

Under what conditions could an AI-first accountable care organization improve outcomes in Medicaid? A microsimulation and economic evaluation

Sanjay Basu

sanjay.basu@ucsf.edu

University of California, San Francisco

Rajaie Batniji

Waymark

Research Article

Keywords: Medicaid, Managed care, Accountable care organization, Artificial intelligence, Microsimulation, Health economic evaluation, Administrative costs, Health equity

Posted Date: June 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9903473/v1>

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

[Read Full License](#)

Additional Declarations: Competing interest reported. Dr Basu receives grants from the NIH and CDC, has patents from Collective Health and Waymark, receives salary from HealthRIGHT360 and Waymark, and serves on the board of Waymark, all outside the submitted work. Dr Batniji receives salary from Waymark, has patents from Collective Health, and sits on the board of Waymark and Collective Health, all outside the submitted work. The authors hold financial interests in organizations that could benefit from the adoption of AI-mediated Medicaid delivery; this potential source of motivated reasoning is discussed in the Limitations and is intended to be addressed through independent prospective evaluation. No other competing interests are declared.

Abstract

Background

Medicaid managed care organizations (MCOs) enroll more than 70% of 48.3 million beneficiaries, with mixed evidence on outcomes and administrative loss ratios averaging 7.9% in 2023 and 10.1% in 2024. State and federal agencies are examining whether artificial intelligence (AI) could enable accountable care organization (ACO)–style coordination at lower administrative cost. We aimed to identify the conditions under which an AI-first Medicaid ACO—replacing private managed care administrative functions with AI—could reduce hospitalizations, emergency department (ED) visits, and per-member-per-month (PMPM) costs, or worsen them.

Methods

We combined a structured multi-agent AI debate (Modified Delphi; 8 agents, 12 parameter domains) to derive a consensus organizational design with a Monte Carlo microsimulation (1,000-iteration probabilistic sensitivity analysis [PSA]; 7 scenarios) of 75,043 Medicaid adults aged 19–64 from the American Community Survey Public Use Microdata Sample (2019–2023). The design specified AI-first virtual primary care handling 58% of encounters with physician oversight and administrative overhead reduced from 8.4% to 3.0% of premium. Primary outcomes were hospitalizations and ED visits per 1,000 person-years; secondary outcomes were PMPM cost, HEDIS gap closure, and engagement. The model was validated against 5 independent natural experiments. Reporting followed CHEERS-AI.

Results

Hospitalizations decreased 15 per 1,000 person-years (95% uncertainty interval [UI], 7–24) and ED visits 39 (95% UI, 20–63); HEDIS gap closure rose from 35% to 50% and PMPM cost fell \$41 (95% UI, 16–76). Administrative automation alone accounted for \$27 PMPM (66% of savings) under zero AI clinical efficacy, and 98.8% of PSA iterations projected net savings from administrative reform alone. The simulation reproduced observed effects in 15 of 15 comparisons. Net savings remained positive when AI clinical effects were discounted to 50% of human-trial benchmarks (\$33 PMPM) and at administrative rates up to 7% (\$18 PMPM).

Conclusions

Administrative overhead reduction—most of it structural to nonprofit versus for-profit managed care rather than AI-dependent—was the dominant driver of projected benefit (Spearman $\rho = 0.82$). An administrative cost rate below 5–6% of premium is a measurable, prospectively testable viability

threshold for Section 1115 waivers, whereas digital access inequity is the primary condition under which an AI-first model could worsen disparities.

Background

Medicaid managed care organizations (MCOs) enroll more than 70% of 48.3 million beneficiaries [1]. Systematic reviews find mixed effects on quality and access [2], mandatory enrollment increased spending without quality gains [3], and managed care has been associated with widened racial disparities in infant health [4]. For-profit MCOs face structural incentives that can be misaligned with patient outcomes [5]: prior authorization and utilization review reduce expenditure but produce no net quality improvement [2, 3], and AI augmentation of utilization review within current MCO structures raises additional concerns about algorithmic opacity and unintended access barriers [6]. Across Medicaid managed care plans nationally, administrative loss ratios averaged 7.9% in 2023 and 10.1% in 2024 in the largest industry survey (n = 184 plans) [7], with statutory medical loss ratio requirements permitting administrative-plus-margin shares up to 15% of premium under federal rules [8]. These administrative shares thus function partly as a care-limitation mechanism rather than a coordination investment in many plans. Nearly 40% of acute care visits are for ambulatory care-sensitive conditions varying 13-fold across counties [9], fewer than half of physicians accept new Medicaid patients [10], and more than half of listed providers are unreachable or not accepting patients [11]. These patterns suggest structural constraints on outcome improvement regardless of clinical tools layered onto existing MCO structures [2, 3, 4].

Accountable care organizations (ACOs) address this differently—replacing utilization management with coordinated, capitated care and aligning incentives with outcomes. The published ACO evidence is mixed and incompletely settled. In Medicare, the Medicare Shared Savings Program (MSSP) produced gross spending reductions of 1% to 2% in early performance years [12], but a 2023 analysis estimated that across performance years 2013–2021 the MSSP was associated with net losses to the Centers for Medicare & Medicaid Services (CMS) of between \$775 million and \$2.063 billion after accounting for bonus payments [13]. Where MSSP savings did occur, they were not concentrated among high-risk patients in whom care coordination would be expected to have its largest effect, with one analysis attributing only 38% of 2012-entrant savings to high-risk patients [14]. Longer-tenure ACOs with experience under newer benchmarking rules have shown more sustained net savings [15], and structurally distinct ACOs—Medicare Pioneer ACOs [16] and the Massachusetts Alternative Quality Contract [17]—achieved larger early spending reductions. In Medicaid, the evidence base is thinner; Oregon’s coordinated care organizations (CCOs) reduced spending 7% through global budgets [18], but a recent scoping review of Medicaid ACOs (2012–2023) characterized the literature as limited in number and heterogeneous in effect [19]. Human-delivered care coordination interventions—community health workers [20], enhanced care management—have reduced hospitalization in Medicaid populations, but the cost of the outreach, care management, and administrative workforce required for coordination-intensive ACO models has limited Medicaid adoption [19, 21].

The federal government and state Medicaid agencies are now actively examining whether AI could enable accountable care–style coordination at lower administrative cost, particularly for administrative processes that are formula-driven and high-volume. The 2024 federal rule on advancing interoperability and prior authorization (CMS-0057-F) extended prior authorization application programming interface and reporting requirements to state Medicaid fee-for-service, Medicaid managed care, and the Children’s Health Insurance Program [22]; the Medicaid and CHIP Payment and Access Commission (MACPAC) initiated a study of automation in Medicaid prior authorization in 2026 [23]; a Kaiser Family Foundation analysis cataloged federal and state consumer protections around AI in prior authorization and claims review and noted that state Medicaid agencies can require contracted plans to disclose AI use [24]; an Urban Institute review of 895 Medicaid agency documents across 45 states found agencies publishing limited detail about AI use, with eligibility determination and improper payment prevention named as primary application areas [25]; and state-level initiatives include California’s Medi-Cal Connect population health management infrastructure designed for AI-mediated member interaction [26]. A complementary administrative-process framing notes that Medicaid eligibility determination, redetermination, and improper-payment prevention are formula-driven processes well suited to AI augmentation [27]. Purpose-built AI clinical tools have also demonstrated technical feasibility: a trial improved hypertension concordance by 15.2 percentage points ($n = 12,137$) [28], a conversational AI was rated superior to physicians on 30 of 32 evaluation axes [29], and large language models have reached expert-level diagnostic performance [30]. General-purpose consumer AI tools, however, under-triaged 52% of emergency presentations [31], indicating that clinical safety may depend on purpose-built architecture with mandatory escalation protocols [32].

The relevant policy questions are what an AI-first Medicaid ACO—an entity adopting ACO-style accountability and capitation while replacing MCO administrative functions with AI—could look like in practice, and under what conditions it could improve patient outcomes, particularly for populations with limited digital access. Because no such organization exists, empirical evaluation is premature; a preclinical approach is to specify an organizational design and evaluate its projected outcomes against calibrated population data with external validation. Designing such a model requires concurrent optimization across clinical, regulatory, financial, and ethical constraints; multi-agent debate—AI instances with distinct expert constraints iterating toward consensus—provides a structured approach to this multi-domain design problem and generates minority-report parameters that bound uncertainty in resulting simulation scenarios [33, 34]. This study adapts this framework to derive a consensus AI-first ACO specification, then evaluates the conditions under which it could improve patient outcomes through Monte Carlo microsimulation calibrated to national Medicaid benchmarks and validated against 5 independent natural experiments.

