BEYOND PATIENT-LEVEL TRIAGE: PRE-ARRIVAL CARE NAVIGATION UNDER SHARED, LIMITED CAPACITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Pre-arrival care navigation, recommending a care setting from a patient's pre-arrival description, offers an opportunity to direct patients to appropriate care before emergency-department (ED) resources are committed. EDs often care for patients whose needs could be met in lower-acuity settings, yet emergency triage largely assesses acuity and resource needs after ED arrival, when care-routing options have already narrowed. The decision serves two interests: each patient seeks capable and timely care, while the health system must keep care timely for patients with urgent needs; under constrained capacity, lower-acuity demand competes with later, more urgent arrivals. The patient's interest can be scored at every routing decision, but the system's interest is determined by a whole run of decisions and observed only after them, so it is hard to learn from single decisions. We introduce Stackelberg-Inspired Commitment Learning (SICL), a hierarchical framework for language-model routing policies that gives the system's interest a decision of its own. A system-level routing posture, held across a segment of arrivals, sets site-level incentives; each patient's destination is then chosen in the patient's own interest under those incentives; and during training the posture is credited with the segment's system-level outcome, according to how closely the routing followed it. We further construct TRIAGEROUTE, a dynamic emergency-care benchmark derived from MIMIC-IV-ED v2.2 that models heterogeneous care sites, travel delays, evolving queues, and workload-sensitive outcomes, where every policy sees the same patient information and care-network state and is given the same instructions. Despite identical information, policies differ widely in timely access for patients with urgent needs; in controlled simulation, SICL achieves the highest timely access among policies given the same information while also achieving the highest patient utility, at the cost of reduced capacity efficiency.

1 INTRODUCTION

Clinical AI systems and large language models are increasingly studied as support for triage and care-seeking decisions (Lu et al., 2024; Tang et al., 2024). They assess the patient in front of them, yet before a patient enters the emergency-care system the actionable question is a different one: where should this patient seek care now? Nurse advice lines, prehospital field-triage protocols (Bunn et al., 2005; Sasser et al., 2012), and patient-facing symptom checkers (Semigran et al., 2015) already address this question, and the answer matters. A substantial share of ED visits could be managed in lower-acuity settings (Weinick et al., 2010), yet which visits these are is hard to tell from the presenting complaint alone (Raven et al., 2013). Once a patient arrives, the ED must screen and stabilize them (U.S. Congress, 1986), so the choice of care setting is effectively made before arrival.

The decision serves two interests. Each patient wants a capable site that is close and will see them soon, while the health system, serving every patient who shares the same sites, needs urgent patients (Emergency Severity Index, ESI, levels 1–2) to be seen in time. When capacity is ample these interests align; under crowding they diverge, as in congested routing, where choices that are best for each individual need not be best for all (Roughgarden & Tardos, 2002). For example, several low-acuity patients sent to a nearby tertiary site because it is closest occupy its beds and clinicians, and a stroke patient arriving twenty minutes later, though first in the queue, waits for a bed (Figure 1;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

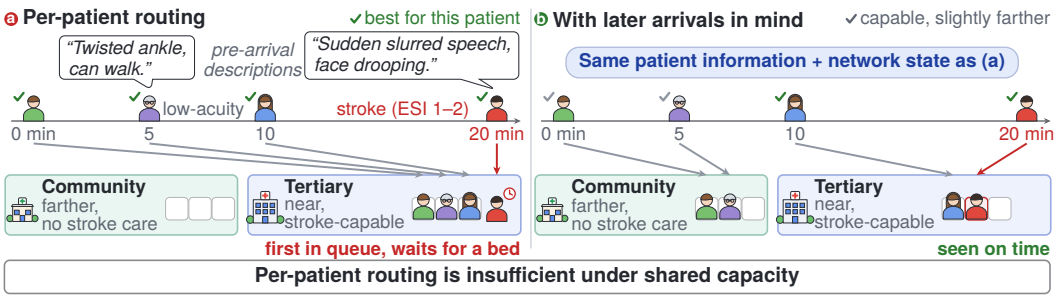

Figure 1: An example of Static Triage Myopia: each routing choice is reasonable on its own, yet together they delay a later urgent patient.

Carter et al., 2014; Morley et al., 2018; Hurwitz et al., 2014). We study *pre-arrival care navigation*: inferring each patient's needs from a pre-arrival description and choosing a destination that serves the patient while protecting the system's ability to see urgent patients in time.

Existing work addresses each side separately. Clinical triage models and language-model agents infer acuity and resource needs for individual patients (Gilboy et al., 2023; Tanabe et al., 2004; Ivanov et al., 2021; Miles et al., 2020; Porto, 2024; Fan et al., 2025; Wang et al., 2025), but are generally evaluated patient by patient. Classical routing methods coordinate across arrivals using congestion, anticipated delay (Winston, 1977; Weber, 1978; Selen et al., 2016), diversion (Laskowski & Mukhi, 2009; Ong et al., 2025), or capacity pricing (Gupta et al., 2013; Mahdavi et al., 2012), but assume clinically structured demand rather than needs inferred from patients' pre-arrival descriptions. Under tight capacity, individually reasonable recommendations can collectively delay later urgent patients (Figure 1), a failure we call *Static Triage Myopia*: making care decisions as if each patient were the only one. Giving every policy the same information is not enough: with identical patient information, network state, instructions and rules, evaluated policies still differ widely in timely access for urgent patients (§6.1). The difference lies not in what a policy sees but in how it decides.

Pre-arrival navigation therefore asks for more than per-patient reasoning. A good policy must infer each patient's needs from a description, and it must also account for how each placement changes the capacity left for later arrivals. The second is hard to learn, because the two interests are scored over different spans. The patient's interest can be scored at every decision, from the site's capability, the trip and the wait. The system's interest cannot: it depends on how a whole run of decisions spreads load across sites, it is observed only after those decisions are made, mixed with whichever patients happened to arrive, and so it is hard to credit to any single decision.

To address this, we propose Stackelberg-Inspired Commitment Learning (SICL), which gives the system's interest a decision of its own. It borrows the move order of Stackelberg games: a leader commits first, and a follower responds in its own interest. The leader sets a system-level routing posture, such as putting urgent patients first or spreading non-urgent demand, and holds it over a segment of arrivals. Like a toll in congested routing, the posture makes the sites it wants to keep free more costly and the sites it wants patients to use less costly. The follower, a patient-level policy, still chooses each destination in the patient's own interest, now counting these incentives, so the system's interest reaches every routing decision while the follower keeps acting for the patient. Holding the posture over a segment gives the segment's system-level outcome a single decision to answer to. During training, SICL measures that outcome for each segment and credits it to the posture through Response-Attributed Credit, which counts that outcome only as far as the routing followed the posture. The posture thus learns which incentives to set, and when, from the responses it observes rather than by solving for an equilibrium.

Our contributions are threefold:

- **Two-interest pre-arrival care navigation.** We formulate pre-arrival routing as a sequential decision that answers to each patient and to the system, and identify Static Triage Myopia as a failure that the same information does not remove.

- **SICL.** A leader–follower framework in which a system-level posture steers routing via site-level incentives, patient-level routing serves each patient's own interest, and Response-Attributed Credit credits system-level outcomes to the posture by how the routing responds.
- **New benchmark and evaluation.** To foster research in pre-arrival care navigation, we introduce TRIAGEROUTE, a dynamic simulation benchmark that turns MIMIC-IV-ED (Johnson et al., 2023) records (chief complaint, triage vital signs, history and a reconstructed pre-arrival interview) into nationally calibrated full-day arrival streams (McCaig & Nawar, 2005) and routes them through a care network whose sites differ in capability, resources, and travel time, scoring policies over whole arrival streams under a same-information protocol. On TRIAGEROUTE, SICL achieves the highest urgent timely access among policies given the same information, improving over PPO (0.532 vs. 0.458) while also achieving the highest patient utility, at some cost in capacity efficiency; without Response-Attributed Credit, the posture stops adapting and the gain disappears.

2 RELATED WORK

Care navigation under shared capacity. Emergency triage and clinical decision support estimate patient-level acuity, anticipated resource needs, disposition, or an appropriate care setting from chief complaints, vital signs, medical history, and triage notes (Gilboy et al., 2023; Tanabe et al., 2004; Ivanov et al., 2021; Miles et al., 2020; Porto, 2024), and language-model systems extend this line through interactive clinical reasoning and multi-agent deliberation (Lu et al., 2024; Tang et al., 2024; Fan et al., 2025; Wang et al., 2025). The operational side of the same decision developed separately. Join-the-shortest-queue is the canonical load-balancing rule (Winston, 1977; Weber, 1978), and shortest-expected-delay routing generalizes it to servers with different service characteristics (Selen et al., 2016). Emergency care adds destination control at the network level: ambulance diversion redirects arrivals away from saturated sites, with mixed reported effects (Ong et al., 2025) and a documented loss of pooling when each site decides on its own (Deo & Gurvich, 2011); agent-based simulations adapt random early detection (Floyd & Jacobson, 1993) into hospital diversion policies that redirect patients with increasing probability as load rises (Laskowski & Mukhi, 2009); and a further line attaches a dynamic price to scarce capacity and updates it online (Gupta et al., 2013; Mahdavi et al., 2012). Queueing models also prioritize patients whose type is observed only through an imperfect signal (Argon & Ziya, 2009; Sun et al., 2018), and proactive diversion uses predicted arrivals to divert before congestion builds (Xu & Chan, 2016). Benchmarks follow the same split: MIMIC-IV-ED benchmarks score per-visit predictions such as hospitalization and critical outcomes (Xie et al., 2022), and clinical agent benchmarks simulate one patient's encounter at a time (Schmidgall et al., 2024; Jiang et al., 2025). Despite their success, the clinical line decides for one patient at a time and the operational line starts from a structured class or its estimate, so neither addresses the challenge set out in §1: inferring need from a patient's own description while protecting the capacity that all patients share. SICL keeps both halves in one policy.

Hierarchical decision-making and credit assignment. Hierarchical reinforcement learning separates decisions that act on different timescales, through temporally extended actions (Sutton et al., 1999) or an upper level that issues goals to a lower one (Vezhnevets et al., 2017; Nachum et al., 2018). For language agents, ArCHer trains a turn-level critic above a token-level policy (Zhou et al., 2024), and HiPER assigns credit across hierarchical decision spans (Peng et al., 2026); SICL is built on the latter. SICL differs in what the upper level is for and how it is held accountable: it holds a routing posture on behalf of the system, and it is credited with the segment's system-level outcome as far as the routing followed that posture. This borrows the commit-then-respond structure of Stackelberg games, in which a leader commits and followers respond in their own interest, as a social planner does in economic policy design (Zheng et al., 2022b) and in Stackelberg formulations of language-model training (Makar-Limanov et al., 2024; Karten et al., 2025; Zheng et al., 2022a). GT-PMARL adds an estimate of team state and a load-based shadow price directly to each agent's payoff (Zhang et al., 2026); in SICL, the leader carries the system's interest and learns from delayed system-level outcomes which site-level incentives to set, so the patient-level payoff gains only that incentive. Constrained reinforcement learning adds the system's interest to each decision's return as a Lagrangian penalty (Altman, 1999; Tessler et al., 2019); when the constraint is on a whole day's urgent access, that penalty reaches every decision alike, much as the shared episode-level reward does for PPO (§6.2).

3 PRE-ARRIVAL CARE NAVIGATION

Task. Let $\mathcal{H}$ be a network of care sites. Patient i arrives at time t_i with a pre-arrival description o_i (chief complaint, early vital signs, relevant history and a pre-arrival interview), and the policy also observes the network state s_{t_i}, which records each site's capabilities, travel time and current load. A policy outputs a destination $a_i \in \mathcal{H}$; language-model policies first state a clinical assessment c_i of acuity and anticipated resource needs, which is provisional and is replaced by formal triage on arrival.

