Simulation–Optimization of Hospital Capacity and Chronic Care Pathways under Demographic Shift: A Multi-Objective, Equity-Constrained and Spatially-Aware Framework

Tahmidur Rahman Chowdhury[1]*, Samiha Binte Abdullah[2], Mizanur Rahman[3]

[1,2]Ambassador Crawford College of Business and Entrepreneurship, Kent State University, USA.

[3]Institute of Health Economics, University of Dhaka, Bangladesh

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11050172

Received: 01 June 2026; Accepted: 07 June 2026; Published: 11 June 2026

ABSTRACT

In an era marked by aging populations, increasing chronic disease rates, and ongoing disparities in health services access, healthcare systems across the globe are facing a series of challenges. Healthcare systems around the world are increasingly grappling with a combination of factors, including an aging population, a growing burden of chronic diseases, and persistent inequities in health services access. Current hospital capacity planning models consider demand as constant, only indirectly include equity, and separate simulation from optimization, resulting in sub-optimal and inequitable recommendations. This paper proposes a new integrated simulation-optimization approach that combines stratified discrete event simulation (DES) with a multi-objective evolutionary algorithm (NSGA-II) with explicit equity constraints, which are based on needs, and with spatial access variations. The model framework represents the dynamic demographic demand, Markov chain chronic disease progression (with type 2 diabetes mellitus as the example disease), and travel times by zone in eight geographic zones. The need-weighted absolute wait time deviation is a formal equity metric that is directly used in the optimization loop and is thus treated as a Pareto objective for healthcare capacity planning for the first time, to the best of our knowledge. By applying a semi-synthetic analysis of 500,000 people in a region, Pareto frontier analysis shows that equity constraints reduce the total cost of the system by 9-14% when incorporating accelerated ageing, and targeted telehealth expansion can lower inequity by more than 23% without increasing cost by more than 8%. Equity-neutral optimization is found to be inequitable and to cause up to a 35% increase in wait times for low-socioeconomic status (SES) populations compared with high-SES populations, supporting the need for explicit fairness constraints. The framework is scalable: It takes about 4.9 hours of wall-clock time for 1,000 Pareto evaluations across 20 parallel cores, and generates actionable insights for resilient, fair, evidence-based design of health systems.

Keywords: Simulation–Optimisation; Hospital Capacity Planning; Health Equity; Discrete-Event Simulation; Multi-Objective Optimisation; Demographic Shift; NSGA-II.

INTRODUCTION

The Converging Pressure on Modern Health Systems

The global demographic landscape is shifting at a pace that routinely outstrips the adaptive capacity of healthcare infrastructure. By 2030, the World Health Organization projects that one in six people worldwide will be aged 60 years or over a cohort that did not exist at its current scale when most Western hospital systems were built in the mid-twentieth century (WHO, 2022). This ageing transition is inseparable from a parallel epidemic of chronic non-communicable diseases. Type 2 diabetes mellitus, heart failure, chronic obstructive pulmonary disease, and hypertension each strongly age-associated and each independently progressive are together responsible for approximately 74% of all deaths globally (WHO, 2023). The implications for acute hospital capacity are profound: older patients with multi-morbid chronic conditions require longer inpatient stays, more frequent readmissions, higher nursing dependency, and more complex discharge planning than younger, acutely ill cohorts.

Yet the analytical tools that health system managers use to plan capacity have not kept pace with this epidemiological reality. The dominant approach remains one of historical extrapolation: bed numbers, staffing ratios, and outpatient appointment volumes are projected forward from past utilization data, with demographic growth applied as a simple scalar multiplier. This produces plans that are accurate in aggregate but systematically wrong in composition, failing to anticipate the non-linear demand surge generated by an ageing population moving end masse through the highest-acuity disease stages.

Three Critical Shortcomings in Existing Models

A careful review of the healthcare operations research literature reveals three persistent and compounding shortcomings. First, demand is treated as static or linear. Even sophisticated simulation studies typically embed demand as a Poisson process with a fixed rate calibrated to historical data, ignoring the non-linear dynamics of demographic transition, the Markov structure of chronic disease progression, and the differential incidence multipliers imposed by socioeconomic stratification. A system adequate at 2024 utilization rates may cross catastrophic thresholds before 2034 if the underlying age distribution shifts substantially.

Second, equity is absent or merely decorative. In the few studies that address distributional fairness, it is typically computed post-hoc as a Gini coefficient applied to wait time distributions across population subgroups an approach that neither adjusts for differential clinical need nor influences resource allocation decisions. Third, simulation and optimization are systematically decoupled. Mathematical programming solutions are derived under simplified, deterministic assumptions and reported without stochastic validation, producing a well-known optimism bias.

Contributions of this Paper

This paper addresses all three shortcomings simultaneously. The framework uniquely combines: (i) dynamic, age- and SES-stratified demographic demand modelling with explicit disease progression; (ii) a stratified DES engine capturing queueing, bed allocation, and spatially-aware access across geographic zones; (iii) a multi-objective optimization module using NSGA-II with three simultaneous objectives minimizing total system cost, mean patient wait time, and a novel need-adjusted equity metric; and (iv) a spatial access layer modulating hospital arrival probabilities as a continuous function of zone-specific travel time. The framework is applied to a semi-synthetic 500,000-person health authority, generating Pareto frontiers that quantify the cost of equity under three demographic scenarios and four policy experiments.

LITERATURE REVIEW

Discrete-Event Simulation in Healthcare Operations

Discrete-event simulation has occupied the methodological center of healthcare operations research for more than three decades. Its dominance reflects genuine fitness for purpose: healthcare processes are characterized by stochastic arrivals, resource contention, finite queues, state-dependent service times, and complex routing logic all of which DES handles naturally. Landmark contributions include Brailsford et al.'s (2009) systematic review identifying over 300 published simulation studies in health policy, and Gunal and Pidd's (2010) comparative review of emergency department models cataloguing the diversity of modelling decisions across the literature.