Methods

Study Design and Reporting

This study combined structured AI-assisted organizational design with Monte Carlo microsimulation evaluation (Fig. 1). Reporting follows the Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022)[35] and its extension for interventions that use artificial intelligence (CHEERS-AI), which retains all 28 CHEERS 2022 items and adds AI-specific reporting items [36].

Multi-Agent Debate

Multi-agent debate is a method in which multiple AI instances, each constrained to a distinct expert perspective, independently propose solutions and then iteratively revise toward consensus—a process that has improved reasoning quality over single-model outputs in prior work[33, 34] and that, to the knowledge of the authors, has not previously been applied to multi-domain healthcare delivery system design. Eight AI agents were instantiated with expert personas (eAppendix A, eTable 1) representing dimensions MACPAC identifies as critical for Medicaid managed care evaluation[37]—access, quality, cost, and beneficiary experience—along with regulatory, ethical, and organizational feasibility. Each agent independently designed a complete AI ACO across 12 parameter domains (eTable 2): clinical model, payment structure, provider reimbursement, organizational structure, regulatory pathway, quality framework, AI clinical architecture, human oversight, social determinants of health (SDOH) integration, rural/urban design, anti-monopoly provisions, and ethical governance. Nine domains correspond to CMS requirements for managed care approval and 1115 waiver evaluation [38]; 3 address challenges specific to AI-delivered care.

The output of the debate is a single fully specified organizational design with numeric values for every parameter (eTable 2): an AI-first virtual primary care model handling 58% of encounters with physician oversight at a 1:45 supervision ratio; full capitation at \$438 per member per month (PMPM) with two-sided risk; provider rates at 125% of Medicare for primary care and 110% for hospital services; nonprofit organizational structure with charter lock on for-profit conversion; Section 1115 demonstration waiver as regulatory pathway; 3-model ensemble AI architecture with mandatory escalation for high-acuity conditions; quarterly algorithmic bias audits; community health workers at 2.5 per 1,000 members; and administrative overhead targeted at 3.0% of premium. This specification was input directly into the microsimulation as the consensus AI ACO scenario.

Agents followed a Modified Delphi process[39]: after independent proposals (Round 0), each agent received anonymized group summary statistics with its own position highlighted, and agents outside the interquartile range were required to justify or revise. Domain convergence required a coefficient of variation (CV) below 0.15 for quantitative parameters and agreement of at least 75% for categorical parameters. The debate terminated when 10 of 12 domains met convergence criteria for 2 consecutive rounds, or at a maximum of 12 rounds. Agents diverging by more than 1 standard deviation from consensus on any parameter submitted structured minority reports identifying the parameters on which they dissented and the direction (higher or lower) of their preferred value. These minority reports parameterized the pessimistic (S3) and optimistic (S2) microsimulation scenarios: S3 used the lower bound of dissenting positions for engagement multipliers and AI clinical efficacy (1.1× outreach, 1.2×

agreement) while holding administrative reform at the consensus value (3.0%), reflecting the patient-advocate and physician minority view that AI engagement and clinical effects would underperform consensus expectations even if administrative reform succeeded; S2 used the upper bound (1.3× outreach, 1.4× agreement, 2.5% admin). All agents used the same underlying AI model (Claude Sonnet, Anthropic); results were robust to varying model temperature settings (eAppendix B).

Study Population

The study population comprised 75,043 Medicaid adults aged 19 to 64 from the American Community Survey (ACS) Public Use Microdata Sample (PUMS), 2019–2023, representing approximately 6.8 million non-institutionalized Medicaid adults nationally (eTable 3). Digital access was modeled using race-by-metropolitan-status broadband penetration rates from Pew Research Center and Federal Communications Commission data [40], ranging from 45% for American Indian/Alaska Native (AIAN) adults in nonmetropolitan areas to 87% for White adults in metropolitan areas.

Microsimulation Model

A microsimulation was chosen rather than a closed-form decision-analytic model for three reasons. First, the engagement cascade and digital-divide penalty operate at the individual level conditional on race-by-metropolitan-status broadband access, which can be represented faithfully only with individual-level heterogeneity. Second, equity stratification by race × metropolitan status × risk tier produces 24 subgroups, several with small sample sizes that require Monte Carlo aggregation to yield stable estimates. Third, parameter uncertainty propagates through nonlinear interactions (e.g., admin-rate × utilization interactions) that closed-form expectations would not capture, and probabilistic sensitivity analysis (PSA) requires individual-level simulation to compute correlation-based parameter-importance metrics.

The microsimulation modeled each individual through 5 sequential stages (Fig. 1): (1) Digital Access, based on broadband and technology adoption rates by race and metropolitan status [40]; (2) Care Engagement, with sequential probabilities of outreach, agreement, engagement, and adherence, calibrated from AHRQ data with a 50% engagement penalty for individuals without broadband access [41, 42]; (3) Clinical Utilization, using risk-stratified hospitalization, ED, and primary care rates from MEPS and HCUP, modified by relative risk reductions anchored to clinical decision support evidence [28]; (4) Preventive Care Quality via HEDIS gap closure from NCQA benchmarks; and (5) Equity, with condition detection and documentation probabilities by race reflecting differential claims visibility [43].

All parameters were drawn from empirically calibrated uncertainty distributions in each of 1,000 PSA iterations (eTable 4). Proportions, rates, costs, and detection probabilities used Beta, Gamma, and Normal distributions, respectively (eTable 4). Stage 1 used Pew and FCC broadband data [40]; Stage 2, AHRQ NHQDR; Stage 3, MEPS and HCUP Statistical Briefs; Stage 4, NCQA; Stage 5, Obermeyer et al [43]. AI engagement multipliers and clinical relative risk reductions were anchored to published human-delivered intervention evidence (community health worker outreach [20], algorithmic care-gap targeting

[44], ACO-attributed hospitalization reductions[15, 16]); the assumption that AI delivery achieves the same effect size as human delivery is a strong one that the AI-to-human discount-factor sensitivity analysis (eTable 5) was designed to test directly. Administrative overhead was modeled at 8.4% for MCOs and 3.0% for the AI ACO, consistent with traditional (non-MCO) Medicaid administrative rates of 2% to 4% [45, 46].

Scenarios and Validation

Seven scenarios were compared (eTable 6): status quo MCO (S0), AI ACO consensus (S1), optimistic (S2), and pessimistic (S3); care coordination without AI (S4); AI ACO under universal eligibility (S5); and administrative reform only (S6), which applied AI administrative cost reduction while holding clinical parameters at status quo values to isolate administrative from clinical effects. The status quo was calibrated to MACPAC PMPM costs [37], HCUP/MEPS utilization rates, NCQA HEDIS scores, and Milliman administrative loss ratios across 184 Medicaid MCOs[7] (eTable 7).

Validation was conducted as out-of-sample backtesting against 5 independent natural experiments spanning diverse delivery reform mechanisms: Oregon CCOs (global budgets in Medicaid) [18], Medicare Pioneer ACOs (shared savings with intensive care management) [16, 47], the IMPaCT community health worker randomized trials (intensive community engagement and SDOH navigation in Medicaid) [20], the Medicare Shared Savings Program (weak shared savings with modest care coordination) [12], and Comprehensive Primary Care Plus (CPC+; primary care transformation without payment reform; selected as a negative control because it tested practice transformation without the payment reform that distinguishes ACOs) [48]. For each reform, intervention parameters were set to match the actual reform characteristics—engagement multipliers, utilization relative risk reductions, administrative cost rates, system-wide spillover effects, and (for CHW) non-utilization cost effects—while disabling AI-specific features (broadband penalty, equity gap closure, AI detection improvement). The microsimulation was run for 200 PSA iterations per reform. Projected percent changes in spending, hospitalizations, and ED visits were compared against observed effects from each reform’s published evaluation. Validation coverage was defined as the proportion of observed effect estimates falling within the simulated 95% UI (eTable 8).