Environment. Every patient goes to the chosen site. Sites serve waiting patients in order of true acuity, and a patient who has been given a bed keeps it. The patient's true needs z_i are hidden from the policy; if the chosen site cannot provide them, the patient is transferred to the nearest site that can, and the time waited plus the extra travel count as waiting. These rules are stated in every policy's instructions. The state then advances as $s_{t_{i+1}} = T(s_{t_i}, a_i, z_i, \xi_i)$, where ξ_i collects concurrent arrivals and service completions, so each placement changes the load later patients face.

Two interests and their measures. The patient's interest is the payoff u_i of a rational patient who prefers a higher-capability site, a shorter wait and a shorter trip, with configurable preference weights (Appendix A.2); we report its normalized mean as patient utility. The system's interest, our primary endpoint, is urgent timely access (UTA):

$$\mathrm{UTA} = \frac{1}{|\mathcal{I}_{\mathrm{urg}}|} \sum_{i \in \mathcal{I}_{\mathrm{urg}}} \mathbb{I}[W_i \le 30 \text{ min}], \qquad \mathcal{I}_{\mathrm{urg}} = \{\, i : z_i^{\mathrm{ESI}} \le 2 \,\},$$

where W_i is the wait after arrival at the site, so travel enters u_i but not UTA. Ground-truth acuity is never shown to a policy: it drives the simulator, and through it every training signal, and it defines who counts as urgent in evaluation, so a policy cannot raise UTA by over-predicting urgency.

Static Triage Myopia. We call a policy myopic when it chooses $a_i \in \arg\max_{h \in \mathcal{H}} \mathbb{E}\,[u_i(h) \mid o_i, s_{t_i}]$: it weighs the site's capability, the trip and the wait, but not the load its choice leaves for later arrivals.

4 STACKELBERG-INSPIRED COMMITMENT LEARNING

The patient's payoff u_i does not count the capacity a choice leaves for later arrivals, so maximizing it patient by patient is exactly the myopic rule of §3. SICL therefore gives each interest its own role. A patient-level follower is paid, at every decision, the patient's payoff u_i and an incentive set by the leader. A leader holds a routing posture over a segment of arrivals and, during training, is credited with the system-level outcome measured over that segment. Both roles are produced by one language model and optimized with the span-level hierarchical PPO of HiPER (Peng et al., 2026); what SICL adds is what the upper level holds, how it reaches the lower level, and what it is credited with.

4.1 FRAMEWORK: COMMIT, RESPOND, ATTRIBUTE

SICL borrows three elements of a Stackelberg game: the leader moves first and commits, the follower observes the commitment before it acts, and the two are paid differently. A rollout of one policy π_θ factorizes accordingly:

$$\pi_\theta(\tau) = \prod_k \Big[\underbrace{\pi_\theta\big(g_k \mid s_{t_k}, g_{k-1}\big)}_{\text{Commit (leader)}} \prod_{i \in \mathcal{I}_k} \underbrace{\pi_\theta\big(c_i, a_i \mid o_i, s_{t_i}, g_k\big)}_{\text{Respond (follower)}} \Big], \tag{1}$$

where segment k starts at t_k and $\mathcal{I}_k$ is the set of patients routed under posture g_k. The follower acts for each patient: it is paid that patient's payoff plus an incentive set by the posture (§4.3). The leader acts for the system: it alone is credited with the segment's system-level outcome, through the credit it is given for the routing it induced (§4.4). Token order is move order, so the two roles share parameters but differ in objective, information and timing. SICL borrows the move order and the incentive structure, not the solution concept: it does not compute an equilibrium, and unlike Stackelberg learning methods that differentiate through the follower's best response (Zheng et al., 2022a), the leader learns which posture to commit to, and when, only from the routing and outcomes it observes.

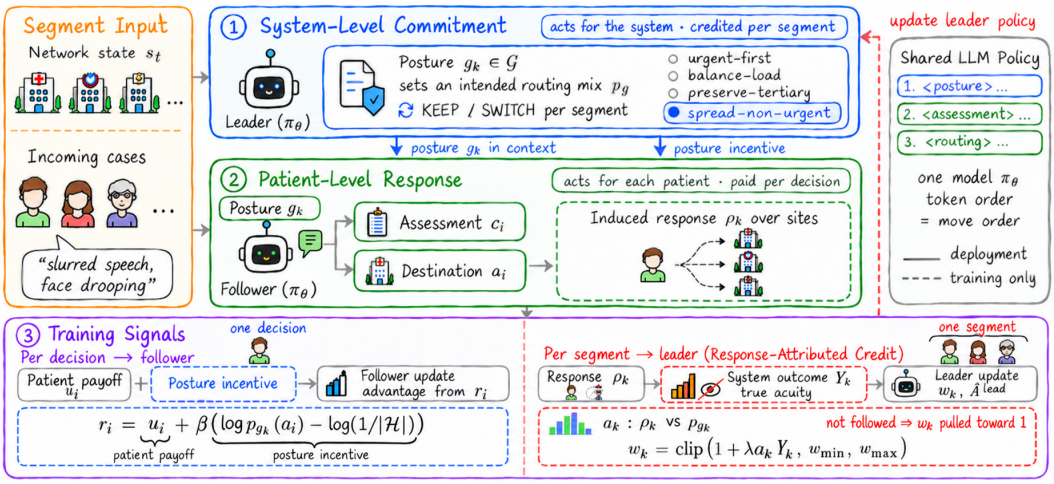

Figure 2: Overview of SICL. The leader commits to a posture for a segment (1); the follower routes each patient for the patient's payoff under the posture's incentive (2; 3, left); in training, the leader is credited with the segment's system-level outcome as far as the routing followed it (3, right).

4.2 Commit: A Routing Posture

The leader's decision persists across arrivals, so that a segment's outcome has a single decision to answer to. The leader chooses from a small family of human-readable postures, $\mathcal{G} = \{\text{urgent-first}, \text{balance-load}, \text{preserve-tertiary}, \text{spread-non-urgent}\}$, each of which fixes a routing prior p_g over sites (Appendix B). The prior depends only on the posture; which posture is chosen depends on the state. At each arrival the leader reads the network state and either keeps its current posture or switches, so it also decides how long a commitment lasts.

4.3 Respond: Routing under Posture Incentives

For patient i in segment k, the follower reads o_i, s_{t_i} and the posture g_k, states an assessment c_i, and then chooses a_i. At each decision during training, its reward is

$$r_i = \underbrace{u_i}_{\text{patient payoff}} + \underbrace{\beta\Big(\log p_{g_k}(a_i) - \log \tfrac{1}{|\mathcal{H}|}\Big)}_{\text{posture incentive}}, \tag{2}$$

which subsidizes the sites the posture favors and taxes the others, much like a toll. The posture never changes u_i; it enters the follower's reward only through this term. The incentive exists only during training: at deployment the follower sees the posture as text in its context and routes as it learned to under that posture's incentive. Unlike action masking or hard constraints, the incentive leaves the final choice to the follower, so whether the routing follows the posture becomes an observable outcome, summarized by the segment's realized site distribution $\rho_k(h) = |\mathcal{I}_k|^{-1} \sum_{i \in \mathcal{I}_k} \mathbb{I}[a_i = h]$.

4.4 Attribute: Response-Attributed Credit

A segment's system-level outcome arrives late, is only partly shaped by the posture, and is mixed with whichever patients happened to arrive. In option-based and feudal hierarchies the upper level is credited with the segment return whether or not the lower level carried out its intent; Response-Attributed Credit (RAC) keeps that ordinary credit and lets the segment's system-level outcome count in proportion to how closely the routing followed the posture. The segment's system-level outcome is approximated by a training-only proxy Y_k, computed by the simulator from true acuity. It combines the share of urgent patients reaching tertiary care, the share of ESI 4–5 patients kept off tertiary care, and the relief of tertiary queue pressure, passing their weighted sum through a $\tanh$ (Appendix B). No policy observes Y_k. The leader's credit weight is

$$w_k = \text{clip}\left(1 + \lambda\, \alpha_k Y_k,\ w_{\min},\ w_{\max}\right), \qquad \alpha_k = \alpha_{\min} + (1 - \alpha_{\min}) \exp\left(-D_{\text{JS}}(\rho_k \,\|\, p_{g_k})/\tau\right), \tag{3}$$

Table 1: TRIAGEROUTE at a glance.

Cases	About 8,000 MIMIC-IV-ED emergency cases across ESI 1–5, reconstructed from de-identified records
Policy input	Demographics, chief complaint, triage vital signs, medical background and earlier ED visits, a pre-arrival interview; the live care-network state; the environment's rules and the routing objective
Hidden from policies	Nurse-assigned ESI, resource needs from the documented ED course, diagnoses, laboratory and imaging results
Arrivals	Full-day arrival streams following NHAMCS 2019–2022 daily and weekly patterns; evaluation arrivals from held-out encounters
Care network	Two tertiary hospitals, two community hospitals, an urgent-care center and a clinic, with capability profiles, effective capacities and road-network travel times
Outcomes	Urgent timely access, patient utility, capacity efficiency and transfer rate, scored in simulation

with $0 < w_{\min} < 1 < w_{\max}$, and it multiplies the leader's ordinary advantage $\hat{A}_k^{\text{lead}}$. RAC reweights rather than re-signs: because $w_k > 0$, a segment's system-level outcome never reverses the direction set by $\hat{A}_k^{\text{lead}}$, but it sets how strongly that segment moves the posture. Notably, a posture the routing did not follow falls back toward its ordinary credit, and because α_k is computed from sampled destinations and enters only as a constant weight, no gradient teaches the follower to deviate in order to change the leader's credit. Unlike HIRO (Nachum et al., 2018), which relabels the upper level's goal to match what the lower level did, RAC keeps the committed posture and scales its credit by the fit.

4.5 TRAINING AND INFERENCE

Both roles are optimized in one rollout with span-level clipped PPO and a KL penalty to a reference policy: the posture tokens use $w_k \hat{A}_k^{\text{lead}}$ with w_k treated as a constant, and the destination tokens use an advantage computed from r_i. Appendix B gives further details. At inference the leader keeps or switches its posture from the network state and the follower routes each patient under it; neither the incentive nor Y_k is computed.

5 EXPERIMENTAL SETUP

5.1 TRIAGEROUTE AND THE SAME-INFORMATION PROTOCOL

Evaluating care navigation requires clinical labels grounded in real records, so that a routing choice can be judged, and a shared care network whose load changes over time, so that each choice affects the patients who come after it. TRIAGEROUTE provides both (Table 1): it turns about 8,000 MIMIC-IV-ED cases into pre-arrival descriptions with record-grounded labels and routes them through a care network whose load evolves with every placement. Because the records carry patient-level labels rather than destination labels, policies are scored by their simulated downstream outcomes (Appendix C). Every policy receives the same inputs and the same statement of the environment's rules and the routing objective; the encounter's true labels are hidden from all policies and used only by the simulator and the metrics (Appendix C.6). Policies therefore differ only in how they decide and what they learn from.

5.2 METRICS AND REPORTING

Urgent timely access (UTA, §3) is the primary endpoint. We also report the two quantities it may trade against: patient utility, the normalized payoff of a rational patient under the default preference profile, and capacity efficiency $E \in [0, 1]$, which combines crowding at high-capability sites, overall throughput, and whether scarce sites serve the patients who need them. Because these measures trade off, we also report safety-constrained routing welfare (SCRW), which combines UTA, E and patient utility as a weighted geometric mean with weights 0.6, 0.2 and 0.2 (Appendix A). The transfer rate is the share of patients whose true needs exceed the capability of the chosen site. Each policy is evaluated in five runs with different random seeds over identical arrival sequences (four for KATE

Table 2: Main results on TRIAGEROUTE, mean ± SD over five evaluation runs (four for KATE with Qwen3-32B). UTA is urgent timely access, the primary endpoint; Utility is patient utility; E is capacity efficiency; SCRW combines the three (§5.2). Prompted policies use Qwen3 (Yang et al., 2025) and GPT-OSS-120B (OpenAI, 2025) backbones; trained policies use Qwen2-1.5B-Instruct (Yang et al., 2024). Definitions are in Appendix A.

