Additional file 2 contains computational inputs, scripts, a case register and verification records. The combined register contains 40 records: 34 optimal results and six infeasible results, including a diagnostic cap below the frontier. Feasibility-maximisation allocations are distinguished from cost-minimising policies.

Each case is identified by its objective weights, confidence level, numerical cap when enabled, and resource inputs. A cap-enabled flag together with the applied numerical bound identifies an adequacy restriction; a case name alone does not establish whether a cap is present. This explicit definition is particularly important for the frontier calculations and for comparisons in which objective weights are held fixed.

Numerical and accounting verification

Verification separates outcome arithmetic from optimisation feasibility. The common-input audit includes 59 checks on demand formulas, scenario probabilities, keys, capacities, budgets and stock construction. Priority and tail settings have a separate 30-check acceptance record. The completed experiment suite passes 102 acceptance checks covering primal feasibility, objective reconstruction, scenario-wise MNH floors, cap satisfaction, common 95% CVaR, monotonic comparisons and agreement with the maximum-floor feasibility envelope. These checks establish computational consistency, not clinical validity.

For the 25-case sensitivity and MNH-protection suite, the largest absolute equality residual is 1.16×10^{-10}, the largest inequality violation is 7.00×10^{-7}, and the maximum objective reconstruction discrepancy is 4.77×10^{-7} cost-index units. Twenty cases are optimal and five return explicit infeasible statuses. Per-case residuals and corresponding checks are retained in the supplement. The priority increment and full-weight shortage diagnostic are reconstructed separately. Residuals are numerical discrepancies in model units, not percentages of demand.

Near-equality comparisons require unrounded output. The presentation tables round loss values to millions and coverage to percentages, while the frontier and cap comparisons retain sufficient decimal places to

distinguish numerical tolerance from substantive slack. A rounded match between policy summaries does not certify identical decision vectors. Reported improvements that are small relative to total loss are interpreted at that scale rather than presented as large practical gains. No p-values or statistical-significance claims are attached to optimisation differences in this fixed scenario experiment.

Priority protected-set and MNH-floor sensitivities

Priority construction is tested by omitting age or access and by assigning the imputed access deficits their endpoints zero and one. The remaining coefficients are not renormalised. Primary-weight diagnostics evaluate every resulting allocation on a common scale. A near-maximum protected-set experiment uses S06, S09 and S12 with conditional weights normalised over their combined 0.15 probability; its cap is the corresponding E00 conditional shortage.

The MNH extension tests floors of 25%, 50%, 60% and 70% at baseline and 120% capacity. Each retains risk weight 0.50, priority weight 0.25, baseline budgets, zero inventory and the same numerical S12 cap. Capacity expansion does not trigger budget or initial-stock increases. No-floor F01 at the matching capacity supplies the comparator. A separate maximum-floor solve at each capacity determines feasibility independently of the selected grid.

Verification boundaries and interpretation

T02, T03 and JT1 are confirmed infeasible under their explicit caps. The two 70% MNH-floor cases are also infeasible. The distinct below-frontier diagnostic provides a further infeasible result in the combined register. These outcomes are reported as infeasible requirements, not implemented allocations. The maximum-floor optimisation independently identifies the limit against which the MNH grid is assessed.

LGA outcomes are exported for the risk-neutral E00 and integrated F01 verification policies. They support the selected distributional comparisons in the Resource specific and geographic outcomes section, but they do not establish the geographic effects of every sensitivity setting. Alternative optimal allocations may change individual LGA outcomes while preserving an aggregate objective. The secondary frontier calculation resolves one such interpretative issue without demonstrating uniqueness across all cases.

Taken together, the experiment design and checks support conditional comparisons within the calibrated system. They do not constitute prospective validation, a test of probability ambiguity, or an assessment of

clinical timeliness. the Results section therefore reports loss protection, cap effects, frontier performance, resource-specific service and sensitivity as separate findings, with conclusions limited to the controls and evidence evaluated here.

Artificial intelligence assistance was used in code development, figure preparation and manuscript drafting and editing, including adaptation of the presentation for a health-services readership. The authors remain responsible for checking the computational materials, source citations and manuscript and for approving the final version. AI assistance does not constitute independent validation of the model or its results.

Results

The results distinguish loss-tail protection, the incremental demand-tail cap effect, frontier feasibility, and service distribution across resources and LGAs. Preference comparisons retain the common physical system; budget and stock comparisons use J01. Capacity and MNH-floor comparisons use the corresponding F01 setting. Tables 2–6 report levels and Figs. 3–9 display changes and distributions. Loss values are cost indices, pooled shortages combine unlike service units, and coverage uses the ratios in equations (24)–(26).

Expected loss and loss tail protection

Table 2. Principal risk and adequacy outcomes

Policy	Expected loss million	CVaR 95 million	PHC %	MNH %
E00	81.043572	98.192488	37.743	70.959
C02	81.043622	98.192246	37.743	70.958
T01	81.043572	98.192488	37.743	70.959
HYBRID_0	81.043622	98.192246	37.743	70.958
J01	81.070910	98.193038	37.743	70.302
F01	81.070612	98.192488	37.743	70.309

The risk-neutral benchmark E00 has expected base loss of 81.043572 million and 95% CVaR of 98.192488 million cost-index units. PHC and MNH coverage are 37.743% and 70.959%. These rates

provide the service baseline. Because every probability atom is at least 0.05, 95% CVaR equals maximum scenario loss; it is not an empirical rare-event frequency.

Across C01–C04, increasing the risk weight from 0.25 to 2 reduces maximum loss while raising expected base loss. Relative to E00, worst-loss reductions are 155.792, 241.148, 389.605, 428.148 cost-index units, with corresponding expected-loss increases of 18.075, 50.265, 155.420, 208.544 units. This ordering agrees with Proposition 4 in Additional file 1. Fig. 3 uses cost-index units because millions conceal this scale. At risk weight 0.50, the tail reduction is only 0.000246% of baseline worst loss.

