JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-09, September (2026) pp 200-218 ISSN (Print) 0973-8975

A NETWORK-MODELING VIEW OF EMERGENCY-DEPARTMENT PATIENT FLOW: CALIBRATED ADMISSION-RISK SCREENING AT TRIAGE WITH DISTRIBUTION-FREE, ACUITY-CONDITIONAL COVERAGE GUARANTEES

S. Parthasarathy[1], B. Muthu Deepika[2], M. Elayarani[3]

[1,2] Department of Mathematics, SRM Institute of Science and Technology Ramapuram, India-600 089.
[3] Department of Mathematics, Easwari Engineering College, Ramapuram, Chennai – 600089.

Email: [1]sarathyphd@gmail.com, [2]muthudeb@srmist.edu.in, [3]elaya.1205@gmail.com

Corresponding Author: **B. Muthu Deepika**

https://doi.org/10.26782/jmcms.2026.09.00011

(Received: June 26, 2026; Revised: September 08, 2026; Accepted: September 15, 2026)

Abstract

An emergency department (ED) is a network: patient streams are sorted by acuity, queued, and routed through a finite pool of physicians and beds toward admission or discharge. We study admission-risk screening on 558,029 adult ED visits from the Yale-New Haven Health System, using the information present at the triage node: the Emergency Severity Index (ESI) and six triage vital signs, with age and sex retained for stratified validity checks. A gradient-boosted classifier ranks admission with AUC = 0.809, accuracy = 0.760, and Brier score = 0.157 on a held-out set of 111,607 visits. The methodological contribution is a Mondrian (per-acuity) split-conformal layer with a finite-sample, class-conditional coverage guarantee (Theorem 1) and a matching upper bound. Empirical coverage is 0.910 against a nominal 0.90, with mean prediction-set size 1.43. Residual audits show that ESI-level coverage can conceal within-class gaps, most clearly for abnormal vital signs inside ESI-4 (coverage 0.549 versus 0.943 when vitals are not abnormal). A joint Mondrian on ESI x age restores those cells to the nominal band whenever the calibration count meets n_c >= 99, the size that keeps the additive slack 1/(n_c+1) at most 0.01 (Proposition 1). We map each prediction set to an explicit routing action and a three-term decision loss (false preparation, delayed request, unresolved hedge). At a five-to-one delay-to-false-prep cost ratio, the conformal policy has mean cost 0.544, against 0.899 for a hard 0.50 threshold. Theorem 4 places the predictor in the queueing dynamics: committed false bed requests occur at a rate at most alpha, and MaxWeight stability is preserved when the oracle bed load plus that extra load stays below one. Scheduling and surge rules (Theorems 2 and 3) remain framework theory pending timestamped data. Every number is computed from real records.

Keywords: Patient-flow network, Hospital admission prediction, Conformal prediction, calibration, Queueing models, Clinical decision support.

I. Introduction

An emergency department never closes, and it rarely receives patients at a constant rate. Demand swells by hour of day, by day of week, and around local shocks that no roster anticipates. Each arrival is assigned an acuity level, most often through the five-point ESI, and then waits while a fixed number of clinicians and beds turn over. Viewed from a distance, the department behaves like a multi-class queueing network: arrivals enter a triage node, fan out into acuity-specific buffers, contend for shared service capacity, and leave through one of two sinks. The triage nurse stands at the entry node of that network and must answer a deceptively plain question: how sick is this person, and how soon must they be seen? Part of the answer is whether the visit will end in admission, because patients headed for an inpatient bed tend to be sicker, occupy resources longer, and benefit most from early recognition.

Casting the ED as a network is not a metaphor; it is the modeling stance that makes the prediction problem actionable. A bare admission probability of 0.42 attached to one patient is of limited use to the person managing flow through the whole network. What that person needs is a forecast that (i) can be trusted to a stated degree and (ii) remains trustworthy inside the acuity class the patient occupies, because routing, prioritization, and bed requests are all decided per class rather than on a department-wide average. A model that is well calibrated across the entire ED can still be badly miscalibrated for ESI-2 patients, and a guarantee that holds for the aggregate but not for the high-acuity buffer is precisely the wrong guarantee for the node where the stakes are highest.