Family	Method (Model)	UTA↑	Utility↑	E ↑	SCRW↑
Prompted per-arrival policies	*Vanilla*				
	Qwen3-4B	0.327±.038	0.665±.006	0.457±.015	0.402±.027
	Qwen3-14B	0.181±.031	0.649±.007	0.446±.023	0.279±.027
	Qwen3-32B	0.287±.039	0.663±.004	0.373±.010	0.357±.031
	GPT-OSS-120B	0.467±.035	0.663±.006	0.503±.009	0.508±.024
	KATE (Liu et al., 2022)				
	Qwen3-32B	0.470±.035	0.667±.004	0.527±.004	0.516±.023
	GPT-OSS-120B	0.462±.034	0.651±.009	0.537±.004	0.509±.023
	TriageAgent (Lu et al., 2024)				
	Qwen3-32B	0.441±.028	0.653±.006	0.557±.013	0.500±.020
	GPT-OSS-120B	0.462±.014	0.648±.005	0.541±.006	0.510±.009
	MedAgents (Tang et al., 2024)				
	Qwen3-32B	0.359±.039	0.657±.012	**0.560**±.014	0.442±.030
	GPT-OSS-120B	0.366±.069	0.649±.005	0.552±.011	0.444±.051
Trained policies	SFT	0.313±.081	0.576±.004	0.402±.008	0.369±.057
	PPO (Schulman et al., 2017)	0.458±.023	0.669±.005	0.407±.005	0.483±.016
	HiPER (Peng et al., 2026)	0.246±.034	0.620±.008	0.235±.015	0.294±.027
	SICL (ours)	**0.532**±.045	**0.680**±.002	0.378±.014	**0.521**±.028

with Qwen3-32B), and we report the mean ± SD. Comparisons use Welch's t-test over runs; the primary comparison is SICL against PPO on UTA.

5.3 COMPARED POLICIES

Prompted per-arrival policies. Frozen language models reason about one arrival at a time: vanilla prompting of Qwen3 (Yang et al., 2025) (4B, 14B, 32B) and GPT-OSS-120B (OpenAI, 2025); KATE (Liu et al., 2022), which retrieves semantically similar cases as demonstrations; and the medical multi-agent frameworks TriageAgent (Lu et al., 2024) and MedAgents (Tang et al., 2024). They test whether reasoning about a single arrival is sufficient for care navigation.

Trained policies. All are fine-tuned from the same base model. SFT imitates clinical-assessment and routing targets. The others are trained with reinforcement learning on each routed patient's payoff u_i together with an episode-level reward for the day's urgent timely access relative to a running baseline, with the same weight for every method. PPO (Schulman et al., 2017) is the flat counterpart, in which that episode-level reward is shared by every routing decision of the day. HiPER (Peng et al., 2026), on which SICL is built, assigns credit across hierarchical decision spans. SICL adds the posture, its incentive and RAC (§4); its ablations are introduced in §6.3.

Classical routing rules. As privileged references, we run capability-constrained queueing, diversion and coordination rules (Appendix D.1) on the acuity and resource needs recorded by the simulator, which no evaluated policy observes.

Implementation. All trained policies use Qwen2-1.5B-Instruct (Yang et al., 2024) and the same simulator interface; SICL additionally emits a posture span, in which the leader keeps or switches its posture. All policies decode at temperature 0.1, and SFT, PPO and SICL produce a valid action for at least 99.9% of arrivals. The classical-rule comparison uses a separate harness with the same SICL checkpoint.

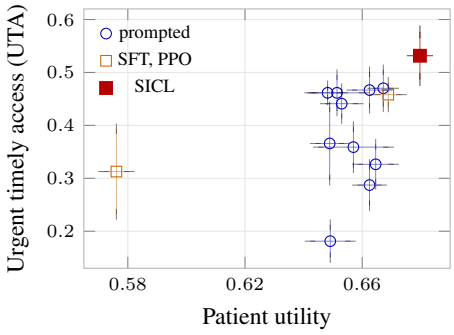

Figure 3: Patient utility does not rank policies by urgent access. Each marker is one policy of Table 2: patient utility (x) against urgent timely access (y), mean $\pm$ SD over runs; SICL is the red square.

6 Results

We answer four research questions. **RQ1**: under identical information, do policies that look alike patient by patient also see urgent patients in time alike (§6.1)? **RQ2**: does SICL raise urgent timely access, and who pays for it (§6.2)? **RQ3**: which part of SICL produces the gain, and where does it arise (§6.3–6.4)? **RQ4**: do the conclusions depend on a particular setting (§6.5)?

6.1 Same Information, Different Outcomes

Under identical information, urgent timely access varies 2.6-fold across prompted policies, from 0.181 to 0.470 (Table 2). Neither scale nor agentic scaffolding closes the gap: Qwen3-14B falls below Qwen3-4B (0.181 vs. 0.327), MedAgents falls below vanilla GPT-OSS-120B (0.366 vs. 0.467), and the five strongest configurations cluster at 0.44–0.47.

Read patient by patient, the same policies look alike: their patient utility spans only 0.019 (0.648–0.667) and does not rank them by urgent access (Spearman $\rho = 0.14$; Figure 3).

Urgent patients do not miss the 30-minute window because they are sent where they cannot be treated: transfers are nearly absent among the misses. For most of the policies we analyze, 77–88% of misses occur when every site able to treat the patient is already full at the moment of the decision (Appendix D.2). Earlier arrivals, as in Figure 1, settled whether a bed is free; the patient's own routing then decides where one frees first. What separates these policies is not what they see but how they route, the Static Triage Myopia of §3.

6.2 Main Comparison

Learning from the environment does not by itself close the gap: PPO (0.458) stays within the range of the strongest prompted configurations, and SFT (0.313) falls below it. SICL reaches 0.532 ± 0.045, the highest UTA among policies under the same-information protocol: 0.074 above PPO (95% CI $[0.018, 0.129]$, $p = 0.018$) and 0.062 above the strongest prompted configuration (KATE with Qwen3-32B). Both receive the same episode-level reward for the day's urgent timely access (§5.3), but in PPO it is shared by every routing decision of the day and credits none of them in particular. SICL additionally credits a segment-level system outcome to the posture held over it, weighted by how closely the routing followed.

Classical rules given privileged acuity and resource needs reach a comparable level. Given the recorded acuity and resource needs, which no policy in Table 2 observes, the best of seven rules reaches 0.513 UTA and the others 0.33–0.50; SICL, reading only the descriptions, reaches 0.509 ± 0.054 in that same harness (Appendix D.1).

The gain is not taken from the routed patient. SICL also has the highest patient utility (0.680 vs. 0.669 for PPO); urgent patients wait less on average (33.9 vs. 37.0 minutes); and non-urgent patients' median wait stays within 3% of PPO's while their 90th percentile is 9% shorter. Its transfer rate (1.0%) is on par with PPO's (1.2%) and well below that of GPT-OSS-120B (4.4%). The cost falls on capacity efficiency (0.378 vs. 0.407). Because the follower routes in each patient's own interest, SICL sends more non-urgent patients to higher-capability sites, which the default preference profile rewards (75% vs. 69%), which raises patient utility and lowers E (Appendix D.3); its gain for urgent

Table 3: Ablation of SICL. Components are added in turn; PPO is the flat reference without a posture. Only with RAC does UTA rise above PPO and the leader switch. Mean ± SD over five runs. *Switch*: share of arrivals at which the leader issues SWITCH.

Variant	Posture	Incentive	RAC	UTA↑	Utility↑	SCRW↑	Switch
PPO (flat reference)				0.458±.023	0.669±.005	0.483±.016	–
SICL w/o incentive & RAC	✓			0.477±.045	0.671±.001	0.476±.027	0%
SICL w/o RAC	✓	✓		0.455±.041	0.663±.003	0.457±.025	0%
SICL	✓	✓	✓	**0.532**±.045	**0.680**±.002	**0.521**±.028	44.6%

patients comes from bedding them faster when the network is full (§6.4), not from turning others away. We call this the access–efficiency trade-off. Weighing the three measures together, SICL also has the highest SCRW (0.521).

6.3 Ablation Study

The leader may commit to any posture, much as a planner may announce any policy; whether the posture works shows only in how the system responds, and RAC is what holds the leader to that response. Table 3 adds SICL's components in turn. Without RAC, the posture and its incentive remain, but nothing asks whether the posture worked: the leader stops adjusting it (switch rate 0% vs. 44.6%), and UTA falls from 0.532 to 0.455 ($p = 0.022$), back to the level of flat PPO (0.458). The variant that also drops the incentive stays at the same level (0.477). The system's interest needs not only a decision of its own but credit for that decision.

6.4 Where the Gain Arises

SICL's gain arises in the hardest moments. When every site able to treat an urgent patient is full at the moment of the decision, SICL brings 52% of such patients to a bed within 30 minutes, against 44% for PPO; otherwise the two are level (57% each; Appendix D.2).

Routing follows the posture at inference, where no incentive is computed: under preserve-tertiary, 53% of arrivals go to non-tertiary sites, against 19% otherwise, and the mutual information between posture and destination tier exceeds its shuffled-label null 4–15-fold in every run ($p < 0.001$). The leader also adjusts with load: it issues SWITCH at 36% of arrivals when tertiary load is in its lowest tertile and at about 49% in the upper two (χ^2, $p < 10^{-15}$). The posture is how the system's interest reaches each routing decision, and the leader adjusts it as the network fills.

6.5 Robustness

SICL leads PPO and the vanilla prompted policies at every threshold from 15 to 60 minutes, and its relative lead is largest at the strictest one (0.195 at 15 minutes, against 0.130 for PPO and 0.154 for GPT-OSS-120B; Appendix D.4). Its lead over PPO holds at every decoding temperature and widens as decoding becomes noisier: $+0.074$, $+0.113$ and $+0.121$ UTA at $T = 0.1$, 0.4 and 1.0 (all $p < 0.02$; Appendix D.5). A second checkpoint of the same run scores 0.533.

7 Conclusion and Limitations

This work studies pre-arrival care navigation, in which the patient's interest is scored at every routing decision and the system's interest only over a run of decisions. SICL gives the system's interest a decision of its own: a posture held over a segment of arrivals and credited with that segment's system-level outcome as far as the routing followed it. In simulation on TriageRoute, it improves urgent timely access and has the highest patient utility, at a cost in capacity efficiency.

Our evidence comes from simulation with full compliance, MIMIC-IV cases from one academic medical center, and possibly incomplete pre-arrival information; partial compliance, other care systems and clinical validation are left to future work (Appendix E).

AI USE STATEMENT

We used generative AI tools in the construction of the benchmark and in the preparation of this manuscript, and disclose these uses below.

Dataset construction. Language models were used to build the benchmark from de-identified MIMIC-IV records (Appendix C): to restructure each record into a case, to extract the emergency-department resource needs it documents (GPT-OSS-120B), and to reconstruct the pre-arrival interview the record implies (GPT-OSS-120B and Qwen3-32B for record-based cases; Qwen3.5-27B for the supplementary low-acuity cases). All of these are open-weight models served locally; no record was sent to an external API. Because the raw cohort contains few low-acuity encounters, the patient pool was supplemented with additional ESI 4–5 cases of the same schema, retained only when two independent clinical-assessment prompts agreed on both ESI and resource need. Ground-truth ESI labels and diagnosis groups are derived from MIMIC-IV records and ICD codes, not from a language model.

Manuscript preparation. Generative AI tools were used for editing and for mechanical restructuring of the manuscript source. All text was reviewed and revised by the authors.

Other uses. Generative AI tools were also used to polish the writing and to refine figures.

Responsibility. Language models are also the object of study in this work: the routing policies we train and evaluate are LLM-based, as described in Section 4. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, claims, and artifacts.