Welfare and Equity Analysis

Net social welfare was computed using a Hicks-Kaldor framework[49] as the sum of three components (eAppendix C): consumer surplus (CS), comprising monetized quality-adjusted life year (QALY) gains from utilization changes at a base-case willingness-to-pay (WTP) of \$100,000 per QALY; government surplus (GS), equal to total PMPM savings annualized (decomposed into administrative and medical subcomponents); and producer surplus (PS), reflecting net primary care volume changes. WTP sensitivity analyses ranged from \$50,000 to \$200,000 per QALY (eTable 9).

Equity analysis proceeded through two mechanisms. First, all primary outcomes were stratified by race and ethnicity and by metropolitan status simultaneously, producing 8 race-by-metropolitan-status

subgroups. Second, welfare gains were weighted using an Atkinson inequality-aversion index[50] at aversion parameters of 0.5, 1.0, and 2.0, giving greater weight to improvements among lower-income and historically disadvantaged populations (eAppendix D). The 50% broadband engagement penalty for individuals without broadband access was the primary modeled mechanism by which AI-first delivery could exacerbate disparities.

Provider Rate, Discount Factor, and Parameter Importance Analyses

Provider reimbursement rate sensitivity was modeled through an explicit causal pathway from payment rate to network adequacy to patient outcomes (eAppendix G; eTable 10) [51]. Provider availability was parameterized as a logistic function of reimbursement as a percentage of Medicare, normalized to 1.0 at the consensus rate of 125%. Only non-AI encounters were affected; the 58% AI virtual encounter share was rate-independent. A two-way encounter share × payment rate sensitivity analysis showed positive net savings across all combinations tested (eTable 11).

To address the strong assumption that human-trial intervention effect sizes translate fully to AI delivery, an AI-to-human evidence-translation discount factor sensitivity analysis was conducted (eTable 5). For discount factors $d \in \{0, 0.10, 0.20, 0.30, 0.40, 0.50\}$, the AI engagement multipliers were rescaled as $1 + (\text{multiplier} - 1) \times (1 - d)$ and the AI clinical relative risk reductions were rescaled as $RR \times (1 - d)$, holding administrative reform unaffected because administrative automation is mechanical rather than behavioral. An AI-administrative-rate sensitivity analysis was also conducted at AI ACO administrative cost rates from 3.0% to 10.0% of premium (eTable 12), holding all other AI parameters at consensus.

Parameter importance for PMPM savings was assessed via Spearman rank correlation of paired iteration-level differences (status quo minus AI ACO) across 1,000 PSA iterations, identifying which inputs most strongly determined outcomes under uncertainty. The Monte Carlo standard error for PMPM savings was \$0.52 (CHEERS 2022 item 20). All analyses used ACS person weights. Code is publicly available (see Availability of data and materials).

Results

Debate Convergence

In Round 0, agents agreed immediately on regulatory pathway (100%, Section 1115 waiver), provider rates (all within a 120–125% of Medicare range), and AI architecture (100%, multi-model ensemble with mandatory human escalation). The most contested domains—measured by variation across agent proposals—were human oversight intensity, clinical model scope, and rural/urban delivery design (eAppendix A). By Round 2, all 12 parameter domains reached consensus and 11 of 12 showed low cross-agent variation ($CV < 0.15$), with human oversight the only domain retaining meaningful disagreement ($CV = 0.23$)—reflecting substantive disagreement about physician supervision ratios (eFigure 1). Seven of eight agents filed structured minority reports; the patient-advocate agent had the

most divergent parameters (27 of 60), primarily related to higher community health worker staffing, lower AI encounter share, and stronger patient consent requirements (eAppendix A).

AI ACO Consensus Design

The consensus specified (eTable 2) AI-first virtual primary care handling 58% of encounters with physician oversight at a 1:45 supervision ratio; a 3-model ensemble architecture with mandatory escalation for high-acuity conditions; full capitation at \$438 PMPM with two-sided risk sharing; community health workers at 2.5 per 1,000 members; FQHC referral for all physical examinations; mandatory quarterly algorithmic bias auditing; an independent ethics board with binding halt authority; and prohibition on for-profit conversion. Provider reimbursement was set at 125% of Medicare for primary care and 110% for hospital services, funded by administrative overhead reduction from 8.4% to 3.0%. The consensus design assumes that core MCO backend functions that are not easily automated—state reporting and compliance, provider network contracting and credentialing, grievance and appeals (with statutory due-process requirements), actuarial review, and executive functions—remain in scope, with AI displacing volume-driven administrative processing (claims adjudication, prior authorization, eligibility verification, care management triage) but not displacing these governance functions. The 3.0% administrative target therefore represents the residual cost of these retained functions plus AI deployment overhead (model validation, ongoing monitoring, vendor contracting, drift detection, regulatory review); this target is below the 5.2% tenth-percentile of Milliman 2024 Medicaid MCO administrative loss ratios[7] and within the 2% to 4% range of traditional (non-MCO) Medicaid agency administrative rates [45, 46].

Validation

The status quo scenario calibrated within published ranges for all 5 metrics (eTable 7). In backtesting against 5 independent natural experiments, the observed effect fell within the simulated 95% UI in 15 of 15 comparisons (Fig. 2). The mean calibration ratio was 0.99 and mean absolute error was 1.1 percentage points. The model reproduced both large reform effects (Oregon CCO: observed – 7.0% spending, simulated – 5.4%; CHW trials: observed – 9.0% hospitalization, simulated – 10.1%) and near-null effects (CPC+: observed 0.0% spending change, simulated – 0.2%). For Medicare backtests (Pioneer ACO, MSSP, CPC+), simulation tested whether the engagement cascade and utilization-reduction mechanisms reproduce relative percent changes through shared reform mechanisms rather than absolute Medicaid-specific rates (eTable 8).

Patient Outcomes

Under consensus design parameters, the simulation projected a decrease in hospitalizations of 15 per 1,000 person-years (95% UI, 7 to 24) and in ED visits of 39 per 1,000 person-years (95% UI, 20 to 63) for the AI ACO scenario (Table 1). HEDIS gap closure was projected to increase from 35% to 50%. These projections were driven by algorithmically prioritized outreach and 24/7 virtual availability, which the model estimated would increase engagement from 12% to 21%. Under pessimistic design assumptions,

the model projected attenuated but directionally consistent improvements (\$35 PMPM savings, 10 fewer hospitalizations per 1,000). Across racial groups, the simulation projected slightly larger absolute reductions among Black and AIAN adults (17 per 1,000) than White (14 per 1,000) and Hispanic adults (15 per 1,000). The Black-White hospitalization gap was projected to narrow from 40 to 37 per 1,000 (8% reduction) and the AIAN-White gap from 51 to 48 (6% reduction). Race-by-metropolitan-status stratification showed projected improvements in all 8 subgroups with no widening of within-group gaps (eAppendix D; eFigure 2; eTable 13).

Table 1
Microsimulation Results: Patient Outcomes and Costs Across Scenarios

Outcome	Status Quo MCO (S0)	AI ACO Consensus (S1)	AI ACO Pessimistic (S3)	Admin Reform Only (S6)	Care Coord. Only (S4)
Patient outcomes					
Hospitalizations per 1,000 PY (95% UI)	191 (147–238)	176 (135–220)	181 (140–228)	191 (147–240)	181 (139–228)
ED visits per 1,000 PY (95% UI)	702 (551–867)	663 (519–817)	678 (528–839)	702 (548–865)	681 (533–839)
PCP visits per 1,000 PY (95% UI)	3,429 (2,709–4,236)	3,549 (2,808–4,370)	3,550 (2,795–4,366)	3,429 (2,702–4,231)	3,600 (2,835–4,443)
HEDIS gap closure, % (95% UI)	35 (26–45)	50 (41–60)	42 (36–49)	35 (26–45)	37 (27–47)
Engagement rate, % (95% UI)	12 (10–15)	21 (12–33)	21 (13–33)	12 (10–15)	14 (11–17)
Cost and administration					
PMPM, \$ (95% UI)	478 (377–593)	437 (344–543)	443 (348–548)	451 (356–559)	458 (360–567)
PMPM savings, \$ (95% UI)	–	41 (16–76)	35 (11–70)	27 (3–63)	21 (12–35)
Admin costs, % of premium (95% UI)	8.4 (6.0–14.5)	2.9 (1.0–6.0)	2.9 (1.0–6.0)	2.9 (1.0–6.0)	5.9 (4.0–10.2)
Values are means from 1,000 probabilistic sensitivity analysis iterations. Uncertainty intervals (95% UI) represent the 2.5th–97.5th percentiles. PMPM savings are computed as paired iteration-level differences. Administrative cost rates shown are PSA iteration means; modeled target values were 8.4% for the status quo MCO and 3.0% for AI ACO scenarios—iteration-level sampling produces means of 8.4% and 2.9%, respectively. Care Coord. Only denotes current MCO with added community health workers and care managers but no AI. ACO indicates accountable care organization; ED, emergency department; HEDIS, Healthcare Effectiveness Data and Information Set; MCO, managed care organization; PCP, primary care physician; PMPM, per member per month; PY, person-years.					

