

Measuring and reducing intervention misalignment in next best action selection for care management: a within-patient natural experiment

Sanjay Basu

sanjay.basu@ucsf.edu

University of California San Francisco

Parth Sheth

University of Pennsylvania

Sadiq Patel

University of Pennsylvania

Article

Keywords:

Posted Date: July 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9545369/v1>

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

[Read Full License](#)

Additional Declarations: Competing interest reported. S.B., P.S., and S.P. are employees of Waymark, a public benefit organization providing free social and healthcare services to patients receiving Medicaid.

Measuring and reducing intervention misalignment in next best action selection for care management: a within-patient natural experiment

Authors: Sanjay Basu, MD PhD^{a,b}; Parth Sheth, MS^{b,c}; Sadiq Patel, PhD^{b,c}

^a Department of Medicine, University of California San Francisco, San Francisco, CA ^b Waymark, San Francisco, CA ^c School of Social Policy and Practice, University of Pennsylvania, Philadelphia, PA

Corresponding author: Sanjay Basu, MD PhD | sanjay.basu@ucsf.edu

Article type: Article (Scientific Reports) Reporting guideline: TRIPOD-AI with the RECORD extension for routinely collected data.

Abstract

Patients in care management typically have multiple chronic conditions and social needs, and care managers must choose a next best action — chronic disease management, behavioral health, medication adherence support, or social needs navigation. We define intervention misalignment as the fraction of patients whose chosen next best action is not the action with the largest expected health gain, and measure it in 34,971 rising-risk Medicaid patients across three states (2023–2025) using within-patient natural experiments from staggered enrollment timing. Intervention misalignment under current behavioral routing was 27.0% (95% CI, 25.9–28.1%) and concentrated in social needs: transport, food security, housing, and financial benefits accounted for 20.9% of within-patient-preferred next best actions but only 1.9% of behavioral routing decisions (12.5- to 25.5-fold gap). Across thirteen pre-registered policies, PEARL (Policy Evolution through Aligned Retrospective Learning) — combining the within-patient signal, group-stratified fairness, and a fourteen-expert router — reduced intervention misalignment to 2.0% ($\Delta = 25.0$ percentage points; $p < 0.0001$), 92.6% of the maximum reduction, and was the best deployable policy on off-policy evaluation. Behavioral cloning on within-patient preferences reduced intervention misalignment to 4.4%, so the within-patient signal alone produced most of the reduction; PEARL variants without group-stratified fairness increased intervention misalignment to 38.7–42.4%, worse than current routing. Intervention misalignment is a measurable, modifiable property of next best action selection in care management.

Introduction

Care management programs serve several million people in the United States — most Medicaid beneficiaries and a growing share of Medicare beneficiaries with multiple chronic conditions and social needs — and care managers must choose, in each patient encounter, which one intervention type to act on next¹. Patients in care management typically have several concurrent needs, so the question is rarely “what does this patient need?” but rather “which need should the care manager address first?” — a *next best action* problem rather than a matching problem. The next best action framing is forced on care managers by three operational realities: care managers are in chronically short supply relative to the population they serve (community health worker and case-manager staffing ratios in many Medicaid programs exceed 1:200, so each patient receives only a few brief encounters per quarter); each encounter is typically fifteen to thirty minutes, and a single new intervention — a medication change, a referral, a new monitoring schedule — requires several follow-up touchpoints, so multiple new actions cannot be initiated simultaneously; and evidence-based behavior-change practice prescribes a single new goal at a time, because patients with limited health literacy or numeracy (common in the rising-risk Medicaid population) show worse adherence and worse outcomes when asked to absorb multiple new changes at once. Care managers answer the next best action question largely as art rather than science: clinical guidelines tell a care manager how to treat one disease at a time, but no guideline tells a care manager whether to address hypertension, diabetes, behavioral health, or a social need first when the patient has all four. Consider a care manager with fifteen minutes on a Friday evening, looking at a patient who needs an antipsychotic prior authorization and who also has uncontrolled stage II hypertension without acute symptoms — both interventions are clinically appropriate, neither guideline indicates which one will produce the larger health gain in this patient, and care managers make this kind of next best action choice millions of times per year, with the quality of those choices unmeasured.

The next best action chosen by a care manager falls into one of several intervention types: chronic disease management (hypertension, diabetes, heart failure, asthma/COPD), behavioral health, medication adherence support, or social needs navigation (food security, housing, transportation, or financial benefits). Care management programs whose next best action choices match each patient’s largest expected health gain should prevent acute care use; programs whose next best action choices do not should not. The opportunity is large: potentially preventable inpatient stays total approximately 3.5 million per year in the United States and cost \$33.7 billion². Yet most programs guide next best action selection primarily through risk scores that flag who needs care without specifying which action to take^{3,4}, and randomized trial results have been inconsistent — some trials show

readmission reduction⁵, others show no reduction⁶. The most cited null result is the Camden Coalition trial, in which all enrolled patients received the same intensive protocol regardless of individual need: 62.3% were readmitted under the protocol vs. 61.7% under usual care ($P = 0.81$)⁶.

We hypothesize that uniform next best action assignment dilutes the average treatment effect because a uniform action helps patients for whom that action is the largest expected health gain, does not help patients for whom a different action would be larger, and the two effects average to near zero — and we therefore define *intervention misalignment* as a measurable, modifiable property of next best action selection policies: the fraction of patients whose chosen next best action is not the action expected to produce the largest health gain. Measuring intervention misalignment from observational data is hard because care managers select next best actions based on unmeasured clinical judgment that is correlated with prognosis^{7,8}. We address this problem with a natural experiment in care management logistics: when a multistate care management program begins serving a new accountable care organization (ACO), patients enroll on a staggered schedule that reflects program logistics — data-feed connectivity, care-team training — rather than patient health, so each patient has a pre-enrollment period (without care management) and a post-enrollment period (with care management) that we compare directly. We call this comparison a *within-patient natural experiment*: each patient is their own control, and confounders that do not change over time cancel out by design⁹.