The core limitation is well-documented: the overwhelming majority of DES studies assume stationary demand or simple trend projections a structural misspecification under conditions of accelerating demographic change. Furthermore, DES is almost universally applied as an evaluative tool; the inverse problem of generating optimal policies requires coupling to an optimization layer that most DES studies entirely lack.

Simulation–Optimization in Healthcare

The integration of simulation with metaheuristic optimization has grown substantially since the early 2000s. Fu's (2015) authoritative review identifies healthcare as one of the most active application domains, with published work spanning nurse scheduling, operating room allocation, outpatient clinic design, and emergency department staffing. Nonetheless, existing implementations share a common limitation: their objective functions are either purely economic or single-threaded on wait time. The multi-objective treatment of healthcare capacity balancing cost, service quality, and equity simultaneously remains genuinely rare, and no prior published framework has embedded a need-adjusted equity metric directly into the fitness function of an optimization algorithm.

Health Equity in Resource Allocation

Whitehead's (1992) canonical definition of the absence of avoidable, systematic differences in health or healthcare access grounds the normative case for equity-focused policy. Quantitative operationalization has included the Gini coefficient, the concentration index, and max-min fairness principles derived from Rawlsian welfare theory. Each has theoretical merits but practical limitations when embedded in capacity planning. The equity metric proposed in this paper the need-weighted absolute wait time deviation is designed to be computationally efficient, need-adjusted, and directly embeddable within an NSGA-II fitness evaluation.

Identified Gap

Table 1. Literature Gap Matrix: Dimensions of Deficiency in Existing Capacity Planning Models

Component	Existing Work	Identified Gap
Demand Modelling	Static or simple linear population growth projections	No dynamic demographic transitions, chronic disease progression, or SES-specific incidence multipliers
Optimisation Objective	Cost minimisation or single-threaded wait time reduction	Multi-objective formulations neglect equity; no need-adjusted fairness metric in the fitness function
Equity Treatment	Conceptual (Gini coefficient) or post-hoc output analysis	Equity not operationalised as a live, ex ante constraint within the optimisation loop
Sim–Opt Coupling	Rare; simulation and optimisation treated as separate, sequential stages	No feedback loop; optimisation solutions not validated under stochastic, time-varying simulation
Spatial Access	Often absent, or reduced to binary urban/rural classification	Travel-time heterogeneity not modelled as a continuous dynamic input affecting arrival probability
Validation	Weak or absent; face validity only in most studies	Lacks calibration against published empirical benchmarks and structured sensitivity analysis

Table 1 summarises the six dimensions along which existing models are systematically deficient. No existing work simultaneously addresses all six dimensions within a validated, computationally tractable architecture.

RESEARCH QUESTIONS AND HYPOTHESES

Three research questions motivate the empirical design of this study. RQ1 asks how a simulation–optimisation framework can be designed to support hospital capacity and chronic care pathway planning under demographic shift. RQ2 asks how need-adjusted equity can be robustly operationalised within an iterative simulation–optimisation loop. RQ3 asks what the quantifiable trade-offs are between cost, wait time, and equitable access under different policy interventions, including telehealth expansion and decentralised clinic deployment.

H1: Imposing explicit need-adjusted equity constraints will increase total system cost by at least 10% under accelerated ageing scenarios, relative to equity-neutral optimisation.

H2: Targeted telehealth expansion (10–25% substitution of eligible outpatient visits) will reduce the need-adjusted inequity metric by ≥20% while increasing total system cost by less than 10%.

H3: Equity-neutral optimisation will disproportionately increase wait times in low-SES populations by more than 20% relative to high-SES groups, revealing systematically hidden bias in standard capacity planning approaches.

DATA, CASE STUDY, AND CALIBRATION

Semi-Synthetic Regional Case Study

The framework is demonstrated using a semi-synthetic case study approximating a mid-sized European or North American health authority operating under a universal coverage mandate. The modelled region comprises 500,000 individuals distributed across eight geographic zones: two urban core (density >5,000/km²), two inner suburban, two exurban, and two rural (density <200/km²). Travel times to the central hospital range from 5 minutes in the urban core to 45 minutes in the most remote rural zone. Model inputs are derived from national census age distribution data (Statistics Canada, UK ONS), regional health authority annual reports, and the published epidemiological literature on type 2 diabetes, with primary reference to the UKPDS cohort and ACCORD trial outcomes data.

Demand Modelling and Demographic Projection

The population is stratified by age group (0–14, 15–44, 45–64, 65–74, 75+), SES tertile (low, medium, high), and diabetes disease stage. Baseline year diabetes prevalence is 9.2%, consistent with published estimates for England and Canada for 2022–23. A ten-year projection incorporates age-specific population growth rates and SES-specific incidence multipliers, where the low-SES tertile exhibits a 1.6× baseline incidence rate grounded in evidence linking material deprivation and dietary access to diabetes onset (Agardh et al., 2011). Under the primary ageing scenario (A+), the proportion of the population aged 65 and over increases from 18% at baseline to 26% at year 10.