ETHICS STATEMENT

Data. Our patient cases are derived from MIMIC-IV (Johnson et al., 2023), a publicly available, de-identified electronic health record dataset distributed under the PhysioNet Credentialed Health Data License. No protected health information or proprietary care-site operational data are included in our simulator, and no raw clinical notes: case text is reconstructed from de-identified records (Appendix C). The simulated care network comprises publicly listed facilities.

Scope of claims and intended use. Our findings come from simulation. On TRIAGEROUTE, SICL improves urgent timely access over PPO from 0.458 to 0.532 (Table 2), at a lower capacity efficiency.

However, we explicitly state that **a significant gap** remains between our simulation-based findings and real-world clinical deployment. A truly functional system requires more than algorithmic optimization; it demands a multi-party coordination across multiple stakeholders:

- **Human-in-the-loop Governance:** AI recommendations must remain strictly advisory. Treating clinicians must retain ultimate authority, and the system must account for the high-stakes, emotional, and physical realities of emergency care that simulators cannot fully capture.
- **Multi-party Cooperation:** Real-world triage routing depends on the active participation of patients (compliance with recommendations), care site operators (real-time capacity sharing), and EMS providers.
- **Infrastructure Hurdles:** Sharing network state requires secure, low-latency data interoperability across competing care systems, an infrastructural requirement our work assumes rather than solves.

Consequently, we view this work as a foundational research step toward understanding capacity-aware routing, rather than a "plug-and-play" clinical tool.

REPRODUCIBILITY STATEMENT

Upon acceptance, we will release the simulator, the care-network configurations, the case schema, the construction prompts and scripts that let credentialed MIMIC users rebuild the pool, the evaluation

code implementing the metric definitions in Appendix A, and synthetic example cases. Patient-level cases, interviews and labels will not be distributed on unrestricted platforms, following the MIMIC data use agreement. Appendix B details SICL's components, Appendix C the benchmark construction and Appendix C.6 its information partition, and Appendix D the additional results.

REFERENCES

ABIM Foundation. Choosing wisely, 2012. URL https://www.choosingwisely.org/. Initiative to reduce unnecessary medical tests and procedures; launched April 2012.

Eitan Altman. *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

Nilay Tanık Argon and Serhan Ziya. Priority assignment under imperfect information on customer type identities. *Manufacturing & Service Operations Management*, 11(4):674–693, 2009. doi: 10.1287/msom.1080.0246.

Anthony B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263, 1970. ISSN 0022-0531. doi: 10.1016/0022-0531(70)90039-6.

Steven L. Bernstein, Vinu Verghese, Winifred Leung, Anne T. Lunney, and Ivelisse Perez. Development and validation of a new index to measure emergency department crowding. *Academic Emergency Medicine*, 10(9):938–942, 2003. ISSN 1553-2712. doi: 10.1197/S1069-6563(03)00311-7.

Donald M. Berwick, Thomas W. Nolan, and John Whittington. The triple aim: Care, health, and cost. *Health Affairs*, 27(3):759–769, 2008. ISSN 1544-5208. doi: 10.1377/hlthaff.27.3.759.

Geoff Boeing. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 2017. doi: 10.1016/j.compenvurbsys.2017.05.004.

Frances Bunn, Geraldine Byrne, and Sally Kendall. The effects of telephone consultation and triage on healthcare use and patient satisfaction: a systematic review. *British Journal of General Practice*, 55(521):956–961, 2005.

Eileen J. Carter, Stephanie M. Pouch, and Elaine L. Larson. The relationship between emergency department crowding and patient outcomes: A systematic review. *Journal of Nursing Scholarship*, 46(2):106–115, 2014. doi: 10.1111/jnu.12055.

Christine K. Cassel and James A. Guest. Choosing wisely: Helping physicians and patients make smart decisions about their care. *JAMA*, 307(17):1801–1802, 2012. ISSN 0098-7484. doi: 10.1001/jama.2012.476.

Centers for Medicare and Medicaid Services. HCAHPS: Hospital consumer assessment of healthcare providers and systems, 2008a. URL https://www.hcahpsonline.org/. National standardized patient experience survey; first publicly reported 2008.

Centers for Medicare and Medicaid Services. Median time from ED arrival to ED departure for discharged ED patients (OP-18), 2008b. URL https://qualitynet.cms.gov/outpatient/measures/throughput/measures. Hospital Outpatient Quality Reporting (OQR) Program throughput measure.

Centers for Medicare and Medicaid Services. Hospital overall star rating, 2016. URL https://qualitynet.cms.gov/inpatient/public-reporting/overall-ratings. Care Compare overall hospital quality star rating; first published July 2016.

Sarang Deo and Itai Gurvich. Centralized vs. decentralized ambulance diversion: A network perspective. *Management Science*, 57(7), 2011. doi: 10.1287/mnsc.1110.1342.

Zhihao Fan, Jialong Tang, Wei Chen, Siyuan Wang, Zhongyu Wei, Jun Xi, Fei Huang, and Jingren Zhou. AI hospital: Benchmarking large language models in a multi-agent medical interaction simulator. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)*, 2025. URL https://arxiv.org/abs/2402.09742.

Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, 1993. doi: 10.1109/90.251892.

Nicki Gilboy, Paula Tanabe, Debbie Travers, and Alexander M. Rosenau. *Emergency Severity Index Handbook, Fifth Edition*. Emergency Nurses Association, 2023. URL https://media.emscimprovement.center/documents/Emergency_Severity_Index_Handbook.pdf. Emergency Severity Index (ESI) handbook.

Iris Gourdji, Lynne McVey, and Carmen G. Loiselle. Patients' satisfaction and importance ratings of quality in an outpatient oncology center. *Journal of Nursing Care Quality*, 18(1):43–55, 2003. ISSN 1057-3631. doi: 10.1097/00001786-200301000-00007.

Anupam Gupta, Ravishankar Krishnaswamy, and Kirk Pruhs. Online primal-dual for non-linear optimization with applications to speed scaling. In *Approximation and Online Algorithms (WAOA 2012)*, volume 7846 of *Lecture Notes in Computer Science*, pp. 173–186. Springer, 2013. doi: 10.1007/978-3-642-38016-7_15.

Joshua E. Hurwitz, Jo Ann Lee, Kenneth K. Lopiano, Scott A. McKinley, James Keesling, and Joseph A. Tyndall. A flexible simulation platform to quantify and manage emergency department crowding. *BMC Medical Informatics and Decision Making*, 14(50), 2014. doi: 10.1186/1472-6947-14-50.

Oleksandr Ivanov, Lisa Wolf, Deena Brecher, Erica Lewis, Kevin Masek, Kyla Montgomery, Yurii Andrieiev, Moss McLaughlin, Stephen Liu, Robert Dunne, Kevin Klauer, and Christian Reilly. Improving ed emergency severity index acuity assignment using machine learning and clinical natural language processing. *Journal of Emergency Nursing*, 47(2):265–278.e7, 2021. doi: 10.1016/j.jen.2020.11.001.

Yixing Jiang, Kameron C. Black, Gloria Geng, Danny Park, James Zou, Andrew Y. Ng, and Jonathan H. Chen. MedAgentBench: A virtual EHR environment to benchmark medical LLM agents. *NEJM AI*, 2025.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo A. Celi, and Roger G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1), 2023. ISSN 2052-4463. doi: 10.1038/s41597-022-01899-x.

Seth Karten, Wenzhe Li, Zihan Ding, Samuel Kleiner, Yu Bai, and Chi Jin. LLM Economist: Large population models and mechanism design in multi-agent generative simulacra. *arXiv preprint arXiv:2507.15815*, 2025.

Marek Laskowski and Shamir Mukhi. Agent-based simulation of emergency departments with patient diversion. In *Electronic Healthcare (eHealth 2008)*, volume 1 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 25–37. Springer, 2009. doi: 10.1007/978-3-642-00413-1_4.

Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. What makes good in-context examples for GPT-3? In *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pp. 100–114, 2022.

Meng Lu, Brandon Ho, Dennis Ren, and Xuan Wang. TriageAgent: Towards better multi-agents collaborations for large language model-based clinical triage. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 5747–5764. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-emnlp.329. URL https://aclanthology.org/2024.findings-emnlp.329/.

Mehrdad Mahdavi, Rong Jin, and Tianbao Yang. Trading regret for efficiency: Online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13(81):2503–2528, 2012. URL https://www.jmlr.org/papers/v13/mahdavi12a.html.

Jacob Makar-Limanov, Arjun Prakash, Denizalp Goktas, Nora Ayanian, and Amy Greenwald. STA-RLHF: Stackelberg aligned reinforcement learning with human feedback. In *RLC Workshop on Cooperative, Competitive, and Multi-Agent Reinforcement Learning (CoCoMARL)*, 2024.

Robert C. Maves, James Downar, John R. Dichter, John L. Hick, Asha Devereaux, James A. Geiling, Niranjan Kissoon, Nathaniel Hupert, Alexander S. Niven, Mary A. King, et al. Triage of scarce critical care resources in covid-19: An implementation guide for regional allocation: An expert panel report of the task force for mass critical care and the american college of chest physicians. *CHEST*, 158(1):212–225, 2020. doi: 10.1016/j.chest.2020.03.063.

Linda F. McCaig and Eric W. Nawar. National hospital ambulatory medical care survey: 2003 emergency department summary. *Advance Data from Vital and Health Statistics*, (358), 2005.

Jamie Miles, Janette Turner, Richard Jacques, Julia Williams, and Suzanne Mason. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. *Diagnostic and Prognostic Research*, 4(16), 2020. doi: 10.1186/s41512-020-00084-1.

Claire Morley, Maria Unwin, Gregory M. Peterson, Jim Stankovich, and Leigh Kinsman. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLOS ONE*, 13 (8):e0203316, 2018. doi: 10.1371/journal.pone.0203316.

Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine. Data-efficient hierarchical reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/e6384711491713d29bc63fc5eeb5ba4f-Abstract.html.

NHS England. Quality and outcomes framework (QOF), 2004. URL https://www.england.nhs.uk/gp/investment/gp-contract/qof/. General Medical Services contract pay-for-performance scheme; introduced 2004.

Jin Han Malcolm Ong, Bernard J. W. Lim, Muhammad Ariffin B. M. Zahrin, Isaac J. S. Yong, Luke L. L. Tan, Ren Hao Desmond Mao, Marcus E. H. Ong, Fahad Javaid Siddiqui, et al. Ambulance diversion and its use as an ed overcrowding mitigation strategy: Does it work? a scoping review. *International Journal of Emergency Medicine*, 18(125), 2025. doi: 10.1186/s12245-025-00933-0.

OpenAI. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*, 2025.

Jiangweizhi Peng, Yuanxin Liu, Ruida Zhou, Charles Fleming, Zhaoran Wang, Alfredo Garcia, and Mingyi Hong. Hiper: Hierarchical reinforcement learning with explicit credit assignment for large language model agents. *arXiv preprint arXiv:2602.16165*, 2026. URL https://arxiv.org/abs/2602.16165.

Bruno Matos Porto. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emergency Medicine*, 24(219), 2024. doi: 10.1186/s12873-024-01135-2.

Maria C. Raven, Robert A. Lowe, Judith Maselli, and Renee Y. Hsia. Comparison of presenting complaint vs discharge diagnosis for identifying "nonemergency" emergency department visits. *JAMA*, 309(11):1145–1153, 2013.

Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.

Scott M. Sasser, Richard C. Hunt, Mark Faul, David Sugerman, William S. Pearson, Theresa Dulski, Marlena M. Wald, Gregory J. Jurkovich, Craig D. Newgard, E. Brooke Lerner, Arthur Cooper, Stewart C. Wang, Mark C. Henry, Jeffrey P. Salomone, and Robert L. Galli. Guidelines for field triage of injured patients: Recommendations of the national expert panel on field triage, 2011. *Morbidity and Mortality Weekly Report (MMWR) Recommendations and Reports*, 61(RR-1):1–20, 2012.