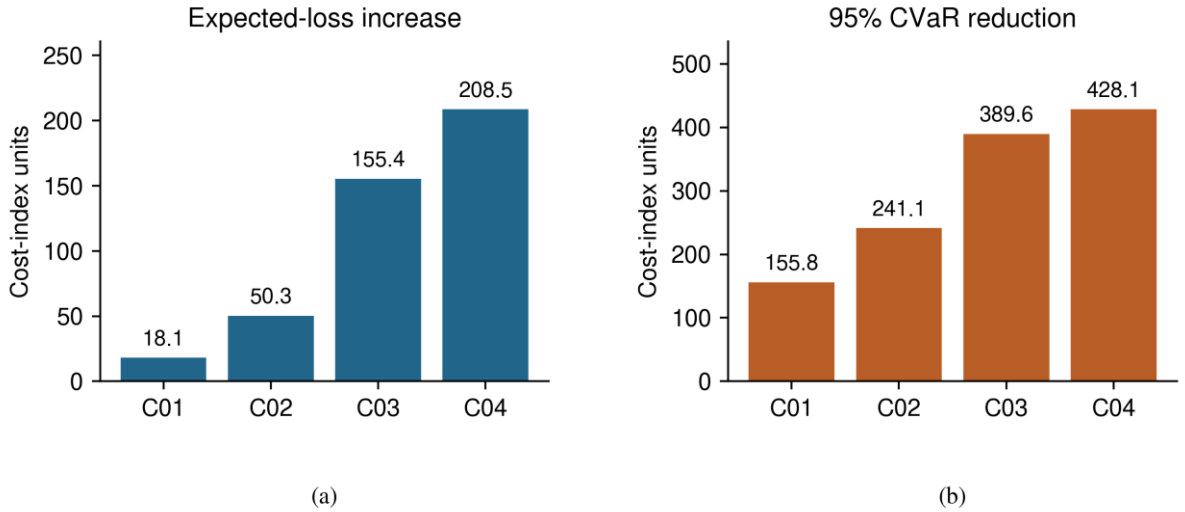


Fig. 3. Expected-loss increases and common 95% CVaR reductions for C01–C04 relative to E00. Both panels use cost-index units, with separate vertical scales. Bars show deterministic optimisation differences, not sampling estimates.

The service response is correspondingly small. PHC coverage is unchanged at the reported precision; MNH coverage moves from 70.959458% under E00 to 70.954445% under C04. Loss protection selects a modest redistribution within the available system rather than expanding service. These differences do not identify a unique physical bottleneck and should not be described as statistical significance or broad clinical improvement.

Incremental effect of the demand tail cap

The cap-only policy T01 reproduces E00 aggregate outcomes. Its protected shortage equals the reference cap of 9,147,367.899574 pooled units; equality alone does not demonstrate an incremental benefit. C02 already satisfies this cap. HYBRID 0 consequently matches C02 at reported aggregate precision and leaves

9.645903 pooled units of slack. The cap leaves the selected CVaR-only optimum feasible and adds no aggregate improvement in this comparison.

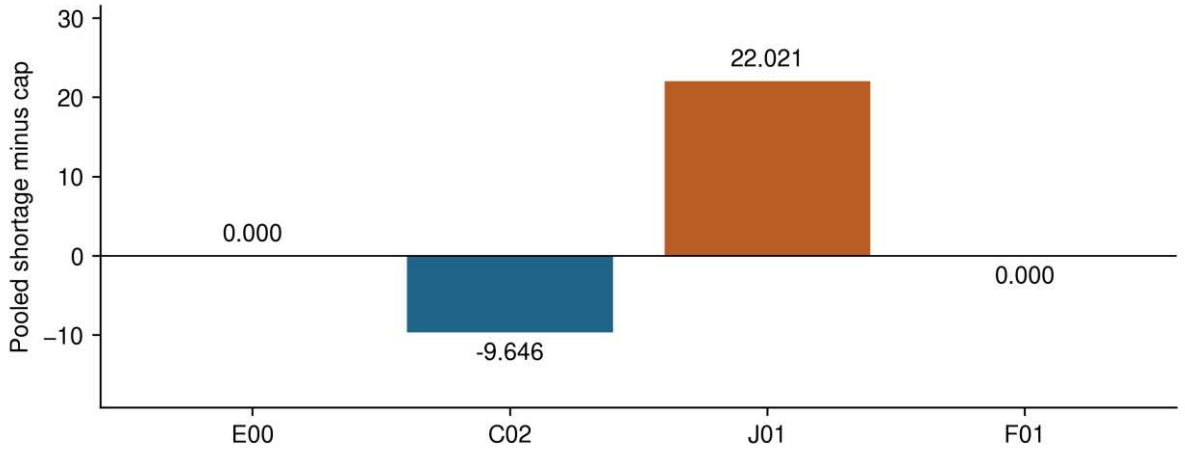


Fig. 4. Protected S12 shortage relative to the baseline cap. Positive values exceed the bound; negative values indicate slack. E00, C02 and J01 do not impose the cap, while T01, HYBRID 0 and F01 do. Values indistinguishable from zero at the displayed precision are labelled 0.000.

With positive priority weight, J01 has protected shortage of 9,147,389.920557, exceeding the bound by 22.020983 units. F01 imposes the cap at the same preference weights and reduces S12 shortage to the bound. Relative to J01, F01 lowers expected base loss by 298.504 and common 95% CVaR by 550.525 cost-index units. MNH coverage rises from 70.302314% to 70.309489%; PHC coverage is unchanged.

These diagnostic gains do not mean that F01 dominates J01. Priority-weighted shortage rises by 200.380 pooled weighted units, and the complete objective increases by 71.468 cost-index units. This agrees with feasible-set inclusion: adding a cap cannot improve the optimum of the same minimisation objective. The demand-defined protected set and the loss-defined tail remain different concepts even when they identify the same scenario.

Against E00, F01 reduces the penalty-weighted priority increment by 4.087% and the pooled priority-weighted shortage diagnostic by 0.168%, while expected base loss increases by 0.033%. Maximum loss is essentially unchanged. This comparison combines risk, priority and cap effects; J01 against F01 isolates the cap effect at fixed weights. The two weighted diagnostics are not interchangeable.

Tail feasibility and secondary allocation

The frontier optimisation gives minimum protected shortage of 9,147,347.633104 pooled units. The reference cap is only 20.266470 units above this minimum, or 0.0002216% of the cap. Little further

reduction in pooled S12 shortage is attainable under fixed inputs. The 90% and 80% reference caps fall below this bound, consistently with the infeasible T02, T03 and JT1 solves.

GSTAR has expected loss of 102.392747 million and pooled coverage of 18.717781%. Its objective does not value outcomes outside the protected set except through shared decisions and feasibility. Low pooled coverage describes one tail-optimal allocation, not a necessary cost of reaching the frontier. A secondary criterion is needed before assigning an efficiency interpretation.