PMPM costs were projected to fall by \$41 (95% UI, \$16 to \$76), from \$478 to \$437 (Table 1; full results across all scenarios in eTable 14). Decomposed by source, the administrative-reform-only scenario (S6) projected a PMPM reduction of \$27 (95% UI, \$3 to \$63) under zero AI clinical efficacy assumptions, accounting for 66% of total projected savings; the remaining \$14 PMPM (34%) was attributable to projected AI clinical effects operating through the engagement cascade (eAppendix F). The administrative reduction reflected automation of claims adjudication, prior authorization, eligibility verification, and care management triage, reducing overhead from 8.4% to 2.9% of premium [7, 45]. The probability that administrative reform alone would produce positive net savings was 98.8% across 1,000 PSA iterations.

The administrative-rate sensitivity analysis (eTable 12) extended these findings to administrative cost rates between 3.0% and 10.0% of premium, holding all other AI parameters at consensus. Probability of net savings was 100% at AI administrative rates of 3.0% through 7.0% (PMPM savings \$37 to \$18), declined to 96.5% at 8.0% (\$14 PMPM savings), and fell below 75% at 9.0% (probability 73.5%, \$8 PMPM). At the empirical 5.2% tenth-percentile of Milliman 2024 Medicaid MCO administrative loss ratios [7], projected PMPM savings was \$28 with 100% probability of net savings.

The discount-factor sensitivity analysis (eTable 5) tested the assumption that AI delivery achieves the same effect size as human-delivered interventions. With AI clinical effects discounted to 80% of human-trial benchmarks, total PMPM savings remained \$36 (12 fewer hospitalizations per 1,000); at 70%, \$35 PMPM (\$25 from administrative reform unchanged, \$10 from attenuated clinical effects, and 11 fewer hospitalizations per 1,000); at 50%, \$33 PMPM and 8 fewer hospitalizations per 1,000. Across the full range, projected savings remained positive because administrative reform is mechanical and was not discounted.

Welfare and Conditions for Improvement

Net social surplus was projected at \$646 per member per year, comprising consumer surplus from health gains (\$151), government savings (\$490), and producer surplus (\$5; eFigure 3 shows the welfare decomposition graphically). Net surplus was positive at all WTP thresholds tested (\$50,000 to \$200,000 per QALY; eTable 9). Across the 7 scenarios, the AI ACO Consensus and Universal scenarios dominated on most performance dimensions in radar comparison (eFigure 4). The parameter importance analysis identified administrative cost rate as the dominant driver of PMPM savings—more strongly correlated with savings than any clinical parameter (Spearman $\rho = 0.82$ versus combined $\rho < 0.25$ for AI clinical parameters; eAppendix F). Figure 3 shows the probability of net improvement by administrative cost rate: at 3%, probability was 100%; at 6%, 78%; at 7%, 61%; and at 8% (the current MCO baseline), 45%.

Discussion

The policy implication of these findings is structural: for an AI-first delivery model in Medicaid, administrative cost reduction is a more reliable lever than AI clinical performance—and, as the decomposition below shows, much of that reduction is structural to organizational form rather than

contingent on AI capability—and it is assessable through a measurable threshold rather than through clinical trial evidence. Administrative cost rate was more strongly correlated with PMPM savings (Spearman $\rho = 0.82$) than any AI clinical parameter (combined $\rho < 0.25$), and 98.8% of PSA iterations projected positive net savings from administrative reform alone. Compared with human care coordination added to the current MCO (S4: projected \$21 PMPM savings, 10 fewer hospitalizations per 1,000 person-years), the AI ACO scenario projected \$41 PMPM savings and 15 fewer hospitalizations—a difference attributable primarily to administrative overhead reduction rather than clinical superiority over human coordinators. The microsimulation’s causal structure reproduced observed effects across 5 independent natural experiments (100% coverage, 15 of 15 comparisons), supporting the plausibility of the engagement-cascade and utilization-reduction mechanisms. The administrative-cost reduction that drives most projected savings, however, is a mechanical accounting relationship rather than a behavioral effect and was not among the mechanisms these experiments tested; its feasibility is examined directly below rather than inferred from the backtests. The Medicaid-specific backtests (Oregon CCO, CHW trials) reproduced both the direction and magnitude of observed reforms, and 3 of 5 backtests used Medicare populations where the mechanisms tested are not Medicare-specific.

The projected \$41 PMPM savings (8.6%) substantially exceed Medicare ACO savings (1 to 3%) [12, 15], reflecting a different mechanism—administrative infrastructure replacement rather than clinical optimization within existing overhead. The administrative-reform-only scenario isolates this: \$27 PMPM in projected savings derives from reducing overhead from 8.4% to 2.9% with no assumed AI clinical efficacy, while leaving hospitalizations and ED visits unchanged from the status quo (Table 1)—confirming that projected utilization reductions operate through the engagement cascade and are not attributable to administrative overhead reduction. The multi-agent debate framework produced an internally consistent specification across 12 interdependent domains; unlike prior work using AI for clinical decision support, this approach addressed concurrent regulatory, financial, and ethical constraints that determine delivery model feasibility. Conflicting hard constraints produced quantifiable disagreement across multiple domains, with human oversight retaining a CV of 0.23 through Round 2.

Administrative Cost Feasibility and Backend Functions

The reduction of administrative overhead from 8.4% to approximately 3% of premium is the load-bearing input behind most projected savings, and no private Medicaid managed care organization has achieved it: in the largest industry survey of Medicaid MCO financial results ($n = 184$ plans, 2024 data), the composite administrative loss ratio was 10.1% of premium, the Medicaid-focused subset 7.7%, the lowest decile approximately 5.2%, and no plan publicly reported sub-3.0% overhead (eTable 15) [7]. The target nonetheless warrants decomposition, because most of the gap is structural rather than contingent on AI: the administrative cost of an AI-first ACO can be built up from three components (eTable 16). First, a substantial share of private Medicaid MCO overhead reflects functions that a nonprofit, charter-locked public-benefit ACO would not incur at all: marketing, broker commissions, profit and risk margin, and shareholder returns. An analysis of 2023 regulatory filings found that administrative spending and profit in states with Medicaid managed care exceeded that in states without managed care by approximately

\$250 per person per year—roughly 4 percentage points of premium at the modeled capitation rate—a difference attributable to managed-care administration and profit rather than to clinical care [52]. Consistent with this, traditional fee-for-service Medicare operates at administrative overhead of approximately 2%—roughly one-sixth that of private Medicare Advantage plans—and the documented rise in U.S. administrative spending has been driven specifically by the growth of private managed care within Medicare and Medicaid, indicating that publicly administered coverage carries far lower overhead without AI [45]. Non-managed-care Medicaid administration is correspondingly low-overhead—on the order of a few percent of program spending, far below private MCO levels—reflecting the same public-administration advantage [45]. Removing for-profit managed-care overhead therefore moves the cost structure toward these public-administration levels before any automation is applied.

Second, the residual volume-driven processing functions—claims adjudication, prior authorization, eligibility verification, claim-status inquiry, and care-management triage—are formula-driven, high-volume, and already substantially automatable with existing technology. As of 2024, only 35% of medical prior authorizations were conducted fully electronically, and full automation of the administrative transactions tracked in the industry index represented an estimated \$20 billion annual savings opportunity [53]; the 2024 federal interoperability and prior-authorization rule (CMS-0057-F) further mandates prior-authorization application programming interfaces for Medicaid [22]. These are precisely the functions the consensus design assigns to AI. Third, the functions that cannot be displaced by AI under current regulation form the residual floor: state reporting and CMS compliance documentation, provider network contracting and credentialing, grievance and appeals with statutory due-process requirements, actuarial certification, executive and board functions, and AI deployment overhead (model validation, ongoing monitoring, vendor management, drift detection, and regulatory review). The approximately 3% target represents this retained floor plus deployment overhead.