Samuel Schmidgall, Rojin Ziaei, Carl Harris, Eduardo Reis, Jeffrey Jopling, and Michael Moor. AgentClinic: A multimodal agent benchmark to evaluate AI in simulated clinical environments. *arXiv preprint arXiv:2405.07960*, 2024.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347.

Jori Selen, Ivo Adan, Stella Kapodistria, and Johan van Leeuwaarden. Steady-state analysis of shortest expected delay routing. *Queueing Systems*, 84(3–4):309–354, 2016. doi: 10.1007/s11134-016-9497-7.

Hannah L. Semigran, Jeffrey A. Linder, Courtney Gidengil, and Ateev Mehrotra. Evaluation of symptom checkers for self diagnosis and triage: Audit study. *BMJ*, 351:h3480, 2015. doi: 10.1136/bmj.h3480.

Olanrewaju A. Soremekun, James K. Takayesu, and Stephen J. Bohan. Framework for analyzing wait times and other factors that impact patient satisfaction in the emergency department. *The Journal of Emergency Medicine*, 41(6):686–692, 2011. ISSN 0736-4679. doi: 10.1016/j.jemermed.2011.01.018.

Zhankun Sun, Nilay Tanık Argon, and Serhan Ziya. Patient triage and prioritization under austere conditions. *Management Science*, 64(10):4471–4489, 2018. doi: 10.1287/mnsc.2017.2855.

Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999. doi: 10.1016/S0004-3702(99)00052-1.

Paula Tanabe, Rick Gimbel, Paul R. Yarnold, and James G. Adams. The emergency severity index (version 3) 5-level triage system scores predict ed resource consumption. *Journal of Emergency Nursing*, 30(1):22–29, 2004.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. MedAgents: Large language models as collaborators for zero-shot medical reasoning. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 599–621. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.33. URL https://aclanthology.org/2024.findings-acl.33/.

Chen Tessler, Daniel J. Mankowitz, and Shie Mannor. Reward constrained policy optimization. In *International Conference on Learning Representations*, 2019.

U.S. Congress. Emergency medical treatment and active labor act (EMTALA). 42 U.S.C. §1395dd, 1986. Public Law 99-272.

Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. FeUdal networks for hierarchical reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3540–3549. PMLR, 2017. URL https://proceedings.mlr.press/v70/vezhnevets17a.html.

Zixiang Wang, Yinghao Zhu, Huiya Zhao, Xiaochen Zheng, Dehao Sui, Tianlong Wang, Wen Tang, Yasha Wang, Ewen Harrison, Chengwei Pan, Junyi Gao, and Liantao Ma. ColaCare: Enhancing electronic health record modeling through large language model-driven multi-agent collaboration. In *Proceedings of the ACM Web Conference (WWW 2025)*, 2025. doi: 10.1145/3696410.3714877. URL https://arxiv.org/abs/2410.02551.

Richard R. Weber. On the optimal assignment of customers to parallel servers. *Journal of Applied Probability*, 15(2):406–413, 1978. doi: 10.2307/3213411.

Robin M. Weinick, Rachel M. Burns, and Ateev Mehrotra. Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Affairs*, 29(9):1630–1636, 2010.

Steven J. Weiss, Robert Derlet, Jeanine Arndahl, Amy A. Ernst, John Richards, Madonna Fernández-Frankelton, Robert Schwab, Thomas O. Stair, Peter Vicellio, David Levy, Mark Brautigan, Ashira Johnson, and Todd G. Nick. Estimating the degree of emergency department overcrowding in academic medical centers: Results of the national ED overcrowding study (NEDOCS). *Academic Emergency Medicine*, 11(1):38–50, 2004. ISSN 1553-2712. doi: 10.1111/j.1553-2712.2004.tb01369.x.

Shari J. Welch, Brent R. Asplin, Suzanne Stone-Griffith, Steven J. Davidson, James Augustine, and Jeremiah Schuur. Emergency department operational metrics, measures and definitions: Results of the second performance measures and benchmarking summit. *Annals of Emergency Medicine*, 58 (1):33–40, 2011. ISSN 0196-0644. doi: 10.1016/j.annemergmed.2010.08.040.

Wayne Winston. Optimality of the shortest line discipline. *Journal of Applied Probability*, 14(1): 181–189, 1977. doi: 10.2307/3213271.

Feng Xie, Jun Zhou, Jin Wee Lee, Mingrui Tan, Siqi Li, Logasan S/O Rajnthern, Marcel Lucas Chee, Bibhas Chakraborty, An-Kwok Ian Wong, Alon Dagan, Marcus Eng Hock Ong, Fei Gao, and Nan Liu. Benchmarking emergency department prediction models with machine learning and public electronic health records. *Scientific Data*, 9:658, 2022.

Kuang Xu and Carri W. Chan. Using future information to reduce waiting times in the emergency department via diversion. *Manufacturing & Service Operations Management*, 18(3):314–331, 2016.

An Yang et al. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2024.

An Yang et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Tong Zhang, Yang Wu, Yufei Shi, Rujing Yao, Zhuoren Jiang, and Xiaozhong Liu. From individual excellence to collective sustainability: Seeking strategic equilibrium in proactive multi-agent teams. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 38552–38565, 2026.

Liyuan Zheng, Tanner Fiez, Zane Alumbaugh, Benjamin Chasnov, and Lillian J. Ratliff. Stackelberg actor-critic: Game-theoretic reinforcement learning algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 9217–9224, 2022a. doi: 10.1609/aaai.v36i8.20908.

Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C. Parkes, and Richard Socher. The AI economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8(18):eabk2607, 2022b.

Yifei Zhou, Andrea Zanette, Jiayi Pan, Sergey Levine, and Aviral Kumar. ArCHer: Training language model agents via hierarchical multi-turn RL. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62178–62209. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhou24t.html.

A Metric Implementation Details

This appendix provides full mathematical specifications, parameter values, and literature anchors for the four evaluation metrics defined in Sections 3 and 5. All metrics are post-processed from raw simulator outputs and are therefore recomputable from saved evaluation traces.

ESI convention. The Emergency Severity Index (ESI) is a five-level triage scale used to stratify patients based on acuity and resource needs. We follow the clinical convention where level 1 denotes the highest urgency (e.g., immediate life-saving intervention) and level 5 denotes the lowest. In our evaluation, we group ESI 1–2 as *urgent* (primary safety focus) and ESI 3–5 as *lower-acuity*, consistent with typical emergency department prioritization logic (Gilboy et al., 2023).

Timely-access threshold. We set $T_{\mathrm{SLA}} = 30$ minutes as the main simulator-level timely-access threshold for urgent patients. This value is a benchmark parameter used to compare routing policies under the same dynamic ED environment, not a clinical guideline or an ESI-prescribed time-to-provider standard. The ESI Handbook defines ESI as a five-level acuity and resource-stratification scale, but explicitly distinguishes ESI from triage systems that prescribe expected time intervals to physician evaluation (Gilboy et al., 2023).

To provide operational context, the 30-minute cutoff lies within broader ED triage-immediacy reporting conventions: Healthy People/NHAMCS defines "urgent" triage immediacy as a 15–60

minute recommended evaluation window (McCaig & Nawar, 2005). We therefore use 30 minutes as a mid-window benchmark cutoff, while evaluating all methods under the same threshold and reporting sensitivity to alternative cutoffs in Appendix D.4.

A.1 Urgent Timely Access (UTA)

For the urgent patient set $\mathcal{I}_{\text{urg}} = \{i : z_i^{\text{ESI}} \leq 2\}$ and post-arrival waiting time W_i (in minutes, measured from physical arrival at the destination ED to bed assignment), we compute

$$\text{UTA} = \frac{|\{i \in \mathcal{I}_{\text{urg}} : W_i \leq T_{\text{SLA}}\}|}{|\mathcal{I}_{\text{urg}}|},$$

with $T_{\text{SLA}} = 30$ minutes. The simulator's ground-truth ESI label is used to prevent reward gaming via over-triage.

A.2 Patient Utility A: Rational-Patient Payoff

Per-patient utility combines care-site tier preference, wait penalty, commute penalty, and a fixed go-bonus:

$$\begin{aligned} u_i = b_{\text{go}} + \beta_{\text{rep}}\,\text{tier}(h_i) \\ - \beta_{\text{wait}}\,\phi_W(W_i) - \beta_{\text{com}}\,\phi_C(C_i), \end{aligned}$$

where $\phi_W(W) = \min(W/W_{\max}, 1)$ and $\phi_C(C) = \min(C/C_{\max}, 1)$ are saturating clip functions on post-arrival wait and pre-arrival commute, respectively. Default coefficients are $(\beta_{\text{rep}}, \beta_{\text{wait}}, \beta_{\text{com}}, b_{\text{go}}) = (1, 4, 1, 5)$, with $W_{\max} = 3000$ minutes and $C_{\max} = 45$ minutes; the aggregate $A = \bar{u}/u_{\max} \in [0, 1]$ uses $u_{\max} = 8$. The saturating clip on waiting time reflects that patient satisfaction reaches a floor beyond a wait threshold (Soremekun et al., 2011). We model each patient as a rational decision-maker whose payoff for being routed to site h_i is u_i. This is an assumed preference model rather than a measured patient-reported outcome, although its components mirror dimensions of patient-experience surveys (Centers for Medicare and Medicaid Services, 2008a; Gourdji et al., 2003): under the default profile, patients prefer higher-tier sites, shorter waits and shorter commutes. The coefficients $(\beta_{\text{rep}}, \beta_{\text{wait}}, \beta_{\text{com}}, b_{\text{go}})$ are configurable benchmark parameters, so alternative profiles, such as more time-sensitive patients, can be instantiated without changing the environment. All results for A are reported under the default profile.

A.3 Capacity Efficiency: E

Let $\rho_{h,t} = (q_{h,t} + b_{h,t})/c_{h,t}$ denote instantaneous occupancy of site h at time t, where $q_{h,t}$ is queue length, $b_{h,t}$ is the number of in-bed patients, and $c_{h,t}$ is effective capacity. Let $\mathcal{H}_{\text{high}} \subseteq \mathcal{H}$ denote the top-tier (high-capability) ED subset.

Crowding component.

$$\text{Crowd} = 1 - \frac{\text{OL}_{\text{high}}}{\text{OL}_{\text{high}} + \kappa_{\text{OL}}},$$

$$\text{OL}_{\text{high}} = \frac{\sum_{h \in \mathcal{H}_{\text{high}},\, t} \Delta t \cdot \max(0,\, \rho_{h,t} - 1)}{\sum_{h \in \mathcal{H}_{\text{high}},\, t} \Delta t},$$

with sigmoid knee $\kappa_{\text{OL}} = 30$ chosen so that the observed overload range maps near the inflection point. Δt denotes the interval between consecutive simulator steps, giving time-weighted aggregation that is robust to irregular event timing. This component is motivated by the NEDOCS occupancy ratio and the EDWIN crowding index (Weiss et al., 2004; Bernstein et al., 2003); restricting to $\mathcal{H}_{\text{high}}$ prevents efficiency gaming via dispersion to low-capability sites.

Throughput component.

$$\text{Through} = 1 - \frac{\text{med}_i\, W_i}{\text{med}_i\, W_i + \kappa_{\text{TTB}}},$$

with $\kappa_{\text{TTB}} = 1500$ minutes. The median (rather than mean) is used to handle heavy-tailed ED wait distributions, matching the CMS OP-18 summary statistic (Welch et al., 2011; Centers for Medicare and Medicaid Services, 2008b).

Appropriateness component.

$$\text{Appro} = 1 - \frac{|\{i : z_i^{\text{ESI}} > 2 \wedge a_i \in \mathcal{H}_{\text{high}}\}|}{|\{i : z_i^{\text{ESI}} > 2\}|}.$$

This penalizes assignment of lower-acuity patients (ESI $\in \{3, 4, 5\}$) to scarce high-capability EDs, following the Choosing Wisely principle of reserving high-tier resources for high-acuity need (Cassel & Guest, 2012; ABIM Foundation, 2012).