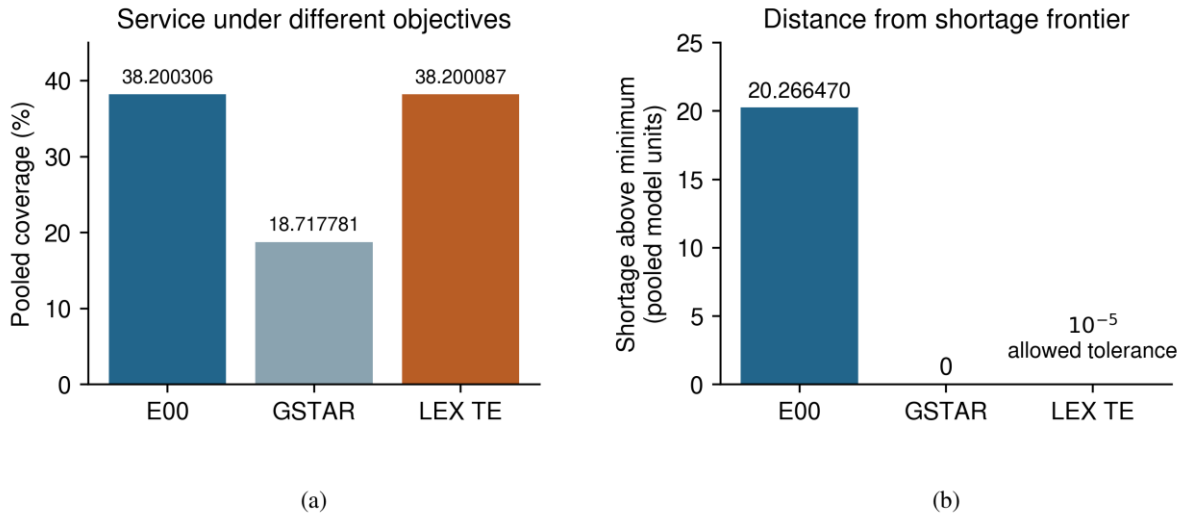


Fig. 5. Pooled coverage and protected shortage above the computed minimum. Coverage labels retain six decimals to distinguish E00 from LEX TE. The right panel explicitly labels the LEX TE allowance as 10^{-5} pooled units; it is not zero. GSTAR is one primary-optimal allocation, whereas LEX TE additionally minimises expected base loss within that tolerance.

LEX TE attains protected shortage of 9,147,347.633114, expected loss of 81.044233 million and pooled coverage of 38.200087%, within 0.00001 pooled units of the frontier. Relative to GSTAR, coverage increases by 19.482305 percentage points and expected loss falls by 21.348514 million. Relative to E00, expected loss increases by 661.589 units, while pooled coverage declines by only 0.000219 percentage points.

The comparison establishes that near-minimum protected shortage is compatible with service close to the expected-loss benchmark. It does not establish a unique allocation, a complete Pareto frontier, or optimal geographic equity. The frontier and the lexicographic tolerance are numerical results for the specified resource system. They must be recalculated if budgets, inventory, capacities, admissible links or the protected set change.

Resource specific and geographic outcomes

F01 retains PHC coverage of 37.743%, while MNH coverage falls from 70.959% to 70.309%, a decrease of 0.649969 percentage points. Pooled coverage falls by only 0.008953 percentage points. PHC dominates the numerical sum of the resource demands, so a small pooled change can conceal a larger MNH change. This weighting does not establish clinical equivalence between visit-equivalents and facility-delivery-equivalents.

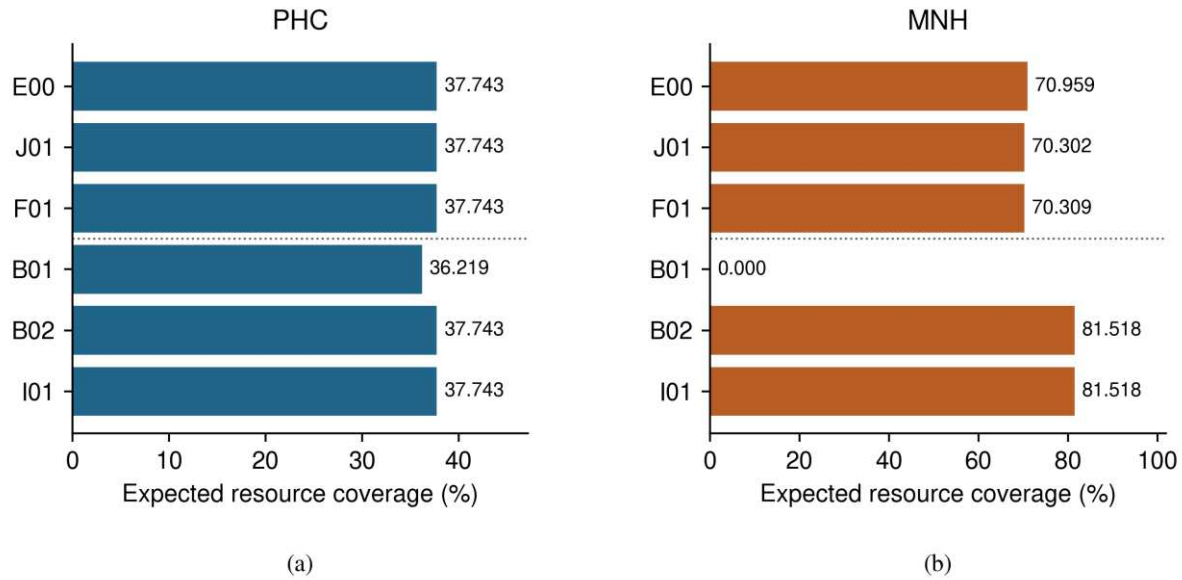


Fig. 6. Expected PHC and MNH coverage, shown separately with numerical labels. The dotted separators distinguish the E00, J01 and F01 preference comparisons from the B01, B02 and I01 resource cases, whose matched comparator is J01. Labels use the rounded values reported in Tables 2 and 4. The horizontal dotted line in each panel is a visual group separator, not a statistical or service threshold.

F01 reduces priority-weighted PHC shortage from 7.775927 million to 7.762177 million, but increases the MNH component from 51,527.947 to 52,121.803. The pooled reduction is driven by PHC and masks the adverse MNH movement. This full-weight diagnostic omits resource penalties, unlike the priority increment in equation (14), and cannot substitute for a geographic fairness assessment.

Table 3. LGA coverage diagnostics for the baseline and integrated policies. SD denotes standard deviation; pp denotes percentage points.

Policy	Resource	Coverage SD pp	Minimum %	Below 20%
E00	PHC	24.852	2.407	27
E00	MNH	24.941	6.086	6
F01	PHC	24.729	1.130	28
F01	MNH	25.189	4.052	4

The geographic outcomes are mixed. PHC dispersion declines from 24.852 to 24.729 percentage points, but its below-20% count rises from 27 to 28 and minimum coverage falls from 2.407% to 1.130%. For MNH, dispersion increases from 24.941 to 25.189 percentage points and minimum coverage falls from 6.086% to 4.052%, although the below-20% count decreases from six to four. Neither lower dispersion nor an improved threshold count guarantees improvement at the least-covered location.