This study makes three contributions: we define intervention misalignment in next best action selection as a causal estimand and measure it in 34,971 rising-risk Medicaid patients; we construct within-patient preference pairs from staggered enrollment timing and use them to compare thirteen pre-registered next best action selection policies on a held-out test set; and we identify which components of a selection policy reduce intervention misalignment and which do not.

Results

The patient population

We studied 34,971 rising-risk Medicaid patients in a multistate care management program (2023–2025; Table 1), having excluded 4,346 of 39,317 assessed patients (11.0%) for incomplete administrative linkage; excluded patients did not differ from included patients on clinical characteristics. Mean age was 36.1 years (SD, 12.3), 60.9% were female, and the mean Charlson comorbidity index was 1.00 (SD, 1.47); the 90-day acute care event rate was 4.1% (1,438 events). Behavioral routing concentrated assignments in two categories — care access coordination (52.5%) and behavioral health referral (27.8%), 80.3%

combined — and assigned social needs intervention types (food security, transport/utilities, housing, financial benefits) to only 1.9% of patients combined.

Within-patient preference pairs

Staggered enrollment yielded 1,707 eligible patient-window pairs in the training set, of which 622 met the primary criterion (favorable post-enrollment outcome, unfavorable pre-enrollment outcome) and 1,085 met a weak-positive criterion (favorable outcome in both windows; Table 2). We constructed two secondary pair types: CHW waitlist windows duplicated the primary pairs and were dropped, and Medicaid eligibility churn windows yielded 221 intent-to-treat pairs. Within-patient pair patients did not differ from the full cohort on Charlson index (1.03 vs. 1.00; $P = 0.42$) or prior-year event rate (4.3% vs. 4.1%; $P = 0.68$).

Six pre-specified falsification tests assessed the within-patient natural experiment (T1–T6). Four tests passed — T1 (covariate balance, maximum standardized mean difference = 0.143), T2 (placebo outcomes, AUC = 0.50), T3 (administrative-predictor F-test, $p = 0.94$), and T4 (density continuity at the enrollment boundary) — while T5 detected local average treatment effect heterogeneity (58.5%) between primary and secondary pair types, triggering a pre-specified contingency that restricted the primary analysis to staggered-enrollment pairs. T6 was inconclusive by design (staggered-enrollment pairs all share the pre-enrollment-then-post-enrollment direction, so bidirectional comparison is infeasible); we used the trajectory-slope covariate from T1 as the stationarity check instead.

Within-patient-preferred next best actions differed substantially from behavioral routing choices: social needs actions accounted for 20.9% of within-patient-preferred pairs (transport/utilities 5.1%, food security 5.2%, housing 2.5%, financial benefits 8.1%) but only 1.9% of behavioral routing next best action choices, with within-patient-preferred pairs 25.5-fold richer in transport/utilities, 17.3-fold richer in food security, and 12.5-fold richer in housing. Conversely, behavioral routing was 1.9-fold richer in care access coordination (52.5% vs. 27.4%) and 2.3-fold richer in behavioral health referral (27.8% vs. 12.1%).

Intervention misalignment under behavioral routing

Intervention misalignment under behavioral routing was 27.0% (95% CI, 25.9–28.1%; Manski partial identification bounds, 26.0–28.4%; effective sample size, 2,670), meaning more than one in four rising-risk patients received a next best action for which a different action was expected to produce a larger health gain — a reduction of at least 2 percentage points in predicted 90-day acute care event probability. The estimate was robust to unmeasured confounding: the E-value¹⁰ was 11.92, so an unmeasured next-best-action-selection confounder would need to predict both action selection and outcome by more

than 11.92-fold to fully explain away the finding — substantially above measured confounder strengths (Charlson hazard ratio ≈ 2.0). We estimated intervention misalignment using doubly robust counterfactual modeling across the full 14 candidate next best actions; positivity coverage was 97.6%, and Manski partial identification bounds accommodated the limited positivity support in transport/utilities and food security (each chosen as the next best action in only 0.2–0.3% of patients under behavioral routing).

Intervention misalignment was pervasive across socioeconomic strata: misalignment by area deprivation index (ADI) quintile was 30.5% (Q1, least deprived), 30.2% (Q2), 30.5% (Q3), 17.7% (Q4), and 25.1% (Q5, most deprived; $N = 4,087$ with ADI linkage; Kruskal-Wallis $q < 0.05$); misalignment by race/ethnicity was 30.8% for patients with unknown race/ethnicity, 20.3% for Black non-Hispanic patients, and 16.5% for Hispanic patients. The non-monotonic pattern indicates that intervention misalignment is not a simple socioeconomic gradient but reflects systematic under-prioritization of social needs as the next best action across multiple strata.

Off-policy evaluation: thirteen pre-registered routing policies

We compared thirteen pre-registered next best action selection policies on the held-out test set ($N = 6,995$; Table 3, Figure 2) using doubly robust off-policy evaluation (DR-OPE), where lower DR-OPE values mean fewer predicted 90-day acute care events. We benchmarked all twelve deployable policies against a theoretical oracle policy that chooses, for each patient, the next best action with the lowest predicted event probability — the oracle is the best possible policy by construction but is not deployable in practice because it uses the same outcome model that defines intervention misalignment (DR-OPE = 0.026; intervention misalignment = 0.0%). Among the twelve deployable policies, PEARL ranked first, achieving DR-OPE of 0.027 (95% CI, 0.026–0.029) and reducing intervention misalignment to 2.0% ($\Delta = 25.0$ percentage points vs. behavioral routing; one-sided bootstrap $p < 0.0001$) — 92.6% of the maximum possible reduction toward the oracle — with coverage of 1.00 relative to the behavioral support set, meaning every action PEARL recommended lay within the historical action distribution.