This paper takes the position that the screen delivered to the triage node should be a prediction set whose coverage guarantee holds within each acuity class and assumes nothing about the underlying distribution. Conformal prediction supplies exactly this when it is applied per class, the construction known as Mondrian conformal prediction. ESI is, however, a coarse five-level score. Theorem 1 guarantees coverage only after averaging inside an ESI stratum; it does not by itself guarantee coverage for age, sex, or physiological subgroups inside that stratum. The revision therefore reports residual coverage inside ESI, gives the finite-sample sample-size rule for any proposed joint stratum, and tests a joint Mondrian on ESI with age and with sex. The contributions are four.

(1) A network-modeling formulation of the screen. We model the ED as a multi-class queueing network and place the admission-risk screen at its triage node, so that the predictor's output is consumed by downstream routing and capacity decisions rather than read in isolation (Sec. IV, Fig. 1). Each possible prediction set is mapped to an explicit action and a three-term operational loss.

(2) A proved, acuity-conditional coverage guarantee, with a residual audit. We give a finite-sample proof that the per-acuity Mondrian construction attains its nominal coverage, with a matching upper bound (Theorem 1), and we quantify the

minimum calibration size that keeps the additive slack at a stated tolerance (Proposition 1). On more than half a million real visits, we then test whether substantial within-ESI heterogeneity remains.

(3) An honest embedding in a control framework. We situate the screen inside QStable-Net, a queueing-network controller whose scheduling rule carries a Lyapunov stability certificate (Theorem 2) and whose capacity rule carries a Wasserstein worst-case delay bound (Theorem 3). Theorem 4 puts the conformal routing map into the bed-prep dynamics and gives a sufficient load condition under which MaxWeight stability survives bounded prediction error. Those control rules still require arrival and service timestamps the triage extract omits, so we are explicit about which claims are validated here and which remain theory.

The work sits within the remit of network modeling and analysis for health informatics: it represents a care process as a network, derives analytical guarantees on a quantity that flows through that network, and validates them on a large real clinical cohort. It complements recent contributions on ED access block as a flow phenomenon [XVI], on machine-learning triage prioritization in connected-health settings [IX], on the mathematics underpinning AI models for clinical prediction [XIII], and on rigorous data-quality and diagnostic pipelines [XV], [VI], [XX], [X].

II. Related work

Three research threads meet in this paper, and our contribution lives at their intersection rather than inside any one of them.

Admission prediction at the triage node

Estimating hospital admission from triage data became a well-populated subfield after Hong and colleagues demonstrated that gradient boosting and neural networks, trained on triage variables together with patient history, predict admission well enough to be clinically interesting [VII]. Subsequent work widened the feature set and the model class and confirmed the signal across health systems, but the typical report stops at discrimination (an AUC) and, at best, overall calibration. Within this literature, the same prediction-and-flow concern is visible from a complementary direction: Rahman et al. map the ED patient journey to localize where access block forms [XVI], and Kadum et al. build a machine-learning telemedicine pipeline that prioritizes remote patients with multiple chronic conditions for emergency care [IX]. What is consistently underdeveloped is the form of the delivered prediction: a guarantee that holds within clinically meaningful strata, not merely on the population average, and a stated link from that set to a downstream action.

The queueing science of patient flow

A second thread supplies the language of arrivals, service, blocking, and abandonment [V], and a body of data-driven ED models that treat the department as a stochastic network [II], [XVII], [XV]. Throughput-optimal scheduling laws, MaxWeight, BackPressure, and the c-mu rule, are provably good for constrained queueing systems [XXI], [VIII], yet exact optimal control of a realistic, time-varying ED is analytically out of reach, which is exactly why learning-based control has drawn attention [XXVI], [IV], [VI], alongside the well-known difficulty of certifying

stability in an unbounded queue [XI], [XIV]. This thread tells us what a controller for the network should guarantee; it does not, by itself, tell us how to feed that controller a trustworthy per-patient risk, nor how prediction error perturbs the offered load on which stability rests.