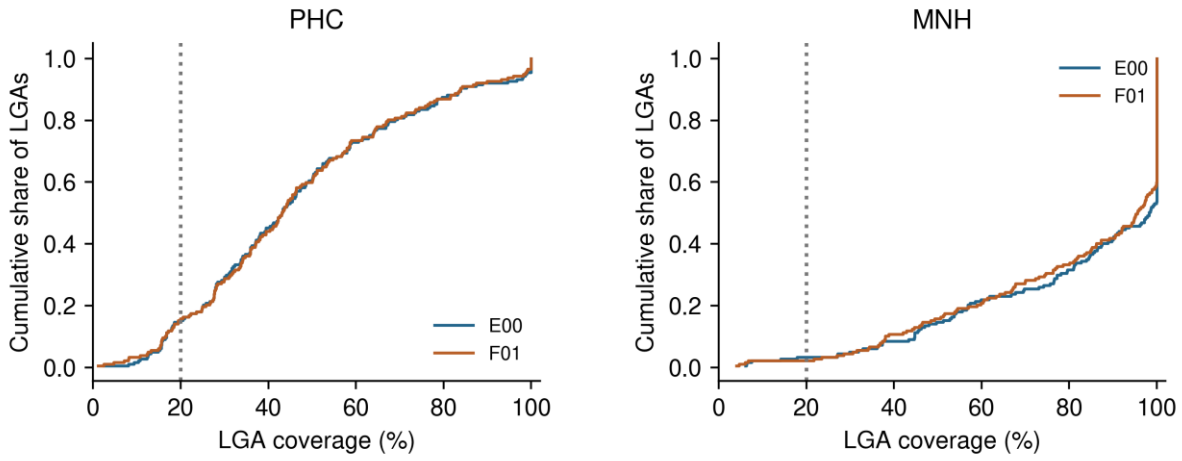


Fig. 7. Empirical cumulative distributions of LGA coverage under E00 and F01, calculated from probability-weighted service and demand for each resource. Each curve includes all 177 LGAs with equal geographic weight. The dotted line marks 20% coverage; Table 3 reports counts strictly below that threshold. The vertical dotted reference line is the 20% LGA-coverage threshold; it is not a fitted trend or confidence limit.

Fig. 7 displays the full distribution rather than only its spread and minimum. At a given coverage threshold, a higher curve means a larger share of LGAs at or below that level. The curves describe the selected allocations and do not show which individual LGAs gained or lost. They are descriptive distributions over the modelled study area, not sampling confidence bands. System coverage weights demand, while these curves weight LGAs equally; their summaries therefore need not move together.

Alternative optimal allocations may also alter individual geographic outcomes without changing aggregate objectives.

Preference and resource sensitivity

Table 4. Sensitivity outcomes evaluated at common 95% CVaR

Case	Expected loss million	CVaR 95 million	PHC %	MNH %
C01	81.043590	98.192332	37.743	70.959
C03	81.043727	98.192098	37.743	70.956
C04	81.043780	98.192059	37.743	70.954
Q01	81.044547	98.193644	37.743	70.936
Q02	81.081069	98.218962	37.743	70.058
Q03	81.210544	98.354971	37.743	66.946
B01	85.182299	102.324774	36.219	0.000
B02	80.686909	97.605873	37.743	81.518
I01	80.142468	97.061431	37.743	81.518
A90	81.067669	98.193295	37.743	70.380
A99	81.070910	98.193038	37.743	70.302

Priority-only policies Q01–Q03 reduce the penalty-weighted priority increment as its coefficient increases, while MNH coverage falls to 70.936%, 70.058%, 66.946%. PHC remains at 37.743%. Q03 has expected base loss of 81.210544 million and worst loss of 98.354971 million. Stronger priority weighting changes the allocation trade-off; it adds no resources and guarantees no coverage floor. The E00–F01 geographic diagnostics do not establish the spatial effects of every sensitivity policy.

Budget and stock effects use J01 because its preference weights match B01, B02 and I01. The budget shares are 90%, 80% and 100% of the same procurement benchmark for J01, B01 and B02. Relative to J01, the changes are −11.111% and +11.111%. Under B01, PHC coverage falls to 36.219% and MNH service is zero. Expected base loss increases by 5.071%.

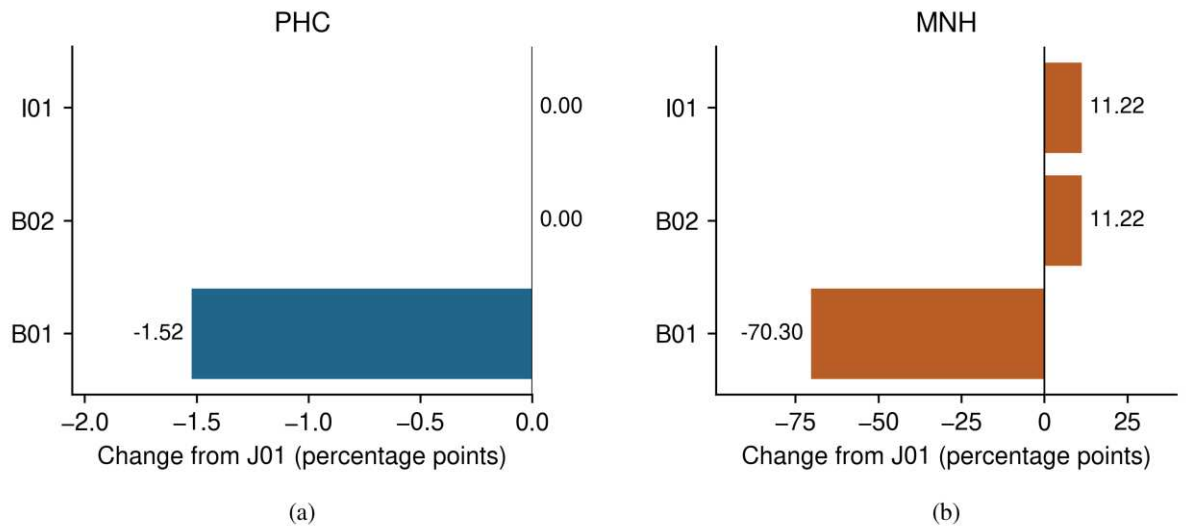


Fig. 8. Resource coverage changes under B01, B02 and I01 relative to the matched J01 policy. Panels use different horizontal scales. B01 and B02 change the budget benchmark share; I01 adds the supplied initial-stock vector at the baseline budget. Demand, capacity and scenario probabilities remain fixed.