PEARL combines three components — the within-patient signal (preference pairs from staggered enrollment), group-stratified fairness (a per-group preference-optimization loss¹¹), and a fourteen-expert router (one expert per intervention type) — and we isolated each using a pre-registered ablation hierarchy. The within-patient signal is necessary: observational DPO without the within-patient signal (C5; rank 12) reduced intervention misalignment only modestly to 20.4% (DR-OPE = 0.044). The within-patient signal is also most of what is needed: behavioral cloning on within-patient preferences (C4; rank 3) reduced intervention misalignment to 4.4% (DR-OPE = 0.030; 95% CI, 0.028–0.031) without any preference-optimization loss, fairness constraint, or expert router. Group-

stratified fairness, however, is necessary on top of preference optimization, because PEARL variants with the within-patient signal but without group-stratified fairness *increased* intervention misalignment above behavioral routing (PEARL base 42.4%, rank 9; PEARL MoE Full 38.7%, rank 5). All three components together reduced intervention misalignment to 2.0%.

Comparator policies confirmed the same pattern: conservative Q-learning (CQL; rank 4) reduced DR-OPE to 0.034 but did not reduce intervention misalignment (32.0%), showing that offline reinforcement learning improves the aggregate event rate without improving next best action selection; risk-score policies (LACE Index, HOSPITAL Score, XGBoost) and Decision Transformer (ranks 6–8 and 11) performed near behavioral routing on DR-OPE and did not reduce intervention misalignment (41–45%); and causal forest (rank 13; DR-OPE = 0.144) failed because its underlying treatment effect estimates extrapolate to unobserved actions. All thirteen policies had effective sample sizes above the pre-specified threshold of 500 (range 2,670–6,995), and conformal prediction empirical coverage was 89.9% (target 90%; mean prediction interval width 0.159; abstention threshold $\tau = 0.082$).

Camden Coalition analog reanalysis

We measured intervention misalignment in 242 rising-risk patients matching the Camden Coalition trial enrollment criteria (Charlson ≥ 4 , prior hospitalization ≥ 2 in 6 months). The Camden Coalition uniform intensive protocol produced intervention misalignment of 90.9% in this matched subgroup because every patient received the same next best action regardless of need, whereas PEARL produced intervention misalignment of 0.0% by construction (PEARL chooses the minimum-predicted-outcome next best action for each patient using the same outcome model that defines intervention misalignment, a circularity we disclose as a limitation). PEARL distributed next best action choices across multiple types — behavioral health referral 43.4%, hypertension management 20.2%, heart failure management 8.3%, housing navigation 6.6%, care access coordination 5.8%, pulmonary management 5.0%, and other types — and the outcome model simulated a 30-day readmission rate 15.1 percentage points lower under PEARL than under the uniform protocol (90% prediction interval, 13.5–16.8%; 1,000 bootstrap iterations; Appendix Figure 3). This simulation is internally consistent with the relationship between intervention misalignment and outcomes in the matched subgroup, but is not a counterfactual claim about the Camden Coalition trial itself.

Sensitivity analysis

We pre-specified twenty sensitivity analyses across seven parameters (Appendix Table 3, Appendix Figure 4). PEARL intervention misalignment remained at 2.0% across all eighteen training-dependent sensitivity analyses (window length, weight clipping, regularization, outcome window, trajectory adjustment, pair-type restriction, and seven

pairwise combinations). The two evaluation-only sensitivities varied the intervention misalignment threshold ϵ : $\epsilon = 0.01$ (a more lenient threshold that counts even tiny modeled differences as misalignment) raised PEARL intervention misalignment to 81.3% — a pre-specified major sensitivity finding, because at this lenient threshold most routing policies appear misaligned regardless of true matching quality — and $\epsilon = 0.05$ (a stricter threshold) yielded 2.9%. PEARL was better than behavioral routing on DR-OPE in all twenty sensitivity analyses.

Discussion

Three findings emerge from the analysis.

First, intervention misalignment reduction requires three components — the within-patient signal, group-stratified fairness, and the expert router — none of which is sufficient on its own. The within-patient signal is necessary, because observational DPO without it reduces intervention misalignment only modestly (20.4% vs. 27.0% under behavioral routing); but the within-patient signal is also most of what is needed, because behavioral cloning on within-patient preferences (the simplest ablation) reduces intervention misalignment to 4.4% without any preference-optimization loss, fairness constraint, or expert router. Group-stratified fairness is then necessary on top of preference optimization, because PEARL variants with the within-patient signal but without group-stratified fairness *increase* intervention misalignment above behavioral routing (PEARL base 42.4%, MoE Full 38.7%) — the contrastive preference-optimization loss amplifies imbalances in the training distribution unless group-stratified fairness counterweights them¹¹. All three components together reduce intervention misalignment to 2.0%, achieve 92.6% of the maximum possible reduction, and produce the best off-policy evaluation rank among deployable policies.

Second, intervention misalignment under behavioral routing concentrates in social needs as the next best action and is not a simple socioeconomic gradient. Behavioral routing under-prioritizes transport/utilities by 25.5-fold, food security by 17.3-fold, and housing by 12.5-fold as the next best action, relative to within-patient-preferred actions, and intervention misalignment by ADI quintile is 30.5% (Q1, least deprived), 17.7% (Q4), and 25.1% (Q5, most deprived) — a non-monotonic pattern that recurs by race/ethnicity (30.8% unknown, 20.3% Black non-Hispanic, 16.5% Hispanic). Intervention misalignment therefore reflects systematic under-prioritization of social needs as the next best action, paralleling earlier findings of unintended demographic gaps in algorithmic risk scoring¹². The T2 falsification test (AUC = 0.50) argues against selection bias and the E-value of 11.92 argues against unmeasured confounding as primary explanations.