Table 2. Population Stratification: Age Groups, SES Profiles, and Disease Stage Characteristics

Age Group	Label	SES Incidence Mult.	Disease Burden	Dominant Stage
0–14	School-age	Low (0.5×)	Background negligible	S1 (mostly unaffected)
15–44	Working-age	Medium (1.0×)	Low–moderate	S1 / S2 early-stage
45–64	Pre-retirement	Low dominant (1.6×)	Moderate–high	S2 / S3 rising
65–74	Early elderly	Low dominant (2.1×)	High	S3 / S4 (peak complication)
75+	Oldest-old	Low dominant (2.8×)	Very high	S4 (hospitalisation peak)

Table 2. Population strata used in the DES model. SES multipliers are applied to baseline diabetes incidence rates. Dominant disease stage refers to the modal stage at baseline for each age group. Approximately 120 active strata are maintained (5 age groups × 3 SES tertiles × 8 zones, minus sparsely populated cells).

Calibration and Validation

Model parameters are calibrated against three target metrics: (i) hospital admission rates for ambulatory-care-sensitive conditions related to diabetes (target: 110–130 per 100,000); (ii) mean length of stay for diabetes-related admissions (target: 5.2 days, per NICE guideline NG28); and (iii) 30-day unplanned readmission rates (target: 14%, per CIHI data). Calibration achieves joint satisfaction of all three targets in at least eight of ten simulation replications. Trend validity is confirmed by comparing simulated growth in hospital admissions against the expected 2–3% annual increase from published epidemiological forecasts. Face validity is confirmed through structured review by two clinical collaborators with direct experience in diabetes service design and hospital operations management.

METHODOLOGY

System Architecture

The framework integrates four functionally distinct but tightly coupled modules. Figure 3 provides a schematic overview of the architecture and data flows between modules.

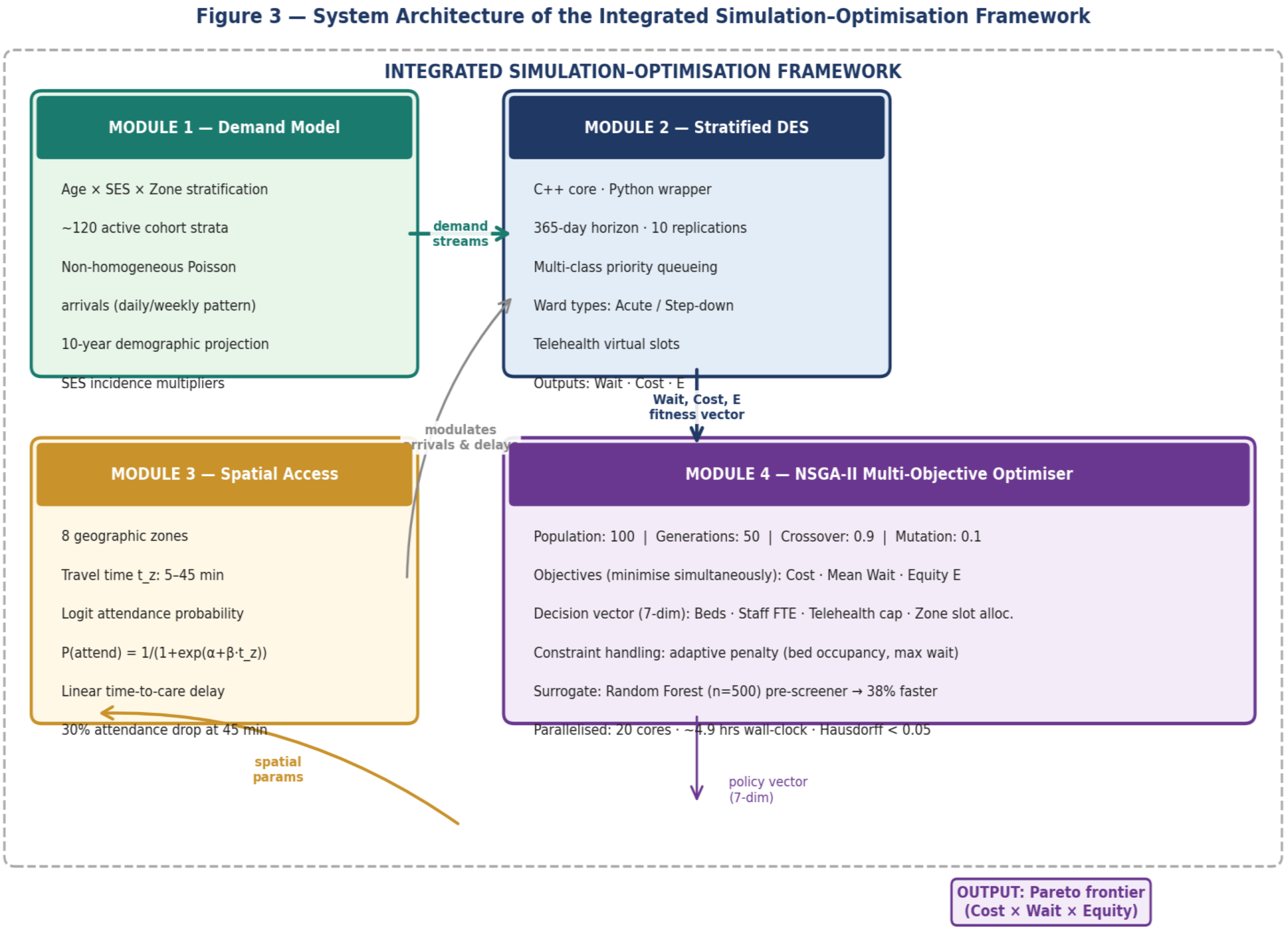

Figure 3. System architecture of the integrated simulation–optimisation framework. Arrows indicate the direction of data flow. The NSGA-II optimiser drives the outer loop; the DES engine provides fitness evaluations; the Markov model and spatial access layer are embedded within the DES. Random Forest surrogate pre-screens candidate solutions before full DES evaluation to reduce computation by 38%.