B02 and I01 reach MNH coverage of 81.518%, an increase of 11.215668 percentage points over J01, with PHC unchanged. Expected base loss falls by 0.474% and 1.145%, respectively. Equal aggregate coverages do not establish identical allocations or identify a unique binding constraint. The zero-MNH outcome under B01 is a model result without a hard service floor, not a clinical recommendation.

I01 has expected base loss 544,441.040 units below B02 despite identical reported resource coverages. Initial stock is an endowment whose pre-horizon acquisition cost is excluded. The difference concerns within-horizon accounting and deployment; it cannot establish that prepositioning is cheaper on an all-in investment basis.

A99 reproduces J01 aggregate outcomes, consistent with 95% and 99% CVaR both equalling maximum loss. A90 has its own 90% CVaR of 98.179575 million, but common 95% CVaR of 98.193295 million, 256.939 units above J01. Its expected loss is 3,241.007 units lower. These are different trade-offs; comparing differently defined CVaR statistics would confound policy performance and tail definition.

Priority construction and protected-set sensitivity

Omitting age or access changes the allocation because each represents a different normative priority rule. Table 5 evaluates every solution with the primary priority increment and reports MNH coverage. Omitting access moves the solution closer to E00; neither endpoint treatment of imputed access establishes which

unobserved values are correct. These cases test sensitivity to the stated construction rather than validate a comprehensive equity index.

Table 5. Priority and protected-set sensitivity under common primary evaluation

Setting	Expected loss million	Primary V (million)	MNH %
Primary F01	81.070612	3.512665	70.309
Without age	81.067988	3.525294	70.373
Without access	81.045319	3.646681	70.917
Imputed deficit 0	81.063939	3.544707	70.470
Imputed deficit 1	81.063200	3.548124	70.488
Near-maximum set	81.062166	3.556460	70.512

For the near-maximum set, minimum conditional pooled shortage is 9,142,998.524548, only 20.374189 units below its own reference cap. The integrated solution has expected loss of 81.062166 million and MNH coverage of 70.512%. It protects a different set and uses a different cap; its result is not a matched estimate of improving S12 protection. The three-state set has planning probability 0.15 and is not a 5% tail.

Explicit MNH protection and capacity trade-offs

At baseline capacity, no-floor F01 achieves expected MNH coverage of 70.309%, but its minimum scenario MNH coverage is 61.568%. The 25%, 50% and 60% floors therefore leave its optimal objective unchanged. At 120% capacity, the no-floor solution has PHC coverage of 40.764% and zero MNH service. Its expected loss is 80.395614 million and pooled S12 shortage is 8,903,526.160, below the reference cap. Compliance with the pooled cap therefore does not ensure any MNH provision.

Table 6. MNH-floor extension with unchanged budgets and pooled cap

Capacity	Floor	Expected loss million	PHC %	MNH %
100%	None	81.070612	37.743	70.309
100%	25%	81.070612	37.743	70.309
100%	50%	81.070612	37.743	70.309
100%	60%	81.070612	37.743	70.309
100%	70%	Infeasible	—	—
120%	None	80.395614	40.764	0.000
120%	25%	80.598941	39.721	25.000
120%	50%	80.802485	38.677	50.000
120%	60%	80.884454	38.259	60.000
120%	70%	Infeasible	—	—

Coverage columns report probability-weighted aggregate service. Every feasible floor was also checked separately in all twelve scenarios. At 120% capacity, expected MNH coverage equals 25%, 50% and 60% in the corresponding floor cases, while PHC coverage declines to 39.721%, 38.677% and 38.259%. Expected loss rises relative to the matching no-floor allocation. Fig. 9 shows this direct resource trade-off.