Third, the Camden Coalition trial is consistent with high intervention misalignment as a mechanism: the Camden Coalition uniform intensive protocol produces intervention misalignment of 90.9% in the matched subgroup of 242 rising-risk patients because every patient received the same next best action regardless of which action would have produced the largest individual health gain, and the outcome model simulates a 30-day readmission rate 15.1 percentage points lower under PEARL personalized next best action selection than under the uniform protocol. Uniform protocols applied to heterogeneous populations produce high intervention misalignment by construction, and high intervention misalignment dilutes average treatment effects toward zero regardless of program intensity. This finding is internally consistent with the relationship between intervention misalignment and outcomes, but is not a counterfactual claim about the Camden Coalition trial itself.

Five limitations apply. First, the within-patient natural experiment identifies the effect of care management receipt versus non-receipt rather than the head-to-head effect of one next best action versus another, so intervention misalignment is a model-based measure of next best action selection quality with partial within-patient causal grounding rather than a fully identified action-level causal estimand; we mitigate this with the T2–T3 falsification tests, a within-pair propensity secondary analysis (Appendix Table 8; intervention misalignment within 5% of primary), and the E-value of 11.92. Second, the T5 falsification test detected local average treatment effect heterogeneity of 58.5% between primary and secondary preference pair types, so we restricted the primary analysis to staggered-enrollment preference pairs (a pre-specified contingency), and T6 was inconclusive by design (we used the trajectory-slope covariate from T1 as the stationarity check). Third, the Camden Coalition analog uses a Charlson criterion absent from the original trial enrollment criteria, and the outcome model that defines intervention misalignment is the same outcome model that PEARL uses for training — a circularity we disclose and address by reporting non-PEARL routing policies on the same outcome model. Fourth, external validation in Medicare Advantage and commercial ACO populations has not been conducted; the within-patient natural experiment framework is payer-agnostic by construction, but population-specific validation is required before generalizability claims. Fifth, ADI linkage was available for 58.4% of held-out test patients, so the intervention misalignment gradient by ADI quintile may be attenuated for unlinked patients.

Taken together, these findings establish intervention misalignment as a measurable property of next best action selection in care management, identify the within-patient signal and group-stratified fairness as necessary components for reducing it, and identify social needs under-prioritization as the dominant correctable gap. We are preparing a stepped-wedge cluster-randomized trial across 24 care teams (Hussey-Hughes design¹³; 80% power to detect a 15% relative reduction in 90-day acute care events) with

prospective intervention misalignment reduction as a co-primary mechanistic endpoint alongside the clinical outcome.

Methods

Study design and reporting

We conducted a retrospective cohort analysis of routinely collected ACO care management data. We reported the analysis per TRIPOD-AI (Appendix Table 9) with the RECORD extension. The study was reviewed and approved by the WCG Institutional Review Board (WCG IRB; protocol 20253751). The WCG IRB granted a waiver of the requirement to obtain informed consent under 45 CFR 46.116(f) because the research involved only secondary analysis of routinely collected, de-identified administrative data, presented no more than minimal risk to participants, could not practicably be carried out without the waiver, and did not adversely affect the rights and welfare of participants. All methods were performed in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki and the U.S. Common Rule (45 CFR 46). We pre-registered the analysis plan on OSF (DOI: 10.17605/OSF.IO/D69M2) before any holdout evaluation.

Setting and participants

We used data from a multistate value-based care management program serving Medicaid beneficiaries in Washington, Virginia, and Ohio (January 2023 – December 2025). The program deploys multidisciplinary teams providing comprehensive care management (CPT G0511), complex care management (G0512), and transitional care management, and identifies rising-risk patients using a validated gradient-boosted risk model¹⁴. Data sources included Medicaid claims, ACO care management records, risk scores, and area-level socioeconomic context (Area Deprivation Index; American Community Survey 5-year estimates).

We studied the rising-risk patient stratum (70th–90th percentile of predicted 12-month acute care utilization trajectory) and included 34,971 of 39,317 assessed patients (89.0%) with complete administrative linkage. The primary outcome was the 90-day acute care event — any unplanned hospitalization or avoidable emergency department visit within 90 days of the index date (Fine-Gray competing risks regression with death as the competing event) — with 30-day readmission as a pre-specified secondary outcome. We split patients 80/20 into training (N = 27,976) and held-out test (N = 6,995); we fit all models on the training set and evaluated all models on the held-out test set.

Within-patient natural experiments

We constructed within-patient preference pairs from three administrative discontinuities: staggered enrollment (the primary discontinuity, in which enrollment timing reflects program logistics — data-feed connectivity, care-team training — rather than patient health), CHW waitlist windows (queue position by referral date), and Medicaid eligibility churn (≥ 60 -day coverage gap; intent-to-treat secondary analysis). We constructed a primary within-patient preference pair when a patient had a favorable 90-day outcome in the post-enrollment window and an unfavorable outcome in the pre-enrollment window, a weak-positive pair when a patient had a favorable outcome in both windows (weight 0.5), and discarded pairs in which the post-enrollment window had an unfavorable outcome (no causal learning signal).

We pre-specified six falsification tests (T1–T6) for the within-patient natural experiment, including a negative-control-outcome test¹⁵; T1 mandates inclusion of a within-patient trajectory-slope covariate, and T5 heterogeneity above 20% triggers a pre-specified restriction to staggered-enrollment pairs only. We disclose a scope qualification: the within-patient natural experiment identifies the effect of care management receipt versus non-receipt rather than the effect of one specific intervention type versus another, which we address with a within-pair propensity secondary analysis (proxy-variable identification approach¹⁶) and an E-value sensitivity analysis (Appendix Tables 6–8).