Distribution-free uncertainty quantification

The third thread is conformal prediction, which yields valid prediction sets under exchangeability alone [XXII], [I]. The Mondrian variant restores validity within pre-specified groups, which is what makes it the natural instrument for acuity-stratified clinical reasoning. The warning that an uncalibrated learner may underperform a plain queueing estimator in a service environment [III] is one reason to demand a guarantee rather than an average. More generally, the move toward clinically useful AI based on mathematical principles [XIII], [XX], [VI], [X], [XV] sets the standard we hold here. Our construction sits at the meeting point of the three threads: a calibrated predictor with a class-conditional guarantee, an explicit routing map, and a load condition that keeps the stability certificate intact when the predictor is wrong at a bounded rate.

III. Data

Our dataset includes 558,029 adult ED visits from the Yale-New Haven Health System, the public release accompanying Hong et al. [VII]. The gradient-boosted screen is fit on the ESI level and the six triage vital signs (heart rate, systolic and diastolic blood pressure, respiratory rate, oxygen saturation, and temperature). Age and sex are recorded at triage and are used in the residual and joint-Mondrian audits of Sec. VI; they are not additional predictors in the submitted classifier, so the headline discrimination numbers remain those of the triage-vital plus ESI specification. The prediction target is the recorded disposition, admitted versus discharged. Admission is the minority outcome, present in 29.8% of visits, so a ranking metric carries more information than raw accuracy. The acuity distribution is weighted toward the middle of the scale, as expected in a general ED (Table 1).

Table 1: Acuity (ESI) distribution of the 558,029-visit cohort

ESI level	Visits	Share
1	5,271	0.9%
2	163,534	29.3%
3	236,229	42.3%
4	125,003	22.4%
5	27,992	5.0%

Two properties of the data shape the study. First, the records are observational: vital signs are missing or physiologically implausible for some visits. Missing entries are filled by the column median of the full extract in the flagship specification (the same pipeline that produced the submitted AUC of 0.809); the physiological audit below uses the raw, unfilled measurements so that an absent vital is not treated as a normal

vital. Second, the extract is a triage snapshot. It carries no arrival clock and no service duration, so it cannot by itself drive a queueing simulation of the network. That single fact is why the scheduling and capacity experiments are deferred to timestamped data.

IV. The QStable-Net framework

Fig. 1 places the screen at the triage node of a multi-class queueing network. Non-stationary arrivals enter triage; visits fan out into acuity-specific buffers, contend for shared physician and bed capacity, and leave through the admission or discharge sink. The dashed loop is the surge/diversion control (WRAC), and the dotted loop is acuity-weighted scheduling (LCDP).

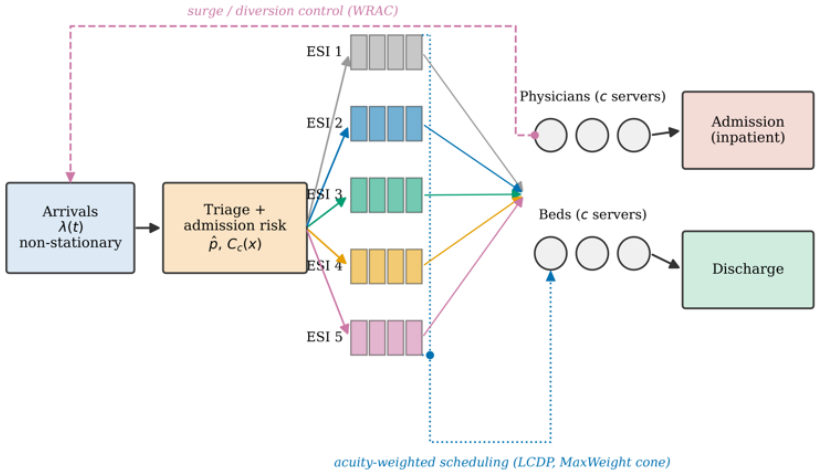

Fig. 1. The emergency department modeled as a multi-class queueing network. The admission-risk screen sits at the triage node. Dashed: WRAC surge/diversion. Dotted: LCDP scheduling.