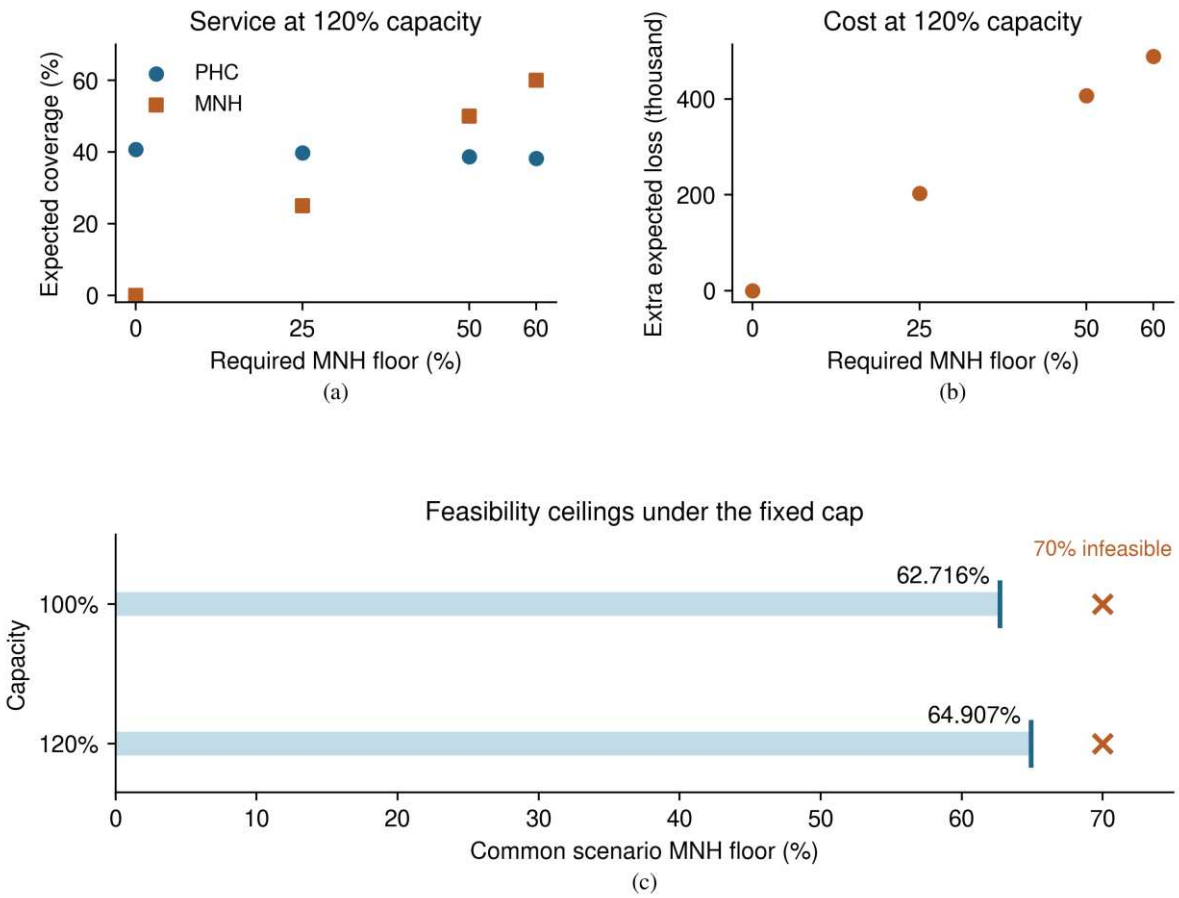


Fig. 9. MNH-floor trade-offs and feasibility. Panels (a) and (b) show only the tested optimal policies at 120% capacity; no continuous trade-off curve is inferred between them. Zero denotes the no-floor comparator. Expected-loss differences are calculated from the displayed million-unit values in Table 6. Panel (c) shows the feasible common-floor intervals established by the separate maximum-floor programmes; their endpoints are feasibility limits, not cost-efficient policies. The 70% floor is infeasible at both capacity levels.

The largest feasible common scenario floor is 62.716% at baseline capacity and 64.907% at 120% capacity. The 70% requirement exceeds both limits. Baseline expected coverage can exceed the maximum common floor because the latter must hold in every scenario. Proposition 3 explains why capacity expansion cannot lower this feasibility ceiling when the other constraints are fixed. The maximum-floor allocations identify feasibility only; no claim is made that their other outcomes are cost-efficient. None of these thresholds is selected as a clinical or ethical standard.

Summary of the empirical findings

Risk aversion yields small worst-loss reductions at a small expected-loss cost. Geographic priority changes resource and LGA outcomes, with weighted gains coexisting with lower MNH coverage and lower minimum LGA coverage. The pooled cap adds no aggregate improvement at zero priority weight but

restricts the preferred allocation under positive priority weights. Secondary expected-loss minimisation approaches minimum protected shortage without the poor non-tail service of a tail-only allocation. Resource conditions have larger effects in the tested ranges, and explicit scenario-wise MNH floors prevent zero service when feasible at a measurable PHC and cost trade-off.

These findings support joint reporting of expected loss, common loss-tail exposure, protected shortage, resource-specific coverage and geographic distributions. No single measure establishes overall policy superiority. The results concern the fixed scenario catalogue and calibrated resource system, with the numerical verification scope set out in the Policy comparisons and verification section. The Discussion section interprets their planning implications without treating them as prospective demand validation or evidence of improved clinical outcomes.

Discussion

The main health-service finding is that protecting an aggregate loss or shortage index does not ensure adequate provision of every service. The integrated policy improved selected priority-weighted measures, but lowered MNH coverage and produced mixed geographic outcomes. Resource-specific minimum requirements addressed a different concern: they prevented zero MNH allocation when feasible, while making the associated PHC and loss trade-offs explicit. These findings favour a planning process that defines the service to protect before choosing a risk or priority weight.

Meaning and scale of loss tail protection

At risk weight 0.50, maximum loss falls by 241.148 cost-index units against a baseline of 98.192 million, while expected loss rises by 50.265 units. Fig. 3 makes this scale visible. The result supports a small preference-dependent allocation change rather than a substantial service-adequacy gain. It does not imply that risk aversion is generally ineffective: its magnitude depends on feasible actions, costs and the scenario distribution.

The discrete distribution is central to this interpretation. Proposition 1 shows that both 95% and 99% CVaR equal maximum loss for every feasible policy under the specified probabilities. Raising the confidence level between these values does not introduce a different tail criterion. The identical J01 and

A99 aggregate outcomes therefore have a structural explanation. The A90 comparison illustrates why policies must be evaluated at a common confidence level: a lower statistic computed over a larger tail mass cannot by itself establish better loss protection.

Risk-sensitive allocation should consequently be assessed alongside resource coverage, rather than judged from the weighted objective alone. The model evaluates CVaR once over total scenario loss; its three-stage decision structure does not make the risk measure recursively dynamic. Changes to scenario probabilities, a richer demand distribution or a nested risk formulation could alter the trade-off and would require separate experiments. The present findings isolate a conditional loss-protection effect within the specified model.

Geographic priority and equity

The integrated F01 policy reduces pooled priority-weighted shortage by 0.168% relative to E00, but MNH coverage declines by 0.650 percentage points and minimum LGA coverage falls for both resources. Table 3 and Fig. 7 show mixed effects: PHC dispersion improves while its low-coverage count worsens; MNH has fewer LGAs below 20% but worse dispersion and minimum coverage. A priority rule can improve an aggregate diagnostic without protecting every distributive dimension.

This distinction is consistent with the criterion-specific treatment of fairness in Bertsimas, Farias and Trichakis [5]. Geographic weights express relative priorities; they do not impose equal service, maximise minimum coverage or guarantee a resource-specific floor. The 0.033% expected-loss increase under F01 is therefore a cost difference for one combination of risk, priority and cap settings, accompanied by mixed outcomes, rather than a universal price of fairness.

Yin and Büyüktahtakın [1] similarly make equity explicit in a multistage epidemic allocation model, using infection and capacity equity measures. Their application links allocation to disease progression, whereas the present study allocates resources against exogenous displacement-derived requirements. The relevant comparison is the need to define the protected outcome, not numerical equivalence between their epidemiological effects and the coverage indices reported here. For this study, resource-specific coverage and LGA diagnostics are needed to judge whether the chosen geographic priorities reflect the intended planning objectives.

Demand tail adequacy as a conditional safeguard

The demand-tail cap has value as an explicit restriction on a specified adverse demand condition, but its incremental effect depends on the objective that selects the allocation. T01 reproduces E00 aggregate outcomes even though the bound is attained. HYBRID 0 reproduces C02 and leaves cap slack. These cases show that a binding equality or the presence of an additional constraint is not sufficient evidence of added service protection. The relevant test is whether the restriction changes the optimum or excludes the preferred allocation under the matched objective.

At positive priority weight, J01 exceeds the S12 cap and F01 changes the allocation to satisfy it. Expected base loss and common CVaR improve relative to J01, while priority-weighted shortage and the complete objective worsen. This redistribution is consistent with feasible-set inclusion. Fig. 4 explains the cap more precisely than E00 against F01, which changes several controls simultaneously.

The safeguard protects a demand-defined state independently of the allocation-induced loss ranking. Its bound is not a probability guarantee, confidence interval or clinical standard. The near-maximum sensitivity also shows that protecting several close-ranked high-demand states is a different planning choice, with probability mass 0.15. Because pooled shortage combines unlike resource units, its protection cannot be interpreted as a guaranteed MNH service level. The separately tested MNH constraint directly addresses that limitation.