Intervention misalignment

We define intervention misalignment as the fraction of patients whose chosen next best action — the intervention type a care manager prioritized in a given encounter — is not the action expected to produce the largest health gain, with the gain measured as a reduction of at least $\varepsilon = 0.02$ in predicted 90-day acute care event probability:

$$\text{IMI}(\pi_b) = \mathbb{E}_i[\mathbf{1}(\exists a \neq A_i: \hat{\mu}_{DR}(X_i, a) < \hat{\mu}_{DR}(X_i, A_i) - \varepsilon)]$$

where A_i is the next best action that behavioral routing assigned to patient i (the dominant intervention type during their post-enrollment window), and $\hat{\mu}_{DR}(X_i, a)$ is the augmented inverse probability weighting (AIPW) estimate of the 90-day acute care event probability if patient i were instead assigned action a ¹⁷. We fit the AIPW nuisance models on the training set only and applied them to the held-out test set, decomposed intervention misalignment into demographic and clinical components (Appendix Theorem S1), applied E-value sensitivity analysis¹⁰ to the within-pair confounding assumption, and computed Manski partial identification bounds¹⁸ for positivity-violating profiles. The 14 candidate next best actions are: care access coordination, clinical other, diabetes management, financial benefits navigation, food security referral, heart failure management, housing navigation, hypertension management, maternal/prenatal support, medication adherence support,

mental health/behavioral health referral, pulmonary management, substance use support, and transport/utilities navigation.

PEARL routing policy and pre-registered comparators

PEARL is a tabular routing policy trained using IPTW-weighted Direct Preference Optimization¹⁹ on within-patient preference pairs, theoretically equivalent to doubly robust off-policy optimization at the behavioral policy initialization²⁰. PEARL training applies demographic-stratified IPTW upsampling and a group-stratified preference-optimization loss (equal group weights, $\beta = 0.10$) to reduce intervention misalignment uniformly across demographic groups¹¹. We pre-specified three PEARL ablations: PEARL (base) trains on within-patient preference pairs without group-stratified fairness or the expert router; PEARL (MoE Full) adds the fourteen-expert router²¹ but not group-stratified fairness; PEARL (MoE Router) adds both (the primary specification). We pre-specified eight comparator routing policies (C1–C8) before any holdout evaluation (Appendix Table 5): LACE Index (C1)³; HOSPITAL Score (C2)⁴; XGBoost on claims features (C3); behavioral cloning on within-patient preferences without preference optimization (C4); observational DPO trained on non-within-patient preference pairs (C5); causal forest heterogeneous treatment effect estimation (C6)⁷; Decision Transformer (C7)²²; Conservative Q-learning (C8)²³.

Statistical analysis

The primary endpoint was intervention misalignment with a pre-specified one-sided bootstrap test (H_0 : PEARL intervention misalignment $\geq$ behavioral routing intervention misalignment; $\alpha = 0.05$; 2,000 resamples), and the secondary endpoint was DR-OPE policy value (Marginalized DR estimator)²⁰, where lower DR-OPE values indicate fewer predicted 90-day acute care events. We pre-specified an effective-sample-size threshold of 500 for DR-OPE reliability, conformal prediction intervals (90% prediction interval; empirical coverage target 88–92%)²⁴, and twenty sensitivity analyses (seven parameters, three values each) before unblinding. All analyses used Python 3.12 (scikit-learn 1.3, NumPy 1.24, pandas 2.0); the complete pipeline code, the synthetic-data demonstration, and the within-patient pair construction code are at <https://github.com/sanjaybasu/pearl>.

Data availability

Patient-level data cannot be redistributed due to Medicaid privacy restrictions. The within-patient preference pair construction pipeline, PEARL model code, aggregate summary statistics (Tables 1–3 and Appendix tables), and a synthetic-data demonstration that requires no data use agreement are publicly available at

<https://github.com/sanjaybasu/pearl>. A data use agreement template for ACO replication is available from the corresponding author. The pre-registered analysis plan is at OSF (DOI: 10.17605/OSF.IO/D69M2).

Code availability

Source code, training scripts, and analysis pipelines are released at <https://github.com/sanjaybasu/pearl> under a permissive open-source license.

Author contributions

S.B.: conceptualization, study design, causal identification framework, statistical analysis, writing — original draft, writing — review and editing, supervision. P.S.: data curation, validation, writing — review and editing. S.P.: data curation, software (risk model integration, outcome ascertainment), validation, writing — review and editing.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Author time was supported by the authors' employers (the University of California San Francisco; Waymark; the University of Pennsylvania) as part of routine academic and operational work.

Competing interests

S.B., P.S., and S.P. are employees of Waymark, a public benefit organization providing free social and healthcare services to patients receiving Medicaid.

Acknowledgements

None.

References

1. Blumenthal D, Anderson G, Burke S, Fulmer T, Jha AK, Long P. Tailoring complex-care management, coordination, and integration for high-need, high-cost patients: a vital direction for health and health care. *NAM Perspect*. 2016. <https://nam.edu/perspectives/tailoring-complex-care-management-coordination-and-integration-for-high-need-high-cost-patients-a-vital-direction-for-health-and-health-care/>
2. McDermott KW, Jiang HJ. Characteristics and costs of potentially preventable inpatient stays, 2017. *HCUP Stat Brief No. 259*. June 2020. Agency for Healthcare

Research and Quality. <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb259-Potentially-Preventable-Hospitalizations-2017.jsp>

3. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ*. 2010;182(6):551–557. <https://doi.org/10.1503/cmaj.091117>
4. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med*. 2013;173(8):632–638. <https://doi.org/10.1001/jamainternmed.2013.3023>
5. Verhaegh KJ, MacNeil-Vroomen JL, Eslami S, Geerlings SE, de Rooij SE, Buurman BM. Transitional care interventions prevent hospital readmissions for adults with chronic illnesses. *Health Aff (Millwood)*. 2014;33(9):1531–1539. <https://doi.org/10.1377/hlthaff.2014.0160>
6. Finkelstein A, Zhou A, Taubman S, Doyle J. Health care hotspotting — a randomized, controlled trial. *N Engl J Med*. 2020;382(2):152–162. <https://doi.org/10.1056/NEJMsa1906848>
7. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113(523):1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>
8. Heckman JJ, Vytlacil E. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*. 2005;73(3):669–738. <https://doi.org/10.1111/j.1468-0262.2005.00594.x>
9. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*. 1994;62(2):467–475. <https://doi.org/10.2307/2951620>
10. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268–274. <https://doi.org/10.7326/M16-2607>
11. Sagawa S, Koh PW, Hashimoto TB, Liang P. Distributionally robust neural networks for group shifts: on the importance of regularization for worst-case generalization. In: *Proc 8th Int Conf Learn Represent*. 2020. <https://arxiv.org/abs/1911.08731>
12. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. <https://doi.org/10.1126/science.aax2342>
13. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182–191. <https://doi.org/10.1016/j.cct.2006.05.007>