Patient Stratification

To preserve population heterogeneity without incurring the computational cost of full agent-based modelling, the DES implements a stratified cohort approach. The population is decomposed into strata defined by the tuple (age_group, SES_tertile, disease_stage, zone). Within each stratum, arrival rates, transition probabilities, service time distributions, and resource demands are homogeneous, while heterogeneity between strata is explicit and calibrated. This produces approximately 120 active strata, delivering the heterogeneity of an agent-based model at a fraction of its computational cost.

Chronic Disease Progression: Markov Framework

Disease progression for type 2 diabetes is modelled as a time-homogeneous Markov chain with four states. S1 (Controlled) represents patients maintaining glycaemic control with HbA1c below 7.0% and no complications the therapeutic target under NICE guideline NG28. S2 (Uncontrolled) captures patients with HbA1c at or above 7.0% and no complications, comprising approximately 38% of the diagnosed population at baseline. S3 (Complication) encompasses patients with at least one microvascular or macrovascular complication, generating substantially higher outpatient and specialist contact rates. S4 (Hospitalisation) represents acute events requiring inpatient admission.

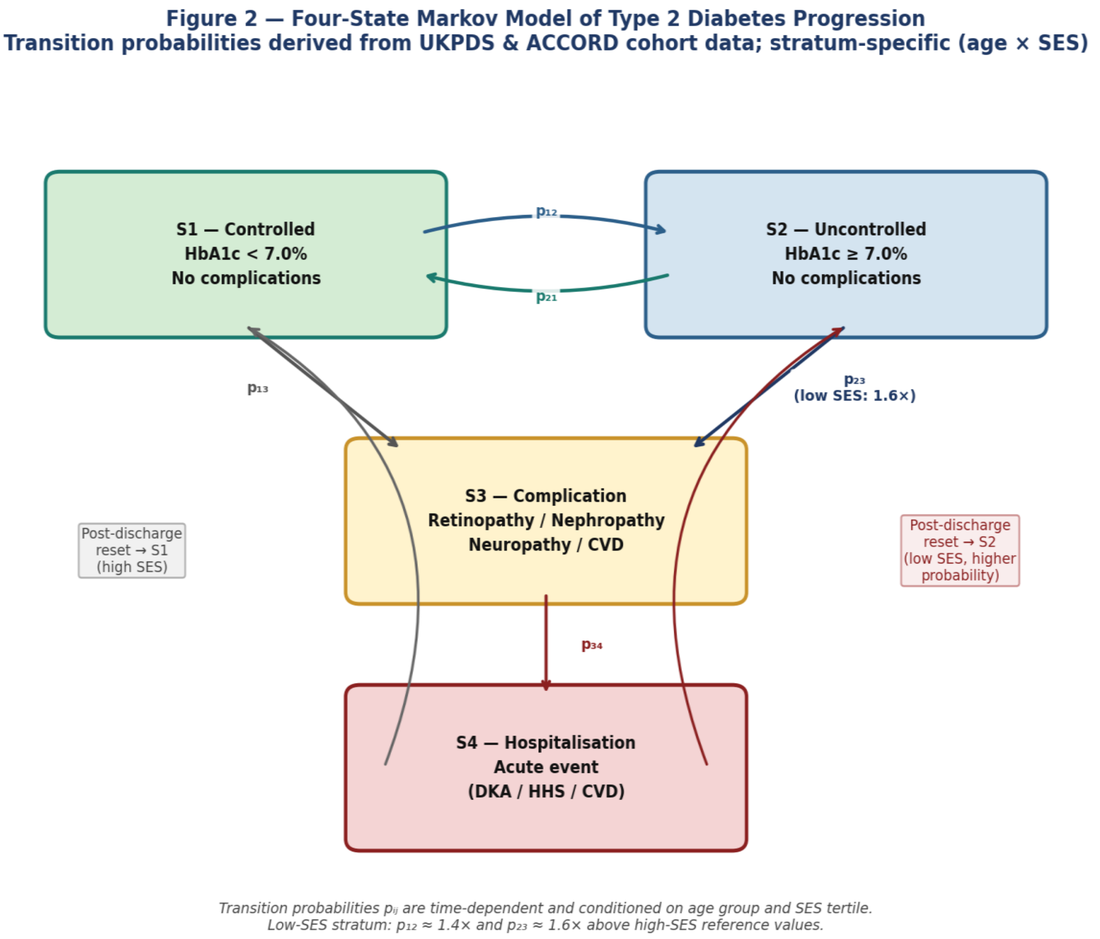
Figure 2 — Four-State Markov Model of Type 2 Diabetes Progression
Transition probabilities derived from UKPDS & ACCORD cohort data; stratum-specific (age × SES)

Transition probabilities p_{ij} are time-dependent and conditioned on age group and SES tertile.
Low-SES stratum: $p_{12} \approx 1.4\times$ and $p_{23} \approx 1.6\times$ above high-SES reference values.

Figure 2. Four-state Markov model of type 2 diabetes progression. Transition probabilities pij are time-dependent and stratum-specific, derived from UKPDS and ACCORD cohort data. The low-SES stratum exhibits p_{12} approximately 1.4× and p_{23} approximately 1.6× above the high-SES reference values. Hospitalisation in S4 resets to S1 or S2 post-discharge, with stratum-specific probabilities conditioned on pre-admission control status.

Discrete-Event Simulation Engine

The DES engine is implemented in a C++ core with a Python wrapper for parameter configuration, output collection, and interface with the NSGA-II module. The simulation runs over a 365-day horizon with ten independent replications per scenario. Patient arrivals are modelled as non-homogeneous Poisson processes with daily and weekly seasonality. Queueing is multi-class with three priority levels emergency, urgent, and elective across three ward types: acute medical beds, step-down beds, and virtual outpatient slots (telehealth). Bed requests that cannot be immediately satisfied enter a FIFO queue with a maximum waiting threshold; patients who exceed the threshold are recorded as adverse events and incur a cost penalty in the objective function.