Aggregate. The three components are equally weighted to mirror the IHI Triple Aim principle of giving equal status to distinct system-quality axes (Berwick et al., 2008):

$$E = \tfrac{1}{3}\big(\text{Crowd} + \text{Through} + \text{Appro}\big).$$

A.4 Safety-Constrained Routing Welfare

The composite welfare is a Cobb–Douglas aggregate:

$$\text{SCRW} = \text{UTA}^{w_S} \cdot E^{w_E} \cdot A^{w_A},$$
$$(w_S, w_E, w_A) = (0.6, 0.2, 0.2).$$

The multiplicative form follows Atkinson's welfare aggregation, which captures non-substitutability between distinct quality axes (Atkinson, 1970): when $\text{UTA} \to 0$, the aggregate $\text{SCRW} \to 0$ regardless of E or A. The choice $w_S = 0.6$ places clinical safety at the lower bound of the 60–66% clinical-domain weighting in CMS Hospital Compare Star Ratings and the NHS Quality and Outcomes Framework (Centers for Medicare and Medicaid Services, 2016; NHS England, 2004), providing a defensible safety-first prior without sacrificing transparency to efficiency and access trade-offs.

A.5 Tail-Risk Indicator

Separately from the four metrics above, we report the 90th percentile of non-urgent patients' wait as a tail-risk statistic (§6.2). It serves as a transparency aid and does not enter the SCRW aggregate.

A.6 Reproducibility

All four metrics are computed by post-processing the `composite_raw` block emitted by our evaluation pipeline. The post-process is deterministic given fixed $W_{\max}$, κ_{OL}, and κ_{TTB} constants, allowing recomputation of any composite from raw fields without re-running the simulator.

B Technical Details of SICL

Postures and priors. The leader chooses among four postures, urgent-first, balance-load, preserve-tertiary and spread-non-urgent. Each fixes a routing prior p_g over the six sites of the care network (two tertiary sites, two community sites, one urgent-care site and one clinic), which depends only on the posture.

Segments. At each arrival the leader emits `KEEP` or `SWITCH`; a posture lasts until the leader switches, so segments have variable length.

Posture incentive. The incentive in Eq. (2) uses $\beta = 0.1$. It is added to the follower's reward during training only, and the follower may choose any site.

System signal. For segment k, let y_k^{urg} be the share of ESI 1–2 patients routed to a tertiary site, y_k^{low} the share of ESI 4–5 patients kept away from tertiary sites, and y_k^{rel} the drop in tertiary queue pressure over the segment, all computed from true acuity. Then

$$Y_k = \tanh\Big(\omega_1\big(y_k^{\text{urg}} - \bar{y}^{\text{urg}}\big) + \omega_2\big(y_k^{\text{low}} - \bar{y}^{\text{low}}\big) + \omega_3\, y_k^{\text{rel}}\Big),$$

where $\bar{y}^{\text{urg}}$ and $\bar{y}^{\text{low}}$ are running means and $\omega_1, \omega_2, \omega_3$ are fixed weights.

Response gate and credit weight. In Eq. (3), $w_{\min} = 0.5$ and $w_{\max} = 1.5$.

Optimization. Both roles are produced by Qwen2-1.5B-Instruct and optimized with the span-level hierarchical PPO of HiPER (Peng et al., 2026), with a KL penalty to a reference policy.

C THE TRIAGEROUTE BENCHMARK

TRIAGEROUTE couples four layers: a case layer that turns de-identified clinical records into pre-arrival presentations, a label layer that carries the clinical ground truth, a demand layer that generates arrival streams, and a care-network layer that defines the sites a patient can be routed to. An event-driven simulator (Section 3) connects them. Three principles guided the construction.

- **Record grounding.** Every patient-specific fact originates in a de-identified MIMIC-IV record (Johnson et al., 2023). Language models are used only to restructure a record, to reconstruct the pre-arrival conversation the record implies, and to map the emergency-department (ED) resource use it documents onto a fixed catalogue. They do not introduce patient-specific facts that are absent from the source record.
- **Pre-arrival realism.** A routing policy observes only what could be known before the patient reaches a site. Results that are produced after arrival, such as laboratory values, imaging findings and clinicians' conclusions, never enter the text a policy reads.
- **Label fidelity.** Evaluation labels come from the record rather than from a model: acuity is the ESI level that the triage nurse assigned during the real encounter, and resource needs are extracted from the ED course documented in the discharge summary, with the supporting evidence attached.

Figure 4 summarizes how the layers are built and how they meet in the simulator.

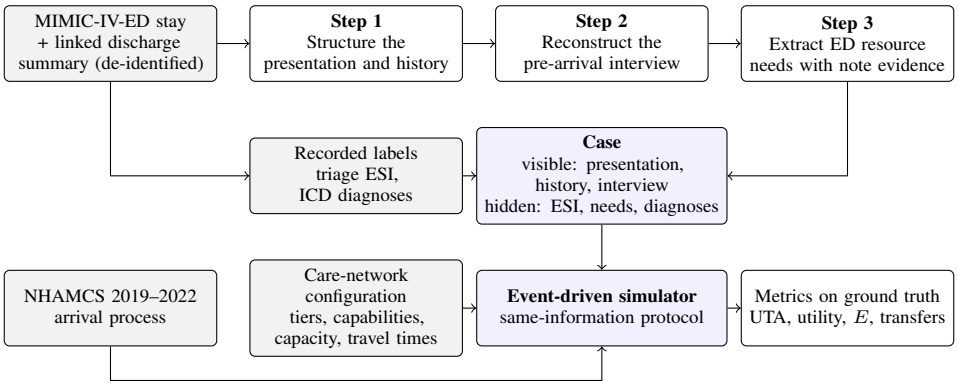

Figure 4: Construction of TRIAGEROUTE. Each case is reconstructed from a real ED encounter in three record-grounded steps; its recorded labels stay hidden from the routing policy and are used only by the simulator and the metrics.

C.1 SOURCE DATA

Table 4 lists every field a case carries, where it comes from, what it is used for, and whether a routing policy can observe it. The clinical sources are MIMIC-IV-ED v2.2 and the MIMIC-IV-Note discharge summaries linked to the same stay (Johnson et al., 2023). Arrival times follow the National Hospital Ambulatory Medical Care Survey (NHAMCS) (McCaig & Nawar, 2005), travel times are shortest driving times on the OpenStreetMap road network (Appendix C.9), and site profiles are compiled from public information about real hospitals and then anonymized. External clinical knowledge (the ESI handbook and the resource catalogue) defines the schema and the label spaces only; it never supplies patient-specific content.

Table 4: Fields of a TRIAGEROUTE case: source, use, and visibility to the routing policy. *Visible only indirectly, through the patient's answers in the pre-arrival interview.

Field	Source	Used for	Visible
Demographics, chief complaint, triage vital signs, pain	MIMIC-IV-ED triage record	Case presentation	✓
Medical background (conditions, medications, allergies, surgical, family and social history)	Discharge summary, restructured (Step 1)	Case presentation; grounding of the simulated patient	✓
Prior ED encounters	Earlier MIMIC-IV-ED stays of the same patient	Case presentation (up to three)	✓
Presenting-illness narrative (onset, character, severity, timing, review of systems)	Discharge summary, restructured (Step 1)	Grounding of the simulated patient	✓*
Pre-arrival interview	Reconstructed (Step 2)	Case presentation	✓
ESI acuity	Triage nurse, MIMIC-IV-ED	Urgent set $\mathcal{I}_{\mathrm{urg}}$; queue priority; stratified reporting	–
Resource needs	ED course in the discharge summary (Step 3)	Feasibility at admission; transfers	–
Diagnoses (ICD-9/10)	MIMIC-IV-ED diagnosis table	Diagnosis groups; scenario sampling	–
Arrival time	NHAMCS-calibrated process	Arrival stream	✓
Origin and travel time to each site	OpenStreetMap drive network (OSMnx)	Commute; transfers; patient payoff	✓
Site tier, capabilities, capacity, hours	Public hospital information, anonymized	Feasibility; network state	✓

C.2 RECORD-GROUNDED CASE RECONSTRUCTION

The source set consists of the $69{,}043$ MIMIC-IV-ED stays that can be linked to a MIMIC-IV-Note discharge summary. Each case is then built in three steps. All three are constrained to content the source record supports, and cases whose extraction fails or is internally inconsistent are removed.

Step 1: presentation and history. A language model restructures the ED record and the discharge summary into a fixed schema. The *presentation* holds the chief complaint, the onset, provocation, quality, radiation, severity and timing of the complaint, positive and negative findings of the review of systems, and the triage vital signs. The *background* holds documented conditions grouped by organ system, medications with dose and frequency, allergies with reactions, surgical, family and social history, and the patient's earlier ED encounters in chronological order.

Step 2: pre-arrival interview. We reconstruct the conversation a patient at home would have with a remote triage assistant before leaving for care. An LLM triage assistant asks one focused question per turn, and a simulated patient answers from the structured case, disclosing information only when asked. Each case carries five independently generated interviews. An interview has 4.9 turns and 128 words on average. Interviews of record-grounded cases were generated with GPT-OSS-120B and Qwen3-32B, and those of supplemented cases (Appendix C.4) with Qwen3.5-27B. The assistant's closing assessment is removed before an interview is shown to a routing policy, so no acuity estimate can be copied from it.

Step 3: resource needs. ED tests and procedures documented in the discharge summary are matched to a fixed catalogue of 29 items in four families: bedside assessments, laboratory tests, imaging, and other diagnostics (Table 6). Each match stores a clinical rationale and a matching rationale that cites the note; in record-grounded cases, 82.2% of matches quote the note verbatim. A test whose ED timing cannot be confirmed is kept in a separate unmatched list (1.6 items per case on average) and does not enter feasibility.

C.3 Clinical Labels

Each case carries three clinical labels. None of them is shown to the routing policy.

ESI acuity. The five-level Emergency Severity Index (ESI) (Gilboy et al., 2023) is the acuity label. It is the level the triage nurse assigned during the real encounter: in all $7{,}966$ pool cases the label equals the triage acuity recorded in MIMIC-IV-ED. ESI 1–2 patients form the urgent set $\mathcal{I}_{\mathrm{urg}}$ of the primary endpoint, and ESI also sets queue priority at the destination site.

Resource need. The resource-need set of a case is the output of Step 3. At admission, the simulator compares it with the capability profile of the assigned site; if the site lacks a needed resource, the patient is transferred to the nearest capable site (Section 3).

Diagnosis group. The ICD-9 or ICD-10 diagnoses of the stay are aligned to ICD-10 chapters and then mapped to seven broad groups with a fixed dictionary: Respiratory & Infection, General & Symptoms, Trauma & Injury, Cardio & Neuro Emergencies, Behavioral Health, Maternal, and Others. Diagnosis groups are used for cohort reporting and scenario sampling; feasibility depends on resource needs only.

C.4 Low-Acuity Supplementation

Linking ED stays to discharge summaries favors patients who were admitted, so low-acuity presentations are nearly absent from the linked set: only 363 of the $69{,}043$ linked stays carry ESI 4 or 5. Because diversion decisions matter most for such patients, we add $1{,}800$ MIMIC-IV-ED stays at ESI 4–5. Their acuity is again the nurse-assigned triage level. Their resource needs come from an LLM-assisted clinical assessment of the ED record, and a case is retained only when two independent clinical-assessment prompts agree on both ESI and resource need. Supplemented cases use a thinner schema than record-grounded cases: they have no structured presenting-illness narrative, 1.3 resource needs on average, and no note-level evidence (Table 5).

C.5 Pool Statistics and Coverage

The resulting pool holds $7{,}966$ cases from $2{,}650$ patients (Table 5). It spans all five ESI levels and a broad diagnostic range: its stays carry $16{,}380$ diagnosis records with $2{,}029$ distinct ICD codes, covering 20 of the 22 ICD-10 chapters and all 19 ICD-9 chapters. Cases are clinically rich: a record-grounded case lists conditions in 5.9 organ systems, 13.5 medications and 6.2 ED resource needs on average, and a typical patient has three earlier ED encounters. These are properties of the case bank; the distribution within an episode is set by the sampling procedure of Appendix C.7.