14. Patel SY, Baum A, Basu S. Prediction of non-emergent acute care utilization and cost among patients receiving Medicaid. *Sci Rep.* 2024;14:824.
<https://doi.org/10.1038/s41598-023-51114-z>
 15. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology.* 2010;21(3):383–388.
<https://doi.org/10.1097/EDE.0b013e3181d61eeb>
 16. Miao W, Geng Z, Tchetgen Tchetgen EJ. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika.* 2018;105(4):987–993.
<https://doi.org/10.1093/biomet/asy038>
 17. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc.* 1994;89(427):846–866.
<https://doi.org/10.2307/2290910>
 18. Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev.* 1990;80(2):319–323.
 19. Rafailov R, Sharma A, Mitchell E, Manning CD, Ermon S, Finn C. Direct preference optimization: your language model is secretly a reward model. *Adv Neural Inf Process Syst.* 2023;36.
 20. Kallus N, Uehara M. Doubly robust off-policy value and gradient estimation for deterministic policies. *Adv Neural Inf Process Syst.* 2020;33.
 21. Shazeer N, Mirhoseini A, Maziarz K, et al. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. In: *Proc 5th Int Conf Learn Represent.* 2017. <https://arxiv.org/abs/1701.06538>
 22. Chen L, Lu K, Rajeswaran A, et al. Decision transformer: reinforcement learning via sequence modeling. *Adv Neural Inf Process Syst.* 2021;34:15084–15097.
 23. Kumar A, Zhou A, Tucker G, Levine S. Conservative Q-learning for offline reinforcement learning. *Adv Neural Inf Process Syst.* 2020;33:1179–1191.
 24. Angelopoulos AN, Bates S. Conformal prediction: a gentle introduction. *Found Trends Mach Learn.* 2023;16(4):494–591. <https://doi.org/10.1561/22000000101>
 25. Dudík M, Langford J, Li L. Doubly robust policy evaluation and learning. In: *Proc 28th Int Conf Mach Learn.* 2011:1097–1104. <https://arxiv.org/abs/1103.4601>
 26. Farajtabar M, Chow YL, Ghavamzadeh M. More robust doubly robust off-policy evaluation. In: *Proc 35th Int Conf Mach Learn.* 2018:1447–1456.
<https://arxiv.org/abs/1802.03493>
-

Tables

Table 1. Demographic and clinical characteristics of the study population (N = 34,971 rising-risk Medicaid patients).

Characteristic	Overall cohort (N = 34,971)
Demographics	
Age, mean (SD), years	36.1 (12.3)
Female sex, n (%)	21,317 (60.9)
Residential ADI linkage available (held-out test set), n (%)	4,087 of 6,995 (58.4)
Clinical characteristics	
Charlson comorbidity index, mean (SD)	1.00 (1.47)
Mental health diagnosis, n (%)	11,086 (31.7)
Hypertension, n (%)	9,442 (27.0)
COPD, n (%)	6,468 (18.5)
Diabetes mellitus, n (%)	3,357 (9.6)
Study outcome	
90-day composite acute care event, n (%)	1,438 (4.1)
Behavioral routing policy assignments	
Care access coordination	18,355 (52.5)
Mental health/behavioral health referral	9,716 (27.8)
Hypertension management	2,260 (6.5)
Pulmonary (asthma/COPD) management	1,494 (4.3)
Medication adherence support	1,275 (3.6)
Diabetes management	620 (1.8)
Financial benefits navigation	426 (1.2)
Heart failure management	292 (0.8)
Clinical other	262 (0.7)
Food security referral	98 (0.3)
Transport and utilities navigation	75 (0.2)
Housing navigation	54 (0.2)
Maternal/prenatal support	42 (0.1)
Substance use support	2 (<0.1)

ADI = area deprivation index; COPD = chronic obstructive pulmonary disease; SD = standard deviation. The full-cohort ADI distribution is not reported because only the held-out test set was geocoded for the equity analysis; the test-set ADI distribution is provided in Appendix Table 1. The 90-day composite acute care event is any unplanned

hospitalization or avoidable emergency department visit within 90 days, ascertained by Fine-Gray competing-risks regression with death as the competing event.

Table 2. WPAD preference pair construction and falsification test results.

Category	N
Type 1: ACO staggered onboarding (primary)	
Total patient-window pairs	2,449
Excluded (insufficient pre-2023 claims history)	742
Eligible patient-window pairs	1,707
Primary preference pairs ($Y_{ON} = 0, Y_{OFF} = 1$)	622
Weak-positive pairs ($Y_{ON} = 0, Y_{OFF} = 0$)	1,085
Type 2: CHW waitlist (primary)	
Total patient-window pairs	1,131
Excluded (overlap with Type 1 pairs)	1,131
Non-duplicate eligible pairs	0
Type 3: Medicaid eligibility churn (ITT secondary)	
Intent-to-treat pairs	221
Cross-patient supplementary pairs	
IPTW-weighted cross-patient pairs	14,380
Effective sample size (DR-OPE, behavioral policy)	2,670
Pre-specified falsification tests	
T1 — covariate balance	PASS — max SMD = 0.143 (threshold < 0.20; trajectory covariate added per T1 protocol)
T2 — placebo outcomes	PASS — AUC = 0.50; lagged health does not predict enrollment timing
T3 — administrative predictors	PASS — F-test $p = 0.94$; health trajectory does not predict onboarding order
T4 — density continuity	PASS — no discontinuity in non-care-