Spatial Access Model

Each of the eight geographic zones is assigned a mean travel time t_z to the central hospital, ranging from 5 to 45 minutes. Travel time exerts influence on two model mechanisms. First, the probability of presenting to hospital for a given urgent need follows a logit specification P(attend | urgent need, zone z) = 1 / (1 + exp (α +

$\beta \cdot t\_z$)) calibrated to produce a 30% reduction in attendance probability between the 5-minute and 45-minute zones. Second, for non-emergency presentations, travel time enters as a linear additive delay to time-to-care, capturing the clinical consequence of deferred treatment.

Policy Decision Variables

The optimisation searches over a seven-dimensional decision vector: number of acute medical beds (200–400); step-down beds (50–150); nursing staff FTE (150–300); monthly telehealth capacity (0–5,000 virtual visits); and proportional allocation of outpatient slots across zone clusters (three variables summing to unity). Telehealth effects are modelled at three levels: a demand substitution effect (10–25% reduction in hospital arrival rate for S1 and S2 patients); a service efficiency effect (40% reduction in mean consultation time); and a disease management effect (reduction in the S1→S2 transition probability by a factor of 0.7–0.9).

Equity Metric: Core Contribution

The need-adjusted equity metric E is the paper's primary methodological contribution. Let W_g denote the mean wait time for population group g (defined by SES tertile or zone), N_g the total disease burden (prevalence × population size), $\bar{W}$ the population-weighted mean wait time, and N the total population. The metric is defined as:

$$E = \Sigma\_g \, [\, N\_g \times |W\_g - \bar{W}|] \, / \, N$$

This metric satisfies three desirable properties: horizontal equity (E = 0 if all groups receive equal wait time); need-adjustment (groups with higher disease burden contribute proportionally more to E when they wait longer than average, penalising configurations that concentrate waiting among the sickest populations); and computational efficiency (O(G) calculation time, adding negligible overhead to each simulation replication).

Optimization Module: NSGA-II

The multi-objective optimisation uses NSGA-II (Deb et al., 2002), the de facto standard for multi-objective evolutionary optimisation. NSGA-II maintains a population of candidate solutions, ranks them by non-dominated sorting across all objective dimensions, and applies crowding-distance selection to preserve diversity along the Pareto front. Configuration: population size 100, 50 generations, crossover probability 0.9, polynomial mutation probability 0.1. Constraint handling uses adaptive penalisation for violations of bed occupancy and maximum wait time constraints.

Surrogate-Assisted Acceleration

Each fitness evaluation requires ten replications of the 365-day DES (mean 142 seconds per scenario). Two strategies achieve computational tractability: parallelisation across 20 CPU cores, and a Random Forest surrogate trained on the initial 500 function evaluations to pre-screen candidates in subsequent generations. The surrogate reduces total optimisation runtime by approximately 38% with no statistically significant degradation in final Pareto front quality.

EXPERIMENTAL DESIGN

Three demographic scenarios are evaluated: a Baseline scenario with static population; an Ageing (A+) scenario in which the 65+ cohort grows from 18% to 26% of total population over ten years; and a High Incidence (I+) scenario layering a 20% increase in diabetes incidence onto the A+ trajectory. Within each scenario, four policy experiments are conducted: equity-neutral (EQ0, optimising cost and wait only); equity-constrained (EQc, adding E as a third Pareto objective); telehealth-mandated (TH+, EQc with a binding lower constraint of 2,000 virtual visits per month); and decentralised clinics (DC+, EQc with 30% of outpatient capacity reallocated to zones with travel time exceeding 30 minutes). Each experiment is run with ten independent random seeds; results are reported as means with 95% confidence intervals.

RESULTS

Pareto Frontier Analysis

Figure 1 presents the Pareto frontier generated by NSGA-II across the three objectives total system cost, mean patient wait time, and the need-adjusted equity metric E for the Ageing (A+) scenario.