This decomposition reframes the feasibility question. The 3% target remains aggressive and unattained by any private MCO, but the conclusion of net savings does not depend on reaching it: the administrative-rate sensitivity analysis (eTable 12) shows a high probability of net savings at every overhead rate materially below the 8.4% status quo—100% probability through 7% overhead, 96.5% at 8%, and \$23 PMPM in projected savings at 6%. The relevant viability question is therefore not whether an AI-first ACO can reach an implausible 3% but whether a nonprofit entity can operate below current private MCO overhead—a level already achieved by public Medicaid administration (2–4%) and approached by the most efficient existing MCOs (5.2%), largely without AI.

Human-to-AI Evidence Translation

The microsimulation’s engagement multipliers and clinical relative risk reductions were anchored to evidence from human-delivered interventions (community health worker trials, ACO evaluations, clinical decision support studies in patient-facing settings), and the AI-to-human discount-factor sensitivity analysis (eTable 5) tests how robust projected outcomes are to attenuated AI efficacy. With AI clinical effects discounted to 50% of human-trial benchmarks—a strong discount approximating the lower bound

of plausible AI clinical performance under deployment-imperfect conditions—projected PMPM savings remain \$33 (8 fewer hospitalizations per 1,000 person-years), still 80% of consensus PMPM savings, because administrative reform is mechanical and not discounted. This robustness is a structural feature of the model: clinical interventions in the engagement cascade can fail, but administrative automation either achieves the targeted overhead reduction or it does not. The asymmetric robustness suggests that early Section 1115 demonstrations should be designed to test the administrative thesis first (an observable cost ceiling that can be verified within the first year) before drawing strong conclusions about AI clinical performance, which requires multi-year clinical endpoint follow-up.

The threshold analysis identifies a measurable administrative target: net improvement with high probability when overhead remains below 5% to 6%—above current traditional Medicaid agency administrative rates (2 to 4%) but well below current MCO overhead (8 to 13%) [7, 8, 45].

Digital Access Inequity and Limits of Delivery Reform

Digital access inequity is the primary condition under which the model projects harm: AIAN beneficiaries in nonmetropolitan areas, where broadband penetration was 45%, face a 50% engagement penalty that limits projected benefit precisely for the most underserved populations (eTable 13). Projected racial hospitalization gap narrowing (6% to 8%) appeared only where digital access was adequate—and even then these improvements operate downstream of the structural determinants that drive disparities: income, housing instability, food insecurity. Delivery system reform cannot close gaps rooted in those determinants; digital access subsidies represent the most proximate modifiable lever, not a solution to underlying structural inequity. Demonstration designs should pair AI-first models with broadband subsidies, mobile-unit coverage in low-broadband areas, and a phone-based clinical pathway with parity standards to avoid concentrating clinical benefit among the digitally connected.

Implementation Pathway

These projections provide a design envelope for prospective evaluation rather than a deployment recommendation. A Section 1115 demonstration should track administrative cost rate as the primary process measure, with hospitalizations and ED visits per 1,000 person-years as co-primary patient outcome endpoints. Digital access rates stratified by race and metropolitan status, care engagement, and HEDIS gap closure should serve as secondary endpoints to test whether the engagement cascade assumptions central to clinical savings hold in practice. Section 1115 waivers offer the appropriate regulatory pathway (eAppendix E); the administrative cost ceiling provides a performance standard for state Medicaid agencies' waiver applications seeking AI-mediated delivery alternatives to private managed care.

Limitations

This study has several limitations. Most importantly, the dominant driver of projected savings—the reduction of administrative overhead from 8.4% to approximately 3% of premium—is a modeled

assumption rather than an externally validated result. The five natural-experiment backtests validate only the second-order clinical and utilization cascade; no U.S. Medicaid managed care organization has publicly demonstrated overhead below approximately 5%, and whether AI can achieve the required automation is the central empirical uncertainty. The decomposition above shows that most of the assumed reduction is structural to nonprofit organizational form rather than dependent on AI capability, and the sensitivity analysis shows the directional conclusion is robust across administrative rates well below the status quo; nonetheless, the magnitude depends on this assumption, which is the primary rationale for tracking administrative cost as the prospectively testable primary measure of the proposed Section 1115 demonstration. The microsimulation uses public survey data rather than claims-level data; claims data would enable individual-level comorbidity attribution and actual utilization rates by plan. The 5 external validation experiments tested care delivery reform mechanisms, not AI-specific clinical parameters, which remain agent-specified and empirically untested. AI instances specified the AI delivery parameters, a design in which the same systems could exhibit structural optimism about AI capabilities; the pessimistic scenario and discount-factor sensitivity analysis partially mitigate this design bias. The authors hold financial interests in organizations that benefit from adoption of AI-mediated Medicaid delivery, a potential source of motivated reasoning that independent prospective evaluation will need to address. The AI virtual encounter share (58%) was parameterized at the population level rather than conditioned on individual-level broadband access; individuals without broadband access plausibly have near-zero AI encounter probability, meaning clinical effects may be more concentrated among digitally connected beneficiaries than the model represents. Consumer surplus estimates used per-event QALY decrements (hospitalization: 0.05; ED: 0.01; primary care: 0.002) within published ranges but not derived from a single validated condition-specific source; QALY-based results should be interpreted alongside the cost and utilization findings. Language barriers, digital literacy, and institutional trust in AI-delivered care are not separately parameterized beyond broadband access. The welfare analysis does not capture transition costs for MCO administrative workers displaced by automation, which would reduce the estimated net social surplus. The model captures steady-state effects without modeling implementation or learning curves. AI clinical models may exhibit differential diagnostic accuracy by race [43]; the specified bias auditing has not been empirically tested as mitigation.

Conclusion

A microsimulation calibrated to national Medicaid benchmarks and validated against 5 independent natural experiments projects that an AI-first Medicaid ACO could reduce hospitalizations, ED visits, and PMPM costs under conditions in which administrative overhead falls below 5% to 6% of premium. Administrative overhead reduction—not AI clinical performance—is the dominant driver of projected benefit and is the structural feature of the design most likely to be either achieved or not achieved early in implementation. Projected savings remain positive even when AI clinical effects are discounted to 50% of human-trial benchmarks. Digital access inequity is the primary condition for worsened outcomes for the most underserved populations. The administrative cost rate offers a measurable, prospectively

testable threshold for Section 1115 demonstration waiver authorization and for ongoing program evaluation.

Abbreviations

ACO

accountable care organization

ACS

American Community Survey

AI

artificial intelligence

AIAN

American Indian/Alaska Native

CCO

coordinated care organization

CHEERS

Consolidated Health Economic Evaluation Reporting Standards

CHW

community health worker

CMS

Centers for Medicare & Medicaid Services

CPC+

Comprehensive Primary Care Plus

CV

coefficient of variation

ED

emergency department

FQHC

federally qualified health center

HEDIS

Healthcare Effectiveness Data and Information Set

MCO

managed care organization

MSSP

Medicare Shared Savings Program

PMPM

per member per month

PSA

probabilistic sensitivity analysis

PUMS

Public Use Microdata Sample
QALY
quality-adjusted life year
SDOH
social determinants of health
UI
uncertainty interval
WTP
willingness to pay.

Declarations

Ethics approval and consent to participate

Not applicable. This study used only de-identified, publicly available secondary data (the American Community Survey Public Use Microdata Sample) and published aggregate statistics. It did not involve human participants, identifiable individual-level human data, or human tissue, and therefore did not require institutional review board approval or informed consent.

Consent for publication

Not applicable.

Competing Interests

Dr Basu receives grants from the NIH and CDC, has patents from Collective Health and Waymark, receives salary from HealthRIGHT360 and Waymark, and serves on the board of Waymark, all outside the submitted work. Dr Batniji receives salary from Waymark, has patents from Collective Health, and sits on the board of Waymark and Collective Health, all outside the submitted work. The authors hold financial interests in organizations that could benefit from the adoption of AI-mediated Medicaid delivery; this potential source of motivated reasoning is discussed in the Limitations and is intended to be addressed through independent prospective evaluation. No other competing interests are declared.

Funding

None.