Category	N
	management healthcare events at WPAD boundary
T5 — LATE heterogeneity by WPAD type	CONCERN — Type 1 vs. Type 3 LATE difference 58.5% (threshold < 20%); primary restricted to Type 1
T6 — direction-of-time	INCONCLUSIVE BY DESIGN — all Type 1 pairs are OFF→ON; trajectory-slope covariate (T1 protocol) serves as stationarity robustness measure

WPAD = Within-Patient Administrative Discontinuity; ACO = accountable care organization; CHW = community health worker; ITT = intent-to-treat; IPTW = inverse probability of treatment weighting; DR-OPE = doubly robust off-policy evaluation; LATE = local average treatment effect; SMD = standardized mean difference; AUC = area under the receiver operating characteristic curve. Y_{ON} and Y_{OFF} indicate 90-day composite acute care event occurrence (1) or non-occurrence (0) in the ON-window (care management active) and OFF-window (pre-enrollment).

Table 3. Off-policy evaluation: thirteen pre-registered policies (N = 6,995 held-out test patients).

Rank	Policy	DR-OPE	95% CI	ESS	IMI (%)	Coverage	Δ vs. behavioral
1	Oracle ($\hat{\mu}$ -optimal)	0.026	0.025–0.027	6,975	0.0	1.00	+33.7%
2	PEARL (MoE Router)	0.027	0.026–0.029	6,991	2.0	1.00	+31.5%
3	Behavioral cloning SFT (C4)	0.030	0.028–0.031	6,995	4.4	1.00	+24.7%
4	CQL (C8)	0.034	0.032–0.037	6,995	32.0	1.00	+13.6%

Rank	Policy	DR-OPE	95% CI	ESS	IMI (%)	Coverage	Δ vs. behavioral
5	PEARL (MoE Full)	0.034	0.029– 0.040	4,101	38.7	1.00	+13.4%
6	XGBoost (C3)	0.036	0.029– 0.044	4,064	42.3	1.00	+10.4%
7	LACE Index (C1)	0.036	0.029– 0.046	4,114	41.2	1.00	+9.3%
8	HOSPITAL Score (C2)	0.036	0.029– 0.045	4,038	42.7	1.00	+8.9%
9	PEARL (base)	0.039	0.031– 0.049	4,027	42.4	1.00	+2.1%
10	Behavioral policy (reference)	0.040	0.014– 0.067	2,670	27.0	1.00	—
11	Decision Transformer (C7)	0.044	0.037– 0.054	4,761	45.3	1.00	−10.8%
12	Observational DPO (C5)	0.044	0.040– 0.052	6,790	20.4	0.84	−11.0%
13	Causal forest CATE (C6)	0.144	0.136– 0.154	6,788	79.7	1.00	−263.8%

DR-OPE = doubly robust off-policy evaluation policy value (Marginalized DR estimator; lower indicates fewer predicted events); ESS = effective sample size (importance sampling); IMI = Intervention Misalignment Index; Coverage = fraction of policy recommendations within the behavioral support set; CI = 2,000-resample bootstrap. The behavioral policy serves as the reference for Δ (positive = improvement). PEARL (MoE Router) is the best deployable policy: it achieves DR-OPE rank 1 among the twelve deployable policies and IMI of 2.0% — 92.6% of the maximum possible reduction toward

the oracle (0.0%). The oracle policy assigns each patient the minimum- $\hat{\mu}$ intervention and represents a theoretical upper bound; the oracle is not deployable in practice because it uses the same outcome model that defines IMI. The paired bootstrap test for IMI reduction ($H_0: \text{IMI}[\text{PEARL}] \geq \text{IMI}[\text{behavioral}]$; 2,000 resamples) yielded one-sided $p < 0.0001$. Effective sample sizes for all thirteen policies exceeded the pre-specified adequacy threshold of 500 (range 2,670–6,995). All policies except observational DPO (C5; coverage 0.84) had full coverage relative to the behavioral support set, indicating recommendations remain within the historical action distribution. C1–C8 labels follow the pre-registered comparator hierarchy in Appendix Table 5.

Figure legends

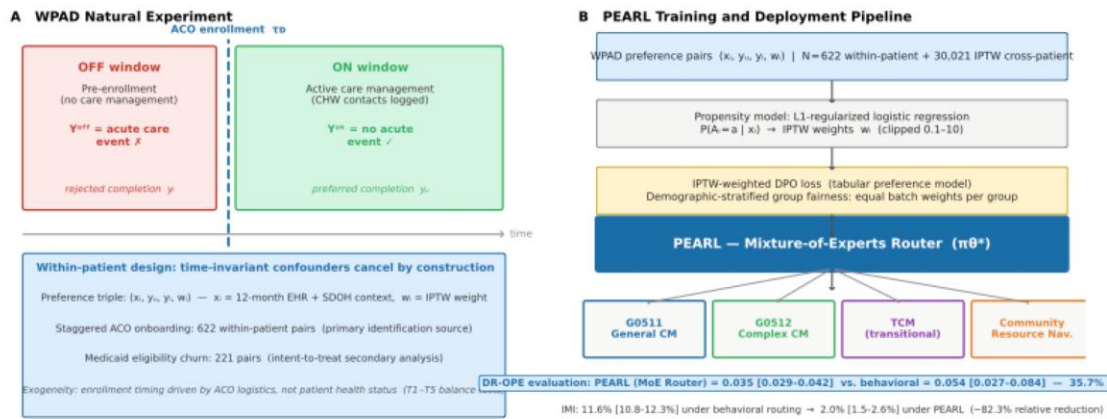


Figure 1. The within-patient natural experiment and the PEARL training pipeline. (A) Staggered enrollment creates a within-patient natural experiment. Each patient has a pre-enrollment window (without care management) and a post-enrollment window (with care management). Each patient is therefore their own control: confounders that do not change over time cancel out by design. When a patient has a favorable 90-day acute care event outcome in the post-enrollment window and an unfavorable outcome in the pre-enrollment window, the patient contributes a within-patient preference pair: the post-enrollment record is the preferred completion (y_w) and the pre-enrollment record is the rejected completion (y_l). Each within-patient preference pair carries an importance weight (w_i) reflecting the propensity of intervention assignment given the patient's covariates (x_i). (B) The PEARL training pipeline. We weight 622 primary and 221 intent-to-treat within-patient preference pairs by an L1-regularized multinomial propensity model (weights clipped to [0.1, 10]). We train a tabular preference-learning model with demographic-stratified group-stratified fairness on those weighted pairs. At inference, the

trained PEARL routing policy assigns each patient to one of fourteen intervention types via the expert router.