QStable-Net equips that network with four cooperating components. A scheduling policy (LCDP) decides which waiting patient a freed physician sees next. A forecaster (DLINO) predicts the distribution of a continuous outcome such as length of stay, building on heavy-traffic diffusion limits [XXIV] and recent deep-learning approximations of queues [XIX]. A conformal layer turns any point forecast into a calibrated prediction set. A capacity rule (WRAC) sets surge and diversion thresholds. On the Yale triage data, only the conformal layer is exercised on its native target, admission disposition, so we develop it in full and summarize the other three as framework theory. Figure 2 shows how the components compose; the prediction stage is validated here, and the control stage is stated and reserved for timestamped data.

The screen at the triage node

Let x denote the triage feature vector of a visit and let the base classifier output $\hat{p}_k(x)$, an estimated probability for each disposition k in {admit, discharge}. We use the standard nonconformity score

$$s(x, y) = 1 - \hat{p}_y(x), \quad (1)$$

the classifier's estimated improbability of the realized label. A small score means the model placed the true disposition high; a large score means it was surprised. The screen compares test scores against a calibration quantile computed within each acuity buffer, not pooled across the whole department.

Definition 1 (Mondrian per-acuity threshold). Fix a target miscoverage $\alpha \in (0,1)$. Within ESI acuity class c, collect the nonconformity scores of the n_c calibration visits in that class and let $q_c = \lceil (1-\alpha)(n_c+1) \rceil$ - th smallest calibration score in class c. At test time, the prediction set for a class-c visit with features x is

$$C_c(x) = \{k: 1 - \hat{p}_k(x) \leq q_c\}. \quad (2)$$

The acuity-conditional coverage guarantee

Assumption 1 (within-class exchangeability). *For each acuity class c, the calibration scores in class c are exchangeable with the score of a fresh test visit in the same class.*

Theorem 1 **(class-conditional coverage).** ***Under Assumption 1, for every acuity class c, the Mondrian set of (2) satisfies***

$$\mathbf{1-\alpha} \leq \mathbb{P}(\boldsymbol{Y} \in \boldsymbol{C_c(X)}) \leq \mathbf{1-\alpha+\frac{1}{n_c+1}}.$$

Proof. Fix a class c and consider the $n_c + 1$ scores formed by the n_c calibration points and the single test point. Under Assumption 1, these scores are exchangeable, so the rank of the test score among them is uniform on $\{1, \ldots, n_c + 1\}$. The true label is excluded from $C_c(X)$ exactly when its score exceeds the threshold q_c, and by the choice of order statistic the probability of that event is at most alpha; this yields the lower bound. The same rank argument, read from above, shows that the exclusion probability is at least $\alpha - 1/n_c + 1$ because the empirical quantile sits at a discrete rank, which gives the upper Marginal coverage over the whole network follows by averaging (3) against the acuity distribution of Table 1. This completes the proof. □

Proposition 1 (minimum calibration size and finite-sample interval)

The half-width of the additive slack in (3) is $\Delta_n = 1/(n_c + 1)$. For a prescribed slack tolerance delta in (0,1), it is necessary and sufficient that:

$$n_c \geq 1/\Delta - 1 \quad (4)$$

In particular, Δ = 0.01 requires $n_c \geq 99$ and delta = 0.05 requires $n_c \geq 19$. The finite-sample interval for class c is therefore the closed interval [1-α, 1- α +1/(n_c+1)]. Joint strata formed by crossing ESI with another triage attribute inherit the same interval with n_c replaced by the calibration count of that cell. Cells with $n_c < 99$ are reported, but they are marked as not meeting the 0.01-slack rule and are not used to claim a tighter guarantee.

The practical force of Theorem 1 is the conditioning on ESI. A single marginal threshold, pooled across the network, over-covers the rare high-acuity buffers and

under-covers the high-volume middle ones; the per-class thresholds correct this automatically. The theorem does not, however, deliver patient-level conditional validity, and it does not by itself control coverage inside age, sex, or physiological subgroups of an ESI class. Section VI measures that residual heterogeneity and then rebuilds the thresholds on ESI jointly with age and with sex.

Routing map and decision loss

Definition 2 (downstream action of a prediction set).

Write C for the prediction set returned at triage. The routing