Feasibility and the role of secondary optimisation

The reference cap lies only 20.266 pooled units above the minimum protected shortage. This narrow interval separates a stronger protection target from what the fixed model can achieve. Substantial cap reductions cannot be delivered simply by increasing shortage penalties: physical conditions would have to change. The result identifies a system-specific limit without identifying the most efficient intervention for relaxing it.

The comparison between GSTAR and LEX TE adds a separate methodological lesson. A tail-only objective leaves service outside the protected set without a direct value in the objective. The low pooled coverage in the returned GSTAR allocation is therefore not an unavoidable cost of reaching the frontier. Secondary expected-loss minimisation retains protected shortage within 0.00001 pooled units of the

computed minimum while recovering coverage close to E00, as Fig. 5 shows. Reporting only the first tail-optimal allocation would materially overstate the efficiency cost of protection.

A defensible frontier assessment should thus distinguish the primary optimum from the secondary allocation selected near it. The second solve answers an allocation question left open by the first; it does not guarantee uniqueness or geographical fairness. Its tolerance and feasibility residuals must remain visible when differences are very small. This interpretation also prevents the nearly identical rounded coverage values for E00 and LEX TE from being mistaken for identical policies.

Resource conditions and planning implications

The matched budget and stock comparisons yield much larger MNH coverage changes than the tested risk weights. B02 and I01 increase MNH coverage by 11.216 percentage points; B01 yields zero MNH service. At 120% capacity with baseline budgets, the no-floor integrated policy also assigns zero MNH service despite lower expected loss and compliance with the pooled cap. More physical capacity can therefore change the preferred resource mix without protecting both services.

The explicit MNH extension demonstrates what a mandatory resource floor adds. At 120% capacity, the feasible 25%, 50% and 60% scenario-wise floors restore MNH service while increasing expected loss and reducing PHC coverage. The 70% target is infeasible at both capacity levels. The maximum common floors, 62.716% and 64.907%, are conditional feasibility limits, not recommended clinical standards. Scenario-wide aggregate protection also leaves individual LGAs unprotected unless a separate geographic requirement is imposed.

Inventory results also require consistent economic accounting. I01 treats stock as an available endowment and excludes its pre-horizon acquisition cost, while B02 expands the operating budget. Their equal resource coverages do not make them equivalent investments. Procurement, storage, expiry, deployment lead times and any displaced use of funds would need to enter an operational appraisal. The present comparison establishes a within-horizon allocation response rather than a recommendation to purchase a particular stock level.

Contribution and conditions for practical use

The contribution is to show how established optimisation controls affect distinct service-planning outcomes within a common displacement-derived system. The comparisons distinguish small loss-tail gains, priority-driven redistribution, the conditional effect of a pooled shortage cap and the value of selecting an efficient allocation near the shortage minimum. The explicit MNH requirement demonstrates why a pooled safeguard cannot substitute for a resource-specific commitment. The study extends the scenario construction and expected-loss allocation reported in our related work [6, 7, 8] by examining these protection mechanisms and their feasibility together; the shared data setting is not a new independent validation sample.

For state health authorities and humanitarian coordinators, the model can support a structured comparison of candidate allocation rules. Planners would first identify mandatory service requirements with service providers and affected communities, then test whether those requirements can be met with available resources. Feasible policies should be compared using expected loss, adverse loss, resource-specific coverage and geographic outcomes. The model can identify incompatible targets and displaced service, but the choice of protected services and acceptable trade-offs remains a policy judgement. The tested coverage floors are illustrations, not recommended clinical standards.

Operational use would require validation of facility readiness, procurement costs, access routes and service demand, followed by testing on demand conditions not used in the calibration. Clinical deadlines and any indivisible resources would also need representation where relevant. Stakeholder-informed service redesign in Bowers et al. [2] and the accessibility-focused allocation study of Wang et al. [3] illustrate ways of connecting model criteria to service-management decisions. The present study does not report such an implementation or stakeholder validation. Its immediate use is to make competing allocation objectives and feasibility limits transparent before operational deployment.

Strengths and limitations

A strength of the study is the use of matched comparisons that distinguish allocation preferences from physical resource changes. Explicit feasibility calculations prevent unattainable shortage targets from being presented as implementable policies. Reporting PHC, MNH and LGA outcomes separately reveals service

losses concealed by pooled or weighted indices. Numerical audits provide evidence of implementation consistency, although they cannot validate the demand proxies or normative priorities.

The model uses calibrated planning demand and effective capacity, four historical archetypes and assumed probabilities. The PHC utilisation proxy combines a programme-facility numerator with a national denominator; population vintage, exposure error, nominal throughput and readiness remain uncertain. Accessibility uses distance and effective-speed proxies, with access-share imputation in two rounds. The age-60-plus priority component is not comprehensive vulnerability or MNH clinical need. Terminal balance imposes no intermediate clinical deadlines. Pooled shortage combines unlike units, while the MNH floor protects scenario totals rather than each LGA or patient.

The results are conditional on these inputs and normative weights. Feasibility thresholds do not establish clinically achievable targets, and LGA allocations can differ across alternative optima. Capacity sensitivity retains baseline budgets and initial stock; it is not an appraisal of funded capacity expansion. Initial-stock comparisons exclude pre-horizon acquisition costs. The study does not claim prospective predictive validation, clinical benefit, distributional robustness or comprehensive fairness.

The priority index includes the share of older displaced people but no direct indicator of obstetric urgency, neonatal risk or other individual clinical needs. Applying the same geographic multiplier to PHC and MNH therefore cannot be interpreted as clinical prioritisation. The imputation of health access in two rounds and the assumed displacement premium introduce additional uncertainty. Existing construction and resource sensitivities examine selected assumptions, but they do not estimate causal displacement effects or eliminate the need for external validation.

Conclusions

Within this scope, risk aversion yields small worst-loss gains, geographic priorities redistribute shortage with mixed service effects, and a pooled extreme-demand cap adds protection only when it excludes the preferred allocation. Secondary expected-loss minimisation prevents an arbitrary tail-optimal allocation from overstating the cost of approaching the shortage frontier. Explicit, feasible MNH floors protect aggregate service in every scenario but introduce PHC and cost trade-offs. A defensible planning process

must therefore specify the protected outcome, establish feasibility and compare resource-specific consequences before selecting a policy.

List of abbreviations

CVaR, conditional value-at-risk; DTM, Displacement Tracking Matrix; GRID3, Geo-Referenced Infrastructure and Demographic Data for Development; IDP, internally displaced person; LGA, local government area; MNH, maternal and newborn health; NDHS, Nigeria Demographic and Health Survey; PHC, primary healthcare.