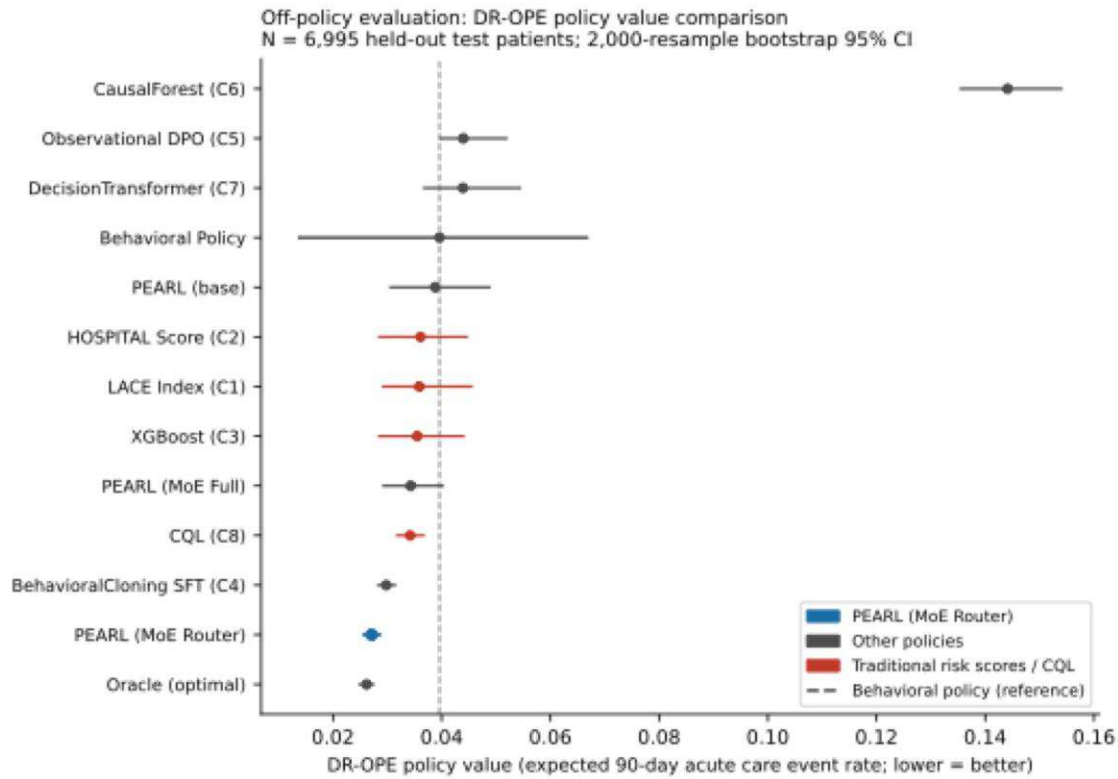


Figure 2. Off-policy evaluation across all thirteen next best action selection policies (N = 6,995 held-out test patients; 2,000-resample bootstrap 95% CI). Lower DR-OPE values mean fewer predicted 90-day acute care events. The dashed reference line marks behavioral routing. The oracle policy is a theoretical upper bound (not deployable in practice because it uses the same outcome model that defines intervention misalignment). PEARL is the best deployable policy: it achieves DR-OPE of 0.027 (95% CI, 0.026–0.029) and reduces intervention misalignment to 2.0% — 92.6% of the maximum possible reduction toward the oracle. Behavioral cloning on within-patient preferences (C4; rank 3) achieves DR-OPE of 0.030 (95% CI, 0.028–0.031) and reduces intervention misalignment to 4.4%, confirming that the within-patient signal alone produces most of the reduction. PEARL ablations without group-stratified fairness (PEARL base rank 9, intervention misalignment 42.4%; PEARL MoE Full rank 5, intervention misalignment 38.7%) increase intervention misalignment above behavioral routing. Observational DPO (C5; rank 12) confirms that the within-patient signal is necessary for intervention misalignment reduction. Causal forest (C6; rank 13; DR-OPE 0.144) fails because the underlying treatment effect estimates extrapolate to unobserved next best actions. The primary reported result is PEARL’s intervention misalignment reduction (27.0% → 2.0%; $\Delta = 25.0$ percentage points; $p < 0.0001$) shown in the intervention misalignment column of Table 3.

Appendix Tables and Figures (Supplementary Information): Appendix Table 1 (IMI by ADI quintile, dual eligibility, language, race/ethnicity); Appendix Table 2 (sensitivity analysis individual specifications); Appendix Table 3 (sensitivity analysis summary); Appendix Table 4 (falsification test details); Appendix Table 5 (comparator specifications); Appendix Table 6 (nuisance model calibration); Appendix Table 7 (per-intervention-type positivity); Appendix Table 8 (within-WPAD secondary IMI); Appendix Table 9 (TRIPOD-AI checklist); Appendix Figure 1 (patient flow); Appendix Figure 2 (IMI by ADI quintile); Appendix Figure 3 (Camden reanalysis); Appendix Figure 4 (sensitivity forest plot); Theorems S1–S2 (IMI Reduction; DPO–DR–OPO equivalence); Algorithms S1–S2 (WPAD pair construction; PEARL training).

Supplementary Files

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

- [PEARLsupplementv4.docx](#)