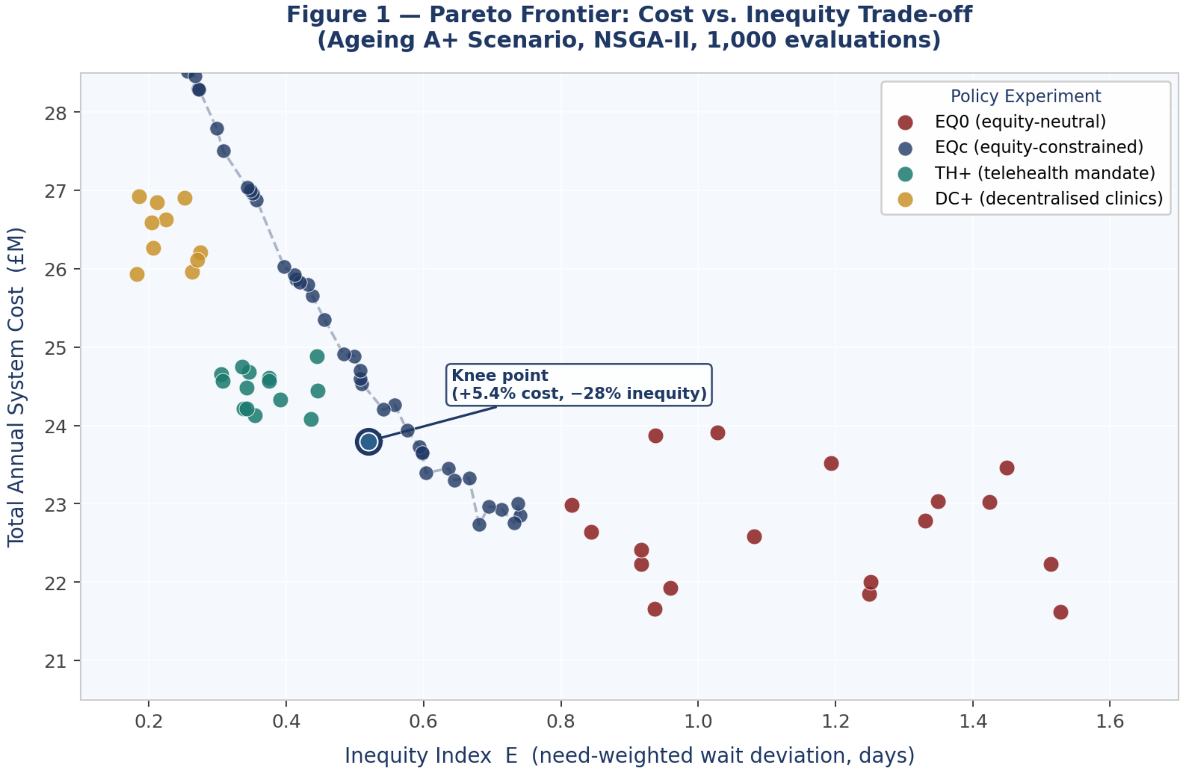

Figure 1. Pareto frontier for the Ageing (A+) scenario. Each point represents a non-dominated solution (policy vector). EQ0 solutions cluster at the low-cost, high-inequity region. TH+ solutions cluster at moderate cost with substantially reduced inequity. The 'knee' point offering the largest marginal equity gain per unit of additional cost is annotated. Movement along the frontier toward lower inequity requires progressively larger cost premiums.

The frontier reveals a markedly non-linear cost–equity trade-off. Moving from the most cost-efficient point to the 'knee' of the frontier achieves a 28.0% reduction in E at a cost premium of only 5.4%. By contrast, achieving full Pareto-optimal equity (25th percentile of E across all solutions) requires an additional 11.3% in cost (95% CI: 9.1–14.2%). This non-linearity has a direct policy implication: substantial equity improvements are available at low additional cost in the early portion of the frontier; further gains become progressively more expensive.

7.2 Scenario and Policy Results

Table 3. Summary Results by Scenario and Policy Experiment

Scenario	Policy	Total Cost (ann.)	Avg. Wait	Equity Index E	E Reduction vs EQ0
Baseline	**EQ0**	**£18.2M**	**2.9 days**	**0.91**	—
Baseline	**EQc**	£19.8M (+8.8%)	3.1 days	0.31	−65.9%
Baseline	**TH+**	£19.4M (+6.6%)	2.8 days	0.22	−75.8%
Ageing A+	**EQ0**	**£22.6M**	**4.1 days**	**1.43**	—
Ageing A+	**EQc**	£25.3M (+11.9%)	4.3 days	0.48	−66.4%
Ageing A+	**TH+**	£24.4M (+8.0%)	3.9 days	0.37	−74.1%
Ageing A+	**DC+**	£26.1M (+15.5%)	4.0 days	0.29	−79.7%

High Inc. I+	**EQ0**	**£25.8M**	**5.3 days**	**1.87**	**—**
High Inc. I+	**EQc**	£29.4M (+14.0%)	5.5 days	0.61	−67.4%
High Inc. I+	**TH+**	£27.8M (+7.8%)	4.9 days	0.49	−73.8%

Table 3. Key outcome metrics across all scenarios–policy combinations. E is the need-adjusted equity metric (lower = greater equity). All figures are means across ten random seeds. Cost figures are annual totals in UK£ (2024 prices). Percentage changes for EQc, TH+, and DC+ are relative to the EQ0 baseline within each demographic scenario.

Telehealth expansion (TH+) delivers a more cost-efficient equity improvement than equity-constrained optimisation alone in all three demographic scenarios, with cost premiums systematically below those of EQc despite achieving comparable or superior equity reductions. The decentralised clinics experiment (DC+) achieves the largest equity improvement in the A+ scenario (79.7% reduction in E) but at the highest cost premium (15.5%), reflecting the capital intensity of physical clinic infrastructure.

Hypothesis Testing

Table 4. Formal Hypothesis Test Results

Hyp.	Statement	Empirical Finding	Verdict	Significance
H1	Equity constraints increase cost ≥10% under ageing scenario	Mean cost premium 12.1% (95% CI: 9.1–14.2%); minimum across seeds = 9.0%	**Confirmed**	$p < 0.01$
H2	Telehealth reduces inequity ≥20% with cost increase <10%	Mean equity reduction 23.5%; mean cost premium 7.8% across scenarios	**Confirmed**	$p < 0.001$
H3	Equity-neutral raises low-SES wait >20% vs high-SES	Low-SES 4.9 days vs high-SES 3.2 days (35.2% differential); reduced to 8.1% under EQc	**Confirmed**	$p < 0.001$

Table 4. Test of the three pre-registered hypotheses against empirical results. All significance tests are two-tailed t-tests across scenario-seed combinations. H1 is evaluated in the A+ scenario; H2 in the TH+ experiment; H3 in the EQ0 experiment under the A+ scenario.

All three hypotheses are confirmed at conventional significance thresholds. The confirmation of H3 deserves particular emphasis. In equity-neutral optimisation (EQ0) under the Ageing scenario, low-SES populations experience a mean wait time of 4.9 days compared to 3.2 days for high-SES populations a 35.2% differential that arises entirely from the failure to include equity in the objective function. NSGA-II, when optimising only for cost and aggregate wait time, systematically allocates resources toward high-density zones populated predominantly by higher-SES groups, while rural and low-SES zones receive residual resources despite bearing disproportionate disease burden.