Author Contribution

SB and RB jointly conceived and designed the study and interpreted the results. SB developed the microsimulation model and multi-agent debate framework, conducted the analyses, and drafted the

manuscript. RB contributed to study design and critically revised the manuscript for important intellectual content. Both authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The analytic code generated and used for this study is publicly available at <https://github.com/sanjaybasu/ai-aco-microsim>. The study population was drawn from the American Community Survey Public Use Microdata Sample (2019–2023), publicly available at <https://www.census.gov/programs-surveys/acs/microdata.html>. All other data sources are publicly available and are cited in the manuscript and supplementary appendices.

References

1. Hinton E, Diep K, Saunders H, Rudowitz R. 10 things to know about Medicaid managed care. KFF; July 1, 2024. <https://www.kff.org/medicaid/issue-brief/10-things-to-know-about-medicaid-managed-care/>
2. Montoya DF, Chehal PK, Adams EK. Medicaid managed care's effects on costs, access, and quality: an update. *Annu Rev Public Health*. 2020;41:537–49.
3. Duggan M. Does contracting out increase the efficiency of government programs? Evidence from Medicaid HMOs. *J Public Econ*. 2004;88:2549–72.
4. Kuziemko I, Meckel K, Rossin-Slater M. Does managed care widen infant health disparities? Evidence from Texas Medicaid. *Am Econ J Econ Policy*. 2018;10:255–83.
5. Arrow KJ. Uncertainty and the welfare economics of medical care. *Am Econ Rev*. 1963;53:941–73.
6. Mello MM, Trotsyuk AA, Djiberou Mahamadou AJ, Char DM. The AI arms race in health insurance utilization review: promises of efficiency and risks of supercharged flaws. *Health Aff (Millwood)*. 2026;45(1):6–13. <https://doi.org/10.1377/hlthaff.2025.00897>.
7. Palmer J, Pettit C, McCulla I, Kinnick C. Medicaid managed care financial results for 2024. Milliman Res Rep June 30, 2025. <https://www.milliman.com/en/insight/medicaid-managed-care-financial-results-for-2024>
8. Centers for Medicare & Medicaid Services. Medicaid and CHIP Managed Care Final Rule (CMS-2439-F): Medical loss ratio standards. *Fed Regist*. 2024;89:41002–295.
9. Patel SY, Baum A, Basu S. Geographic variations and facility determinants of acute care utilization and spending for ambulatory care-sensitive conditions. *Am J Manag Care*. 2024;30:e329–36.

10. Hsiang WR, Lukasiewicz A, Gentry M, et al. Medicaid patients have greater difficulty scheduling health care appointments compared with private insurance patients: a meta-analysis. *Inquiry*. 2019;56:46958019838118.
11. Zhu JM, Charlesworth CJ, Polsky D, McConnell KJ. Phantom networks: discrepancies between reported and realized mental health care access in Oregon Medicaid. *Health Aff (Millwood)*. 2022;41(7):1013–22. <https://doi.org/10.1377/hlthaff.2022.00052>.
12. McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare spending after 3 years of the Medicare Shared Savings Program. *N Engl J Med*. 2018;379:1139–49.
13. Ryan AM, Markovitz AA. Estimated savings from the Medicare Shared Savings Program. *JAMA Health Forum*. 2023;4(12):e234449. <https://doi.org/10.1001/jamahealthforum.2023.4449>.
14. McWilliams JM, Chernew ME, Landon BE. Medicare ACO program savings not tied to preventable hospitalizations or concentrated among high-risk patients. *Health Aff (Millwood)*. 2017;36(12):2085–93. <https://doi.org/10.1377/hlthaff.2017.0814>.
15. Bond AM, Civelek Y, Schpero WL, et al. Long-term spending of accountable care organizations in the Medicare Shared Savings Program. *JAMA*. 2025;333:1897–905.
16. Hsu J, Price M, Vogeli C, et al. Bending the spending curve by altering care delivery patterns: the role of care management within a Pioneer ACO. *Health Aff (Millwood)*. 2017;36:876–84.
17. Song Z, Ji Y, Safran DG, Chernew ME. Health care spending, utilization, and quality 8 years into global payment. *N Engl J Med*. 2019;381:252–63.
18. McConnell KJ, Renfro S, Lindrooth RC, Cohen DJ, Wallace NT, Chernew ME. Oregon's Medicaid reform and transition to global budgets were associated with reductions in expenditures. *Health Aff (Millwood)*. 2017;36:451–9.
19. Holm J, Pagán JA, Silver D. The impact of Medicaid accountable care organizations on health care utilization, quality measures, health outcomes and costs from 2012 to 2023: a scoping review. *Med Care Res Rev*. 2024;81:355–69.
20. Vasan A, Morgan JW, Mitra N, et al. Effects of a standardized community health worker intervention on hospitalization among disadvantaged patients with multiple chronic conditions: a pooled analysis of three clinical trials. *Health Serv Res*. 2020;55(suppl 2):894–901.
21. Sen AP, Chen LM, Samson LW, Epstein AM, Joynt Maddox KE. Performance in the Medicare Shared Savings Program by ACOs disproportionately serving dual and disabled populations. *Med Care*. 2018;56:805–11.
22. Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Patient Protection and Affordable Care Act; Advancing Interoperability and Improving Prior Authorization Processes. Final Rule Fed Regist. 2024;89:8758–943. CMS-0057-F.
23. Medicaid CHIP, Payment, Access Commission. Automation in Medicaid prior authorization: recommendations. MACPAC; 2026. <https://www.macpac.gov/publication/automation-in-medicaid-prior-authorization-recommendations/>.

24. Pestaina K, Wallace R, Lo J, Long M. Regulation of AI in prior authorization and claims review: a look at federal and state consumer protections. KFF; May 6, 2026. <https://www.kff.org/patient-consumer-protections/regulation-of-ai-in-prior-authorization-and-claims-review-a-look-at-federal-and-state-consumer-protections/>
25. Stern A, Ramos C, Prinvil J, November. 2025. <https://www.urban.org/research/publication/exploring-artificial-intelligence-and-automation-medicaid>.
26. California Health Care Foundation. Medi-Cal bold idea: AI-first Medi-Cal health data utility. CHCF; 2025. <https://www.chcf.org/resource/medi-cal-bold-idea-ai-first-medi-cal-health-data-utility/>
27. Cho T, Miller BJ. Using artificial intelligence to improve administrative process in Medicaid. *Health Aff Sch*. 2024;2(2):qxae008. <https://doi.org/10.1093/haschl/qxae008>.
28. Song J, Wang X, Wang B, et al. Learning implementation of a guideline-based decision support system to improve hypertension treatment in primary care in China: pragmatic cluster randomised controlled trial. *BMJ*. 2024;386:e079143.
29. Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence. *Nature*. 2025;642:442–50.
30. Singhal K, Tu T, Gottweis J, et al. Toward expert-level medical question answering with large language models. *Nat Med*. 2025;31:943–50.
31. Ramaswamy A, Tyagi A, Hugo H, et al. ChatGPT health performance in a structured test of triage recommendations. *Nat Med*. 2026;32(5):1671–5. <https://doi.org/10.1038/s41591-026-04297-7>.
32. Basu S, Sheth P, Muralidharan B, et al. Comparative evaluation of AI architectures for medical triage safety: a real-world validation study. *JMIR Preprints*. 2026. <https://doi.org/10.2196/preprints.94081>.
33. Du Y, Li S, Torralba A, Tenenbaum JB, Mordatch I. Improving factuality and reasoning in language models through multiagent debate. In: *Proceedings of the 41st International Conference on Machine Learning (ICML)*. 2024:11733–63.
34. Yuksekgonul M, Koceja D, Li X et al. Learning to discover at test time. *arXiv preprint arXiv:2601.16175*. 2026.
35. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Value Health*. 2022;25:3–9.
36. Elvidge J, Hawksworth C, Avşar TS, et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). *Value Health*. 2024;27(9):1196–205.
37. Medicaid CHIP, Payment, and Access Commission. MACStats: MedicaidCHIP data book. 2024. <https://www.macpac.gov/macstats/>
38. Centers for Medicare & Medicaid Services. Medicaid and Children’s Health Insurance Program (CHIP) programs; Medicaid managed care, CHIP delivered in managed care, and revisions related to third party liability. Final rule. 42 CFR Parts 431, 433, 438, 440, 457, and 495. *Fed Regist*. 2016;81:27498–901.

39. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol*. 2014;67:401–9.
40. Pew Research Center. Internet/broadband fact sheet. 2024.
<https://www.pewresearch.org/internet/fact-sheet/internet-broadband/>
41. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. *NEJM Catalyst*. 2020;1(3).
42. Rodriguez JA, Khoong EC, Lipsitz SR, Lyles CR, Bates DW, Samal L. Telehealth experience among patients with limited English proficiency. *JAMA Netw Open*. 2024;7:e2410691.
43. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–53.
44. Patel SY, Barnett ML, Basu S. Predicting quality measure completion among 14 million low-income patients enrolled in Medicaid. *npj Digit Med*. 2025;8:393. <https://doi.org/10.1038/s41746-025-01797-7>.
45. Himmelstein DU, Campbell T, Woolhandler S. Health care administrative costs in the United States and Canada, 2017. *Ann Intern Med*. 2020;172:134–42.
46. Centers for Medicare & Medicaid Services. Medical loss ratio annual reporting. 2024.
<https://www.cms.gov/ccio/resources/data-resources/mlr>
47. McWilliams JM, Chernew ME, Landon BE, Schwartz AL. Performance differences in year 1 of Pioneer accountable care organizations. *N Engl J Med*. 2015;372:1927–36.
48. Singh P, Fu N, Dale S, et al. The Comprehensive Primary Care Plus model and health care spending, service use, and quality. *JAMA*. 2024;331:132–46.
49. Kaldor N. Welfare propositions of economics and interpersonal comparisons of utility. *Econ J*. 1939;49:549–52.
50. Atkinson AB. On the measurement of inequality. *J Econ Theory*. 1970;2:244–63.
51. Polsky D, Candon M, Saloner B, et al. Changes in primary care access between 2012 and 2016 for new patients with Medicaid and private coverage. *JAMA Intern Med*. 2017;177:588–90.
52. Buxbaum JD, Arnold DR, Fuse Brown EC, Whaley CM, Ryan AM. Substantial variation in administrative spending and profit across state insurance markets, 2023. *Health Aff (Millwood)*. 2026;45(3):331–40. <https://doi.org/10.1377/hlthaff.2025.00779>.
53. CAQH. 2024 CAQH Index: from transactions to trust—building better care through healthcare automation. Washington (DC): CAQH; 2024.
https://www.caqh.org/hubfs/Index/2024%20Index%20Report/CAQH_IndexReport_2024_FINAL.pdf

Figures

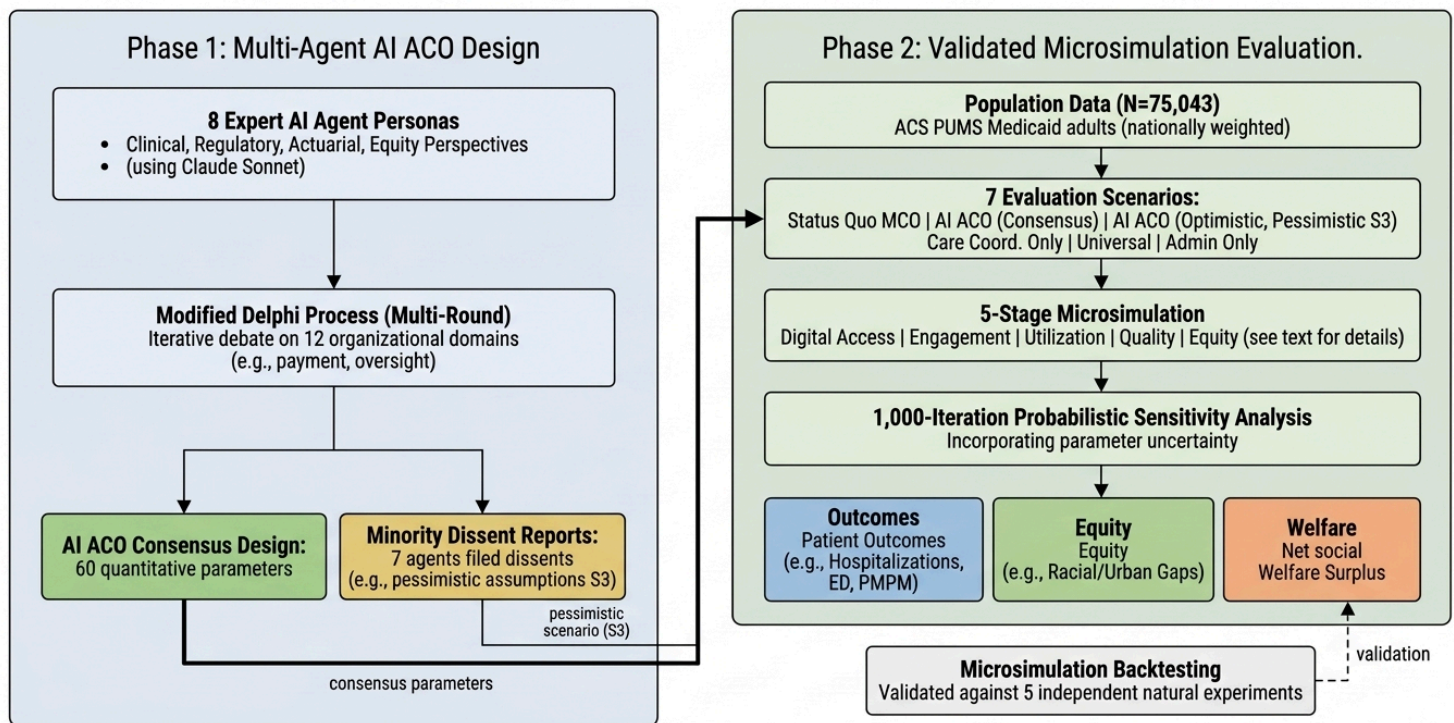


Figure 1. Study Design Overview: AI ACO Design and Evaluation.
Phase 1 (left) designs the AI ACO through multi-agent debate and a Modified Delphi process. Phase 2 (right) evaluates the designs using microsimulation across 7 scenarios, validated by backtesting.

Figure 1

Study Design. Two-phase design. Phase 1 (left panel): 8 AI agents with distinct expert personas (health economist, Medicaid regulator, patient advocate, physician, FQHC operator, health ethicist, actuary, health services researcher) independently proposed AI accountable care organization (ACO) designs across 12 parameter domains, then engaged in a Modified Delphi process (3 rounds) with iterative convergence assessment (coefficient of variation threshold <0.15 for quantitative parameters; $\geq 75\%$ agreement for categorical parameters). Consensus required ≥ 10 of 12 domains to meet criteria for 2 consecutive rounds. Agents outside 1 SD of consensus submitted structured minority reports. Phase 2 (right panel): the consensus design was evaluated via 5-stage Monte Carlo microsimulation—digital access, care engagement, clinical utilization, preventive care quality, and equity—across 7 scenarios with 1,000-iteration probabilistic sensitivity analysis. ACO indicates accountable care organization; AI, artificial intelligence; CV, coefficient of variation; FQHC, federally qualified health center; IQR, interquartile range; SD, standard deviation.