Declarations

Ethics approval and consent to participate

The study uses secondary aggregate geographic data and computational allocation models; no participants were recruited and no individual clinical records were analysed.

Consent for publication

Not applicable. The manuscript contains no identifiable individual participant information.

Availability of data and materials

Derived computational inputs, model and analysis scripts, numerical results, mathematical proofs, and verification records are available. Original third-party datasets are identified in the references and remain subject to their providers' access and redistribution terms. No public repository DOI is asserted.

Competing interests

The authors declare that they have no financial or non-financial conflicts of interest, financial encumbrances, or obligations that could influence the conduct, findings or publication of this research.

Funding

This research was self-funded by the authors. No grants or external funding were received.

Authors' contributions

CBO led conceptualisation, methodology, data curation, computation, analysis, visualisation and drafting. BOO provided principal methodological guidance, supervision and critical review. ML contributed to the methodology and critical reviews

Acknowledgements

The authors acknowledge the International Organization for Migration Displacement Tracking Matrix and the Geo-Referenced Infrastructure and Demographic Data for Development programme for the source data used in this study.

References

1. Yin X, Büyüktahtakın IE. A multi-stage stochastic programming approach to epidemic resource allocation with equity considerations. Health Care Manag Sci. 2021;24:597–622. https://doi.org/10.1007/s10729-021-09559-z.

2. Bowers J, Cheyne H, Mould G, Miller M, Page M, Harris F, et al. A multicriteria resource allocation model for the redesign of services following birth. BMC Health Serv Res. 2018;18:656. https://doi.org/10.1186/s12913-018-3430-1.

3. Wang Q, Weng L, Li J, Shi L. An analysis and optimization model for healthcare equity of medical resources allocation in an aging community. BMC Health Serv Res. 2026;26:239. https://doi.org/10.1186/s12913-026-14068-2.

4. Rockafellar RT, Uryasev S. Conditional value-at-risk for general loss distributions. J Bank Financ. 2002;26:1443–1471. https://doi.org/10.1016/S0378-4266(02)00271-6.

5. Bertsimas D, Farias VF, Trichakis N. The price of fairness. Oper Res. 2011;59:17–31. https://doi.org/10.1287/opre.1100.0865.

6. Okechukwu CB, Osu BO. Data-driven spatial demand scenarios for multi-stage stochastic healthcare resource planning under displacement uncertainty. International Journal of Development Mathematics. In press.

7. Okechukwu CB, Osu BO. Multistage stochastic healthcare resource allocation under displacement-driven demand uncertainty in Nigeria. Lapai Journal of Applied and Natural Sciences. In press.

8. Okechukwu CB, Osu BO, Onwukwe I. Multi-stage stochastic healthcare resource allocation under displacement-driven demand uncertainty: evidence from Nigeria. Journal of Physical Sciences: Theories, Methods and Applications. 2026;1(1):1–26. https://doi.org/10.68839/jpstma.2026.zn2nxmmz.

9. Fattahi M, Keyvanshokooh E, Kannan D, Govindan K. Resource planning strategies for healthcare systems during a pandemic. Eur J Oper Res. 2023;304:192–206. https://doi.org/10.1016/j.ejor.2022.01.023.

10. Yu X, Shen S. On the value of risk-averse multistage stochastic programming in capacity planning. INFORMS J Comput. 2025;37(5):1143–1162. https://doi.org/10.1287/ijoc.2023.0396.

11. Heyhat S, Rahmani D, Shahparvari S. Data-driven distributionally robust optimization for resilient healthcare resource planning and crisis mitigation. Transp Res E Logist Transp Rev. 2026;205:104449. https://doi.org/10.1016/j.tre.2025.104449.

12. Huang S, Zhu H, Wang H, Liu K, Liu X, Zhang ZH. Balancing the trade-off between efficiency and equity in a stochastic emergency supplies allocation problem. Appl Math Model. 2025;148:116242. https://doi.org/10.1016/j.apm.2025.116242.

13. Dik MNY, Lo CKY, Lai KH. Fairness, equity, & equality in humanitarian logistics: a systematic review analysis. Transp Res E Logist Transp Rev. 2026;212:104927. https://doi.org/10.1016/j.tre.2026.104927.

14. Alturki I, Lee S. A systematic survey of multicriteria models in humanitarian logistics. Int J Disaster Risk Reduct. 2024;102:104209. https://doi.org/10.1016/j.ijdrr.2023.104209.

15. Latunde T, Richard JO, Esan OO, Dare DD. Sensitivity of parameters in the approach of linear programming to a transportation problem. J Niger Soc Phys Sci. 2019;1:116–121. https://doi.org/10.46481/jnsps.2019.14.

16. International Organization for Migration. Nigeria. Displacement Tracking Matrix. https://dtm.iom.int/nigeria. Accessed 8 Oct 2026.

17. Center for International Earth Science Information Network, Columbia University. GRID3 NGA – Health Facilities v2.0. 2024. https://doi.org/10.7916/kv1n-0743.

18. GRID3. Geospatial data Nigeria. https://grid3.org/geospatial-data-nigeria. Accessed 8 Oct 2026.

19. Federal Ministry of Health and Social Welfare. Nigeria health sector-wide joint annual review 2025: speech by Muhammad Ali Pate. 2025. https://health.gov.ng/nigeria-health-sector-wide-joint-annual-review-2025-abuja-nigeria-12-14-november-2025-speech-by-muhammad-ali-pate-the-coordinating-minister-of-health-and-social-welfare-federal-republic-of-nigeria/. Accessed 8 Oct 2026.

20. World Bank. Population, total—Nigeria, 2025 (SP.POP.TOTL). World Development Indicators. 2026. https://api.worldbank.org/v2/country/NGA/indicator/SP.POP.TOTL?date=2025&format=json. Accessed 8 Oct 2026.

21. World Bank. Birth rate, crude—Nigeria, 2024 (SP.DYN.CBRT.IN). World Development Indicators. 2026. https://api.worldbank.org/v2/country/NGA/indicator/SP.DYN.CBRT.IN?date=2024&format=json. Accessed 8 Oct 2026.

22. Federal Ministry of Health and Social Welfare of Nigeria, National Population Commission, ICF. Nigeria demographic and health survey 2024. 2025. https://www.dhsprogram.com/pubs/pdf/FR395/FR395.pdf. Accessed 8 Oct 2026.

23. Huangfu Q, Hall JAJ. Parallelizing the dual revised simplex method. Math Program Comput. 2018;10:119–142. https://doi.org/10.1007/s12532-017-0130-5.