Figure 5 visualizes the pool along four axes: acuity, the pre-arrival interviews, the ED resources cases require, and diagnostic coverage.

C.6 Information Partition and Leakage Control

For each arrival, the routing policy reads the patient's demographics, chief complaint and triage vital signs; a summary of the medical background, medications, allergies and up to three earlier ED encounters; and the pre-arrival interview with its closing assessment removed. It also receives the live care-network state, the routing objective and the environment's rules (Section 5). The acuity of the current encounter, its resource needs, its diagnoses, the discharge-summary narrative, and all laboratory and imaging results remain hidden; the simulator and the metrics use them, and every policy is scored against them. Scoring urgent timely access on the nurse-assigned ESI rather than on a policy's own estimate prevents a policy from inflating the endpoint by over-triage.

Evaluation arrival streams are drawn from encounters held out from the pool: no evaluation encounter appears in the pool, and record-grounded pool cases share no patient with the evaluation encounters.

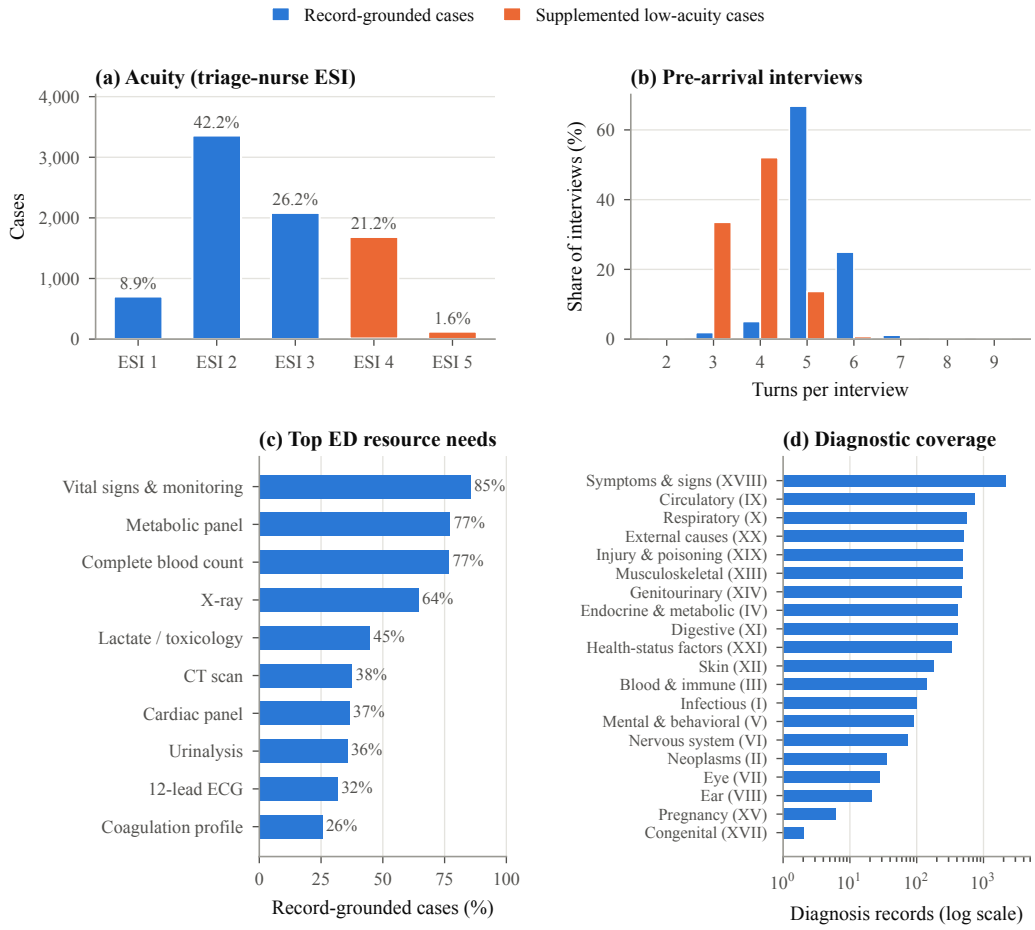

Figure 5: The TRIAGEROUTE patient pool. (a) Cases by nurse-assigned ESI level; shares of the pool are printed above the bars. (b) Number of question–answer turns per pre-arrival interview (five interviews per case). (c) The ten most frequent ED resource needs among record-grounded cases. (d) Diagnosis records by ICD-10 chapter (ICD-10-coded stays; 20 of the 22 chapters are covered, and ICD-9-coded stays cover all 19 ICD-9 chapters).

C.7 ARRIVAL SEQUENCE

Each episode exposes the routing policy to a temporally realistic stream of incoming patients. Formally, an episode is an ordered sequence

$$\mathcal{S} = \{(p_1, t_1), \ldots, (p_N, t_N)\}, \quad t_1 \leq t_2 \leq \cdots \leq t_N,$$

where p_i is a sampled patient and t_i is its arrival time. At each step, the routing policy observes the current patient and care-network state, recommends a care site, and the simulator updates queues and workload before the next patient arrives.

Arrival times are generated from the National Hospital Ambulatory Medical Care Survey (NHAMCS) (McCaig & Nawar, 2005) using 2019–2022 ED records: a weighted Poisson process based on day-of-week and hour-of-day distributions produces hour-level arrivals, with minute-level timestamps sampled Poisson-uniformly within each hour, and a small fraction of patients grouped to represent co-arriving individuals. The resulting curve matches the diurnal pattern of real EDs (peak in late morning to early afternoon, trough overnight) and is stable across months (peak-window coefficient of variation 0.087; mean pairwise Jensen–Shannon divergence 0.009). Figure 6a plots the curve. The same curve serves as the base for the optional scenario settings in Appendix C.8.

Table 5: Pool statistics. Means per case unless stated otherwise; medians in parentheses. †Diagnosis-group shares are computed on the ICD-anchored subset.

	Record-grounded	**Supplemented**	**All**
Cases	6,166	1,800	7,966
Patients	1,255	1,458	2,650
ESI distribution (%)			
ESI 1 (most urgent)	11.5	–	8.9
ESI 2	54.5	–	42.2
ESI 3	33.8	–	26.2
ESI 4	0.1	93.2	21.2
ESI 5 (least urgent)	0.0	6.8	1.6
Diagnostic coverage			
Diagnosis records	12,491	3,889	16,380
Distinct ICD codes	1,327	1,065	2,029
ICD-10 chapters covered (of 22)	20	18	20
ICD-9 chapters covered (of 19)	18	18	19
Patients and presentations			
Age, years	65.9 (66)	54.8 (55)	63.4 (64)
Female (%)	51.1	53.5	51.6
Triage vital signs available (%)	97.4	100.0	98.0
Interview turns	5.2	3.8	4.9
Interview length, words	142	81	128
Clinical content			
Resource needs	6.2	1.3	5.1
Matches quoting the note (%)	82.2	–	–
Organ systems with conditions	5.9	1.0	4.8
Medications	13.5	9.5	12.6
Earlier ED encounters	4.6 (3)	9.6 (3)	5.7 (3)
Diagnosis-group distribution (%)†			
General & Symptoms			67.2
Respiratory & Infection			14.5
Cardio & Neuro Emergencies			13.2
Trauma & Injury			3.8
Behavioral Health			1.1
Maternal			0.1
Others			0.1

C.8 Scenario Sampling

Beyond the baseline arrival process, the benchmark provides a distribution-control mechanism that varies case mix and arrival pressure jointly, enabling evaluation under qualitatively different demand regimes (seasonal-respiratory wave, outbreak surge, mass-casualty incident, MCI). The main experiments use the baseline regime only; the settings below are an affordance for future distribution-shift evaluation.

Four scenario components are defined (Figure 6). Baseline uses the default case mix and the NHAMCS arrival curve. Flu and COVID raise the share of respiratory cases and scale daily volume by 1.3 and 1.2; COVID also shifts acuity toward ESI 1–2. MCI raises the share of trauma and of urgent cases and adds a two-hour surge of arrivals equal to a quarter of the baseline daily volume. These parameterizations qualitatively reflect documented patterns such as respiratory predominance during COVID-19 critical-care surges (Maves et al., 2020) and trauma dominance under field-triage protocols for mass-casualty incidents (Sasser et al., 2012), but they are not empirical estimates of specific events. For reproducible stress testing, five preset scenarios (Normal day, Flu dominant, COVID dominant, MCI event, and Mixed outbreak) fix the component weights instead of sampling them.

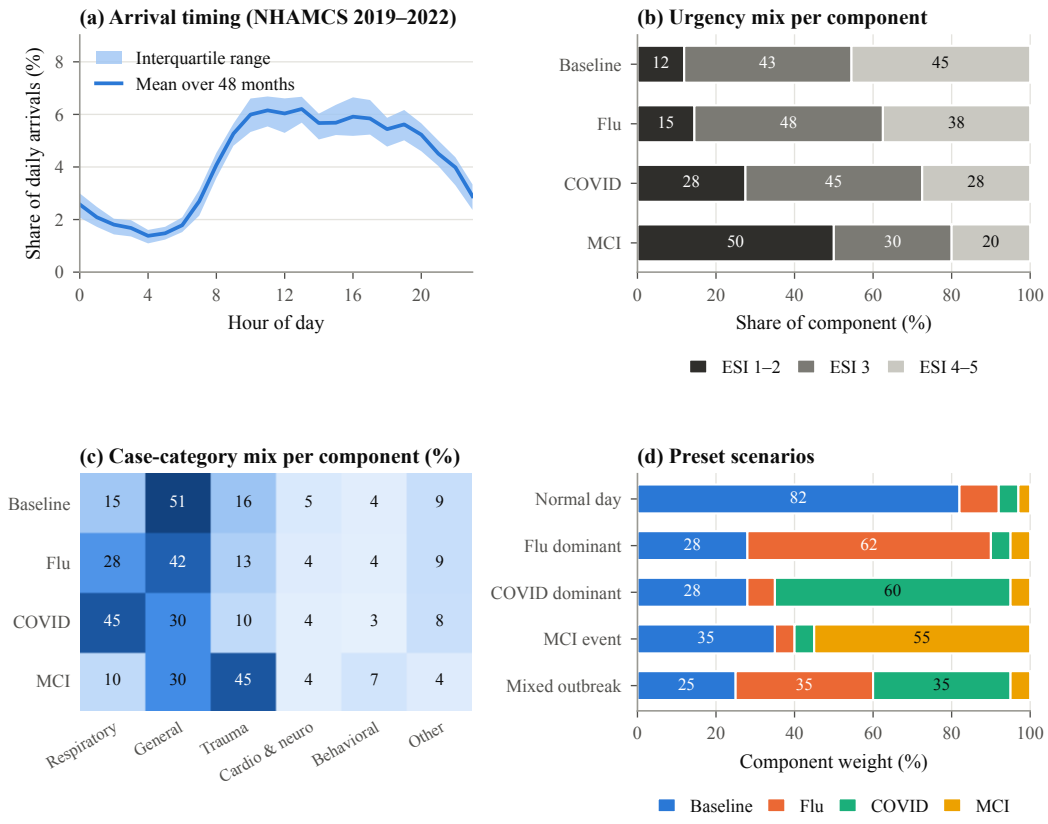

Figure 6: Demand layer. (a) Share of daily ED arrivals by hour of day in NHAMCS 2019–2022: the mean over 48 months and the interquartile range across months. (b) Urgency mix and (c) case-category mix of the four scenario components. (d) Component weights of the five preset scenarios. Flu and COVID additionally scale daily volume, and MCI adds a two-hour surge (see text).

C.9 Care-Network Component

The care-network layer is configurable. A configuration lists the candidate sites of a region, each with a capability tier (from tertiary academic centers through general and community hospitals to urgent-care centers and walk-in clinics), a capability profile over the resource catalogue, an effective capacity, operating hours, and road-network travel times from patient origins. Different configurations instantiate the care networks of different cities and medical-resource levels, so the same patient streams can be simulated under different capacity environments. Site identities are anonymized; only capability and capacity profiles enter the simulator.