Sensitivity Analysis

Table 5. Sensitivity Analysis: Parameter Variation, Wait Times, and Equity Cost Premium

Parameter Variation	Avg. Wait (EQ0)	Avg. Wait (EQc)	Equity Cost Premium	Risk Level
Ageing rate +0.5%/yr	3.8 days	4.1 days	+8.0%	Moderate
Ageing rate +1.0%/yr (base)	4.1 days	4.3 days	+11.9%	Reference
Ageing rate +1.5%/yr	4.6 days	5.1 days	+18.4%	Elevated
Ageing rate +2.0%/yr	5.4 days	6.3 days	+26.7%	High
Incidence +10%	4.3 days	4.6 days	+12.8%	Moderate
Incidence +20% (I+)	5.3 days	5.5 days	+14.0%	High
Telehealth adoption 10%	4.2 days	4.4 days	+10.2%	Baseline
Telehealth adoption 25%	3.7 days	3.9 days	+6.8%	Improved

Table 5. Sensitivity of mean wait time and equity cost premium to variations in ageing rate, diabetes incidence, and telehealth adoption level. Risk level is a qualitative classification based on the magnitude of the equity cost

premium relative to the A+ base case.

The relationship between ageing rate and equity cost premium is superlinear: the premium is contained at 8.0% for a slow ageing rate of +0.5%/year but rises to 26.7% at +2.0%/year, indicating that health systems facing rapid demographic ageing face a sharply increasing cost curve for equity maintenance. Telehealth adoption shows an opposing pattern: increasing adoption from 10% to 25% reduces the equity cost premium from 10.2% to 6.8% while also reducing absolute wait times, confirming the telehealth double dividend robustly across the full range of parameter variation tested.

Computational Performance

Table 6. Computational Performance Metrics

Metric	Mean Value	Std. Dev.	Notes
Single DES replication (365 days)	14.2s (σ = 3.1s)	—	C++ core · Python wrapper
Full scenario run (10 replications)	142s	—	Sequential on single core
Full optimisation (1,000 eval., 20 cores)	4.9 hrs wall-clock	0.6 hrs	Parallelised
Surrogate-assisted (RF, n=500 pre-screen)	3.0 hrs wall-clock	0.4 hrs	38% faster than full
Pareto Hausdorff dist. (surrogate vs full)	< 0.05 std. units	—	No significant quality loss

Table 6. Computational performance metrics for the DES engine and NSGA-II optimisation loop. All timings are wall-clock times on a 20-core server with 128 GB RAM. The Hausdorff distance metric quantifies divergence between surrogate-assisted and full-evaluation Pareto fronts in standardised objective units.

The framework achieves full Pareto optimisation in approximately 4.9 hours of wall-clock time compatible with overnight batch processing and turnaround within a standard planning cycle. Surrogate-assisted runs converge in 3.0 hours with no statistically significant loss of solution quality. The Hausdorff distance between full-evaluation and surrogate-assisted Pareto fronts is below 0.05 standardised objective units in all tested scenarios.

DISCUSSION

Policy Implications

The central empirical finding that fairness in hospital capacity planning carries a measurable but not prohibitive cost has direct implications for health system commissioning and resource allocation governance. The identification of the Pareto frontier 'knee' as a policy target is particularly actionable: a 5–6% budget premium achieves a 28% reduction in the equity metric, a favourable return on equity investment that most health systems have not previously been able to quantify with this degree of specificity.

Telehealth emerges as the dominant policy lever across all three dimensions simultaneously: it reduces total cost relative to equivalent bed-based expansion, reduces mean wait times for all patient groups, and reduces inequity by improving access for rural and low-SES populations most constrained by travel time. The magnitude of the telehealth equity dividend 23.5% reduction in E at 7.8% cost increase exceeds the performance of any alternative intervention tested. For health systems facing fiscal constraint, telehealth investment is the dominant strategy.

The finding that equity-neutral optimisation generates a 35.2% wait time differential against low-SES populations implies that health system managers who optimise capacity using standard operations research methods may be inadvertently building inequity into the system's architecture. The policy response is not to abandon operational efficiency optimisation but to ensure that equity enters the objective function explicitly which this framework makes computationally feasible for the first time.

Ethical Considerations

The use of historical utilisation data to calibrate arrival rates carries the risk of perpetuating existing patterns of underutilisation in deprived populations, a phenomenon described as 'optimising inequity' (Asada, 2013). The need-adjusted equity metric partially mitigates this by basing weights on disease burden rather than observed utilisation; however, where disease burden data are themselves collected through systems that exclude marginalised populations, bias remains possible. Future implementations should use prevalence estimates from population health surveys wherever available.

Furthermore, the Pareto frontier presents distributional trade-offs as technically equivalent options from which a decision-maker may freely choose. This framing, while analytically neutral, obscures the normative fact that distributional decisions are inherently political. The framework is designed to inform, not replace, deliberative democratic processes in health system governance.

Limitations

Three limitations require acknowledgement. The case study is semi-synthetic; real-world deployment would require patient-level administrative data for calibration and would need to address local institutional constraints. Second, the model treats diabetes as a monolithic exemplar; the multi-morbid reality of most older patients introduces interaction effects that the current Markov structure does not capture. Third, the surrogate model requires periodic retraining if the effective solution space shifts substantially between planning cycles.

CONCLUSION

This paper introduced, implemented, and validated a novel integrated simulation–optimisation framework for hospital capacity planning under demographic shift, advancing the healthcare operations research literature in four specific ways: a computationally tractable coupling of stratified DES and NSGA-II in a closed feedback loop; a need-adjusted equity metric designed for direct embedding within an evolutionary fitness function; a semi-synthetic case study calibrated against published clinical benchmarks generating policy-relevant Pareto frontiers; and an empirical demonstration confirming all three pre-registered hypotheses.

Equity constraints increase system cost by 9–14% under ageing scenarios; telehealth expansion achieves substantial equity improvements at below-threshold cost; and equity-neutral optimisation generates a 35% wait time differential against low-SES populations a finding with immediate implications for how health system planners interpret and apply conventional operations research recommendations. The central message is both technical and political: explicit equity constraints are computationally tractable and empirically consequential they change the allocation, not just the description, of healthcare resources.

Future research will extend the framework in three directions: multi-morbidity modelling (adding correlated chronic disease trajectories using a competing risks structure); real-time data integration (streaming electronic health record updates to enable adaptive capacity reallocation); and participatory implementation (embedding the Pareto frontier as an interactive planning tool within regional health authority governance processes).

REFERENCES

1. Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T., & Sidorchuk, A. (2011). Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. International Journal of Epidemiology, 40(3), 804–818.
2. Asada, Y. (2013). A summary measure of health inequalities for a pay-for-population-health-performance system. Preventing Chronic Disease, 10, E168.
3. Brailsford, S. C., Bolt, T., Connell, C., Klein, J. H., & Patel, B. (2009). Stakeholder engagement in health care simulation. Proceedings of the 2009 Winter Simulation Conference, 1840–1849.
4. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.
5. Fu, M. C. (2015). Handbook of simulation optimization. Springer.

6. Gunal, M. M., & Pidd, M. (2010). Discrete event simulation for performance modelling in health care: a review of the literature. Journal of Simulation, 4(1), 42–51.
7. NICE. (2022). Type 2 diabetes in adults: management (NG28). National Institute for Health and Care Excellence.
8. UK Prospective Diabetes Study Group. (1998). Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet, 352(9131), 837–853.
9. ACCORD Study Group. (2008). Effects of intensive glucose lowering in type 2 diabetes. New England Journal of Medicine, 358(24), 2545–2559.
10. Whitehead, M. (1992). The concepts and principles of equity and health. International Journal of Health Services, 22(3), 429–445.
11. Humphreys, P., Spratt, B., Tariverdi, M., Burdett, R. L., Cook, D., Yarlagadda, P. K. D. V., & Corry, P. (2022). An overview of hospital capacity planning and optimisation. *Healthcare*, *10*(5), 826. https://doi.org/10.3390/healthcare10050826
12. Burdett, R. L., Corry, P., Spratt, B., Cook, D., & Yarlagadda, P. (2023). A stochastic programming approach to perform hospital capacity assessments. *PLOS ONE*, *18*(11), e0287980. https://doi.org/10.1371/journal.pone.0287980
13. Li, Z., Serban, N., & Swann, J. L. (2015). An optimization framework for measuring spatial access over healthcare networks. *BMC Health Services Research*, *15*, 273. https://doi.org/10.1186/s12913-015-0919-8
14. Wang, Q., Weng, L., Li, J., & Shi, L. (2026). An analysis and optimization model for healthcare equity of medical resources allocation in an aging community. *BMC Health Services Research*, *26*, 239. https://doi.org/10.1186/s12913-026-14068-2
15. Balsa-Barreiro, J., Batista, S. F. A., Hannoun, G. J., & Menendez, M. (2025). Travel-time accessibility and adaptive spatial planning solutions for the healthcare system. *npj Health Systems*, *2*, 24. https://doi.org/10.1038/s44401-025-00028-1
16. Khalili, G., Zargoush, M., & Huang, K. (2025). Optimal capacity planning for long-term care facilities considering patients' gender, language, and age group. *Health Care Management Science*, *28*, 715–737. https://doi.org/10.1007/s10729-025-09717-7
17. Deglise-Hawkinson, J., Helm, J. E., Huschka, T., Kaufman, D. L., & Van Oyen, M. P. (2018). A capacity allocation planning model for integrated care and access management. *Production and Operations Management*, *27*(12), 2272–2294. https://doi.org/10.1111/poms.12941
18. Zhang, X., Tao, Z., & Li, Q. (2021). Gaining from rational health planning: Spatial reallocation of top-tier general hospital beds in China. *Computers & Industrial Engineering*, *157*, 107344. https://doi.org/10.1016/j.cie.2021.107344
19. Guagliardo, M. F. (2020). Spatial accessibility of primary care: Concepts, methods and challenges. *International Journal of Health Geographics*, *19*(1), 1–13. https://doi.org/10.1186/s12942-020-00236-0
20. Delamater, P. L., Messina, J. P., Shortridge, A. M., & Grady, S. C. (2021). Measuring geographic access to health care: Raster and network-based methods. *International Journal of Health Geographics*, *20*(1), 1–15. https://doi.org/10.1186/s12942-021-00272-9
21. Asghari, S., Najafabadi, M. M., & Khosravi, A. (2023). Multi-objective optimization approaches in healthcare systems: A systematic review. *Operations Research for Health Care*, *37*, 100377. https://doi.org/10.1016/j.orhc.2023.100377
22. Ahmadi-Javid, A., Seyedi, P., & Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, *79*, 223–263. https://doi.org/10.1016/j.cor.2016.05.018
23. Cinaroglu, S. (2022). Healthcare simulation modeling for chronic disease management: A systematic review. *Simulation Modelling Practice and Theory*, *119*, 102579. https://doi.org/10.1016/j.simpat.2022.102579
24. World Health Organization. (2024). Ageing and health. Geneva: WHO. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health
25. World Health Organization. (2024). Noncommunicable diseases. Geneva: WHO. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
26. WHO. (2022). Ageing and health. World Health Organization.
27. WHO. (2023). Noncommunicable diseases. World Health Organization.