Figure 2. Calibration and External Validation

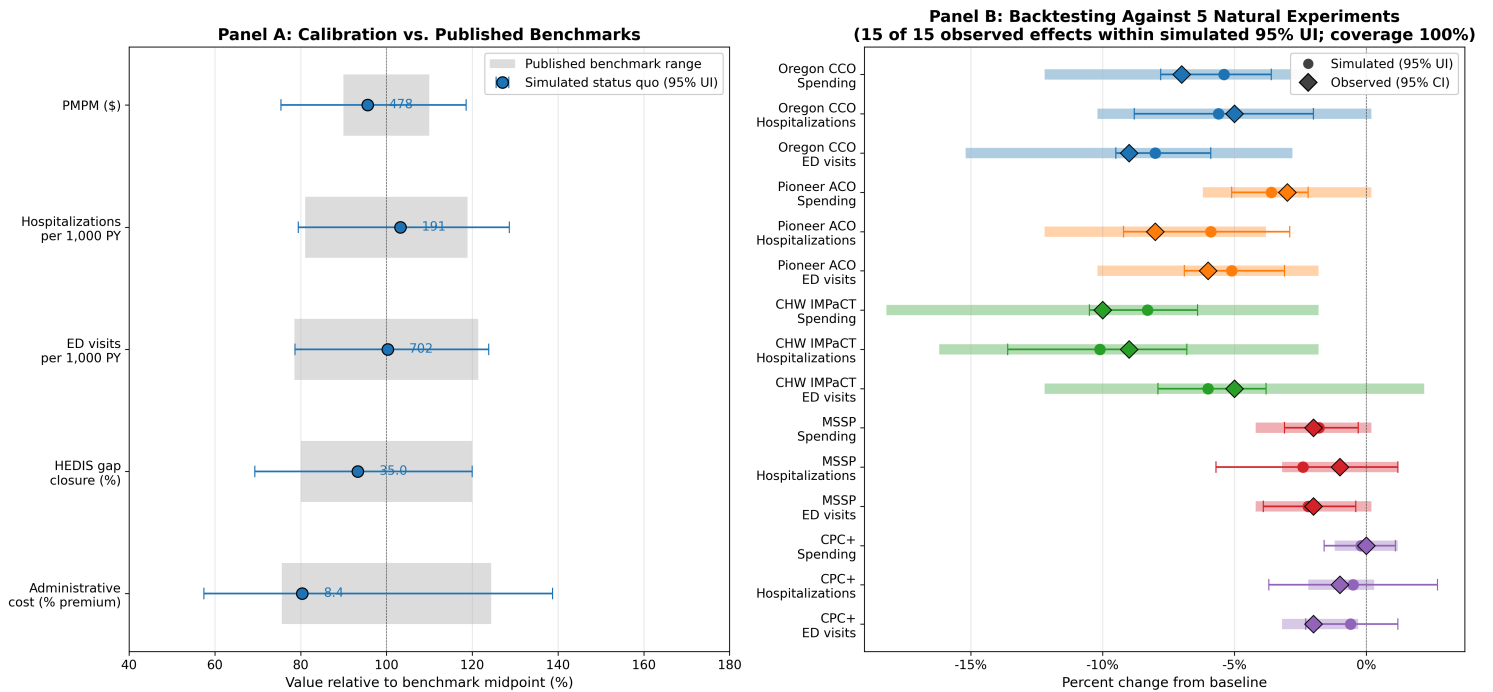


Figure 2

Calibration and External Validation. Left panel: simulated status quo values (points with 95% uncertainty interval [UI] whiskers) plotted against published Medicaid benchmark ranges (shaded bands) for per-member-per-month (PMPM) cost, hospitalization rate, emergency department (ED) visit rate, Healthcare Effectiveness Data and Information Set (HEDIS) gap closure, and administrative cost share. Right panel: backtesting against 5 independent delivery reform natural experiments (Oregon coordinated care organizations [CCOs], Pioneer accountable care organizations [ACOs], community health worker [CHW] randomized trials, Medicare Shared Savings Program [MSSP], and Comprehensive Primary Care Plus [CPC+]). The microsimulation was parameterized to match each reform's intervention characteristics with AI-specific features disabled; projected percent changes (points with 95% UI) are plotted against observed effect sizes from published evaluations (diamonds with 95% confidence intervals). The observed effect fell within the simulated 95% UI in 15 of 15 comparisons (100% coverage), with mean calibration ratio 0.99 and mean absolute error 1.1 percentage points. Figure regenerated at increased label size and axis font for revised submission. ACO indicates accountable care organization; CCO, coordinated care organization; CHW, community health worker; CPC+, Comprehensive Primary Care Plus; ED, emergency department; HEDIS, Healthcare Effectiveness Data and Information Set; MSSP, Medicare Shared Savings Program; PMPM, per member per month; UI, uncertainty interval.

Figure 4. Conditions for Improvement:
Probability of Net Savings by Administrative Cost Rate

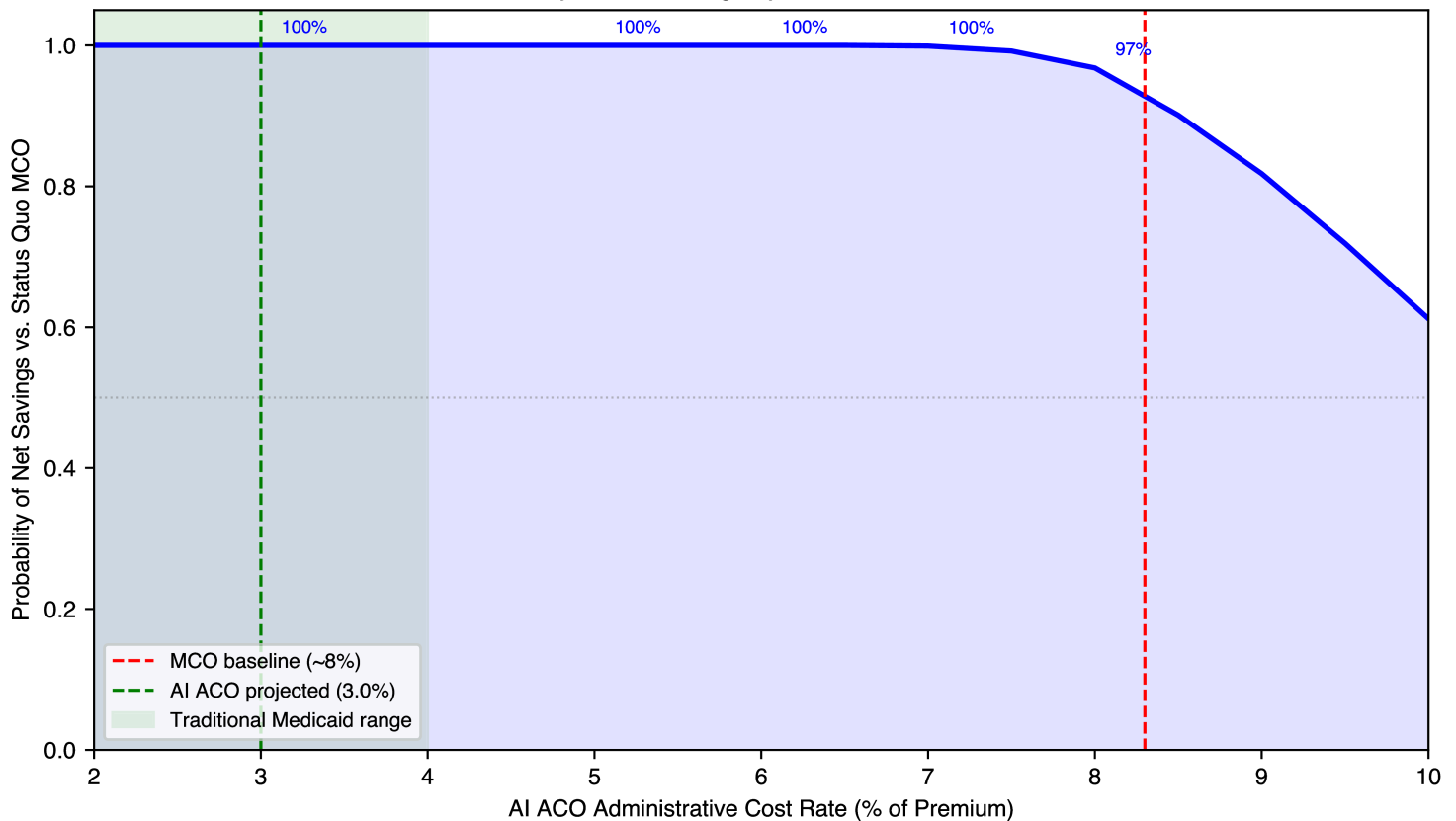


Figure 3

Probability of Net Improvement by Administrative Cost Rate. Probability that net social surplus is positive (y-axis) as a function of AI ACO administrative overhead as a percentage of premium (x-axis), based on 1,000 probabilistic sensitivity analysis iterations holding AI clinical efficacy at consensus levels. The dashed horizontal line indicates 50% probability. The shaded region indicates the range consistent with traditional (non-managed care organization) Medicaid administrative rates (2–4%). The current managed care organization (MCO) overhead reference (8.4%) is shown with a vertical reference line. At 3% overhead, probability of net savings was 100%; at 6%, 78%; at 7%, 61%; at 8%, 45%. ACO indicates accountable care organization; AI, artificial intelligence; MCO, managed care organization; PMPM, per member per month; PSA, probabilistic sensitivity analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [BasuAIACOMedicaidAdditionalFile1Appendix.docx](#)
- [BasuAIACOMedicaidCHEERSAIReportingChecklist.docx](#)