Travel times. Each configuration includes the drive network of its region, extracted from OpenStreetMap with OSMnx (Boeing, 2017) and restricted to its largest strongly connected component. Edge travel times start from free-flow speeds and are scaled by fixed congestion multipliers that depend on the road class and on five time-of-day periods (weekday night, daytime and peak; weekend daytime and peak). A patient's commute to a site is the shortest congested driving time from the patient's origin at the time of arrival, computed with Dijkstra's algorithm from every site; the same computation gives transfer times between sites. Patient origins are sampled uniformly over the nodes of the drive network.

Routing feasibility is determined by comparing the resource needs of a case (Appendix C.3) with the capability profile of a site. Table 6 lists the resource categories of the shared catalogue, which both the case labels and the site profiles use.

Table 6: Resource categories of the shared catalogue. Case resource needs and site capability profiles are both expressed over these items; a site can admit a patient only if its profile covers the patient's needs.

Category	Examples
Bedside assessments	Vital signs & cardiac monitoring; point-of-care testing; physical examination; focused neuro/respiratory exams
Laboratory tests	CBC; BMP/CMP; cardiac panel (troponin/BNP); inflammatory markers; lactate; coagulation profile; ABG/VBG; urinalysis; microbiology cultures
Imaging	X-ray; CT; ultrasound (POCUS and formal); MRI; nuclear medicine
Other diagnostic	12-lead ECG; Holter/EEG; angiography; endoscopy; lumbar puncture/paracentesis

C.10 Illustrative Case

Figure 7 shows one case in the benchmark schema. Because MIMIC-derived text may not be redistributed, the case is *synthetic*: it was written for this appendix to illustrate the layers and does not reproduce any patient record.

Synthetic illustrative case (constructed for exposition; not a MIMIC record)

Visible to the policy. *Presentation:* 58-year-old woman; chief complaint "chest pressure"; BP 158/92 mmHg, HR 104 bpm, RR 20/min, SpO_2 96%, temperature 98.4 °F, pain 6/10. *Background:* hypertension, type 2 diabetes, hyperlipidemia; metformin, lisinopril, atorvastatin; penicillin allergy (rash); one earlier ED visit for palpitations eight months ago. *Pre-arrival interview (excerpt):*

Assistant: When did the pressure start, and does it spread anywhere?
Patient: About an hour ago while I was carrying groceries. It goes into my left arm.
Assistant: Are you short of breath, sweaty, or light-headed?
Patient: A little short of breath, and I've been sweating.

Hidden from the policy. *ESI:* 2 (triage nurse). *Resource needs:* 12-lead ECG; cardiac panel; CBC; BMP; chest X-ray. Example evidence for the ECG match: the ED course states that "an ECG was obtained on arrival". *Diagnosis group:* Cardio & Neuro Emergencies.

Consequence in the simulator. Only sites whose capability profile covers all five needs can admit this patient; routing her to an urgent-care center or a walk-in clinic triggers a transfer at admission. Because she is urgent, her wait counts toward urgent timely access (UTA).

Figure 7: A synthetic case in the TriageRoute schema, showing which layers a routing policy observes and which it does not.

C.11 Access, Release, and Intended Use

TriageRoute is derived from credentialed-access MIMIC data and follows the PhysioNet data use agreement. All language-model processing of MIMIC content (Steps 1–3) was run with open-weight models served locally; no record was sent to an external API. Patient-level cases, interviews and labels will not be distributed on unrestricted platforms. Upon acceptance, we will release the simulator, the care-network configurations, the case schema, the construction prompts and scripts that let credentialed MIMIC users rebuild the pool, the evaluation code, and synthetic example cases. The benchmark is intended for research on routing policies in simulation. It is not a clinical decision tool, and none of its policies should be deployed without prospective clinical validation.

D Additional Results

D.1 Classical Routing Rules

The classical references take a structured clinical state and select a destination without any learned component. Nearest-capable sends each patient to the closest site whose capability profile covers their resource needs, which mirrors the closest appropriate facility rule of prehospital field triage

(Sasser et al., 2012). LEAST-LOADED and SHORTEST-DELAY are capability-constrained forms of join-the-shortest-queue and shortest-expected-delay routing (Winston, 1977; Weber, 1978; Selen et al., 2016). GUIDED RED adapts the random early detection diversion policy of Laskowski & Mukhi (2009), itself derived from congestion control (Floyd & Jacobson, 1993), to capacity-normalized load. CCMD, a capability-constrained myopic dispatcher, and two dual-pricing routers in the spirit of online primal-dual methods (Gupta et al., 2013; Mahdavi et al., 2012), one with bounded prices and one without, are constructed for this environment and inherit no competitive-ratio or regret guarantee. Each rule reads the acuity and resource needs recorded by the simulator, which no evaluated policy observes, and is deterministic, so it is run once. This comparison uses a separate evaluation harness; SICL is evaluated in it with the checkpoint of Table 2 over five runs (Table 7).

Table 7: Classical routing rules on recorded acuity and resource needs, and SICL on pre-arrival descriptions, in the same harness.

Policy	**UTA**	**Utility**	E	**SCRW**	**Transfer**
DUAL PRICING (BOUNDED)	0.513	0.665	0.464	0.529	0.0%
SHORTEST-DELAY	0.504	0.643	0.526	0.534	0.2%
CCMD	0.453	0.663	0.359	0.467	0.0%
NEAREST-CAPABLE	0.436	0.632	0.458	0.474	0.0%
LEAST-LOADED	0.385	0.653	0.499	0.450	0.1%
DUAL PRICING	0.385	0.656	0.498	0.451	0.0%
GUIDED RED	0.325	0.655	0.497	0.407	0.1%
SICL (descriptions only)	0.509$\pm$.054	0.677	0.377	0.507	1.1%

D.2 WHERE URGENT PATIENTS MISS THE WINDOW

For every urgent patient (true ESI 1–2) who is not given a bed within 30 minutes, we record the first of the following that applies: the patient was transferred on admission; at the moment of the routing decision, another open site able to meet the patient's true needs had a free bed and was at most 15 minutes farther away (*re-routable*); every site able to meet the patient's needs was full (*saturated*); otherwise, *other*. Table 8 reports the shares over five runs. For all policies except SFT, which never uses one of the two tertiary sites, and Qwen3-32B, most misses are saturated. The last column restricts attention to urgent patients who meet a saturated network; outside saturated moments, SICL and PPO reach the same UTA (0.567).

Table 8: Urgent patients who miss the 30-minute window, by cause (share of misses, five runs pooled), and UTA among urgent patients who meet a saturated network (per-run mean count).

		Share of misses (%)				**UTA when**
Policy	**Misses**	**Transfer**	**Re-routable**	**Saturated**	**Other**	**saturated (count)**
Qwen3-4B	394	0.0	12.7	87.3	0.0	–
Qwen3-14B	479	0.4	21.3	77.2	1.0	–
Qwen3-32B	417	0.2	43.9	54.4	1.4	–
GPT-OSS-120B	312	0.3	10.9	87.8	1.0	0.472 (104.2)
SFT	402	0.2	99.8	0.0	0.0	–
PPO	317	0.0	12.3	87.7	0.0	0.438 (99.0)
w/o incentive & RAC	306	0.0	14.4	85.0	0.7	0.450 (94.6)
SICL	**274**	0.0	16.1	83.6	0.4	**0.524** (96.2)

D.3 CAPACITY-EFFICIENCY COMPONENTS

Capacity efficiency E averages three components (Appendix A): *Crowding* (overload at high-capability sites), *Throughput* (time to bed across all patients) and *Appropriateness* (whether high-capability sites serve the patients who need them). Table 9 reports them. SICL's lower E relative to PPO comes from Crowding and Appropriateness; its Throughput matches PPO's. SICL sends 75% of ESI 3–5 patients to high-capability sites, against 69% for PPO.

Table 9: Components of capacity efficiency E (mean over five runs).

Policy	Crowding	Throughput	Appropriateness	E
Qwen3-4B	0.442	0.533	0.397	0.457
Qwen3-14B	0.432	0.517	0.391	0.446
Qwen3-32B	0.374	0.486	0.260	0.373
GPT-OSS-120B	0.474	0.572	0.463	0.503
SFT	0.400	0.388	0.419	0.402
PPO	0.402	0.509	0.310	0.407
w/o incentive & RAC	0.343	0.475	0.189	0.336
SICL	0.370	0.519	0.247	0.378

D.4 Sensitivity to the Access Threshold

Table 10 reports urgent timely access at 15, 30, 45 and 60 minutes. SICL is highest among these policies at every threshold. Its lead over PPO is significant at 15 minutes ($p = 0.009$) and 30 minutes ($p = 0.018$), and its lead over vanilla GPT-OSS-120B at 30, 45 and 60 minutes ($p = 0.035$, 0.006, 0.044; Welch's t-test over runs).

Table 10: Urgent timely access at alternative thresholds (mean over five runs).

Policy	15 min	30 min	45 min	60 min
Qwen3-4B	0.123	0.327	0.550	0.682
Qwen3-14B	0.048	0.181	0.362	0.545
Qwen3-32B	0.075	0.287	0.482	0.638
GPT-OSS-120B	0.154	0.467	0.723	0.858
SFT	0.070	0.313	0.571	0.744
PPO	0.130	0.458	0.725	0.867
w/o incentive & RAC	0.179	0.477	0.725	0.838
SICL	**0.195**	**0.532**	**0.766**	**0.892**

D.5 Decoding Temperature

Table 11 evaluates SICL and PPO at three decoding temperatures, five runs each. SICL leads at every temperature, and the gap widens as decoding becomes noisier because PPO degrades faster. At $T = 1.0$, SICL produces a valid action for 97.9% of arrivals.

Table 11: Urgent timely access of SICL and PPO at three decoding temperatures (mean $\pm$ SD over five runs; Welch's t-test over runs).

Temperature	PPO	SICL	Difference	p
0.1	$0.458_{\pm.023}$	$0.532_{\pm.045}$	$+0.074$	0.018
0.4	$0.419_{\pm.048}$	$0.532_{\pm.041}$	$+0.113$	0.004
1.0	$0.359_{\pm.039}$	$0.480_{\pm.026}$	$+0.121$	0.0006

E Extended Limitations and Future Work

Our results come from simulation and should be read within its assumptions.

- **Simulation-to-Reality Gap:** Our evaluation relies on a discrete-event simulation environment derived from MIMIC-IV data. While this allows for dynamic workload modeling, real-world emergency department (ED) operations are subject to highly unpredictable disruptions (e.g., sudden staff shortages, equipment failures, ambulance diversions) that are not fully captured by our synthetic queues and state transitions.

- **Assumption of Patient Compliance:** Our simulation assumes that every patient goes to the recommended site (a_i). In reality, patient autonomy, transportation limitations, and insurance network constraints strongly influence care-seeking behavior. Future work must model partial compliance and probabilistic patient routing acceptance.
- **Data Distribution Bias:** The clinical patient profiles and resource need assessments in our benchmark are derived exclusively from the MIMIC-IV dataset, which reflects the demographic and operational characteristics of a single urban academic medical center in the United States. Consequently, the generalizability of the trained policy to rural settings, community hospitals, or non-US healthcare systems remains untested.
- **Reliance on Noisy Pre-Arrival Information:** The patient-level routing response relies on pre-arrival context (e.g., initial chief complaints, self-reported history, early vitals). In clinical practice, this information is often incomplete, noisy, or subjective compared to in-person triage assessments. Although the environment enforces site capability at admission, the risk of under-triage due to missing pre-arrival information is an inherent limitation of any pre-arrival routing system.

These limitations point to the next steps: modeling partial and probabilistic compliance with a recommendation, evaluating on further care networks and patient populations, and validating routing recommendations under clinical supervision before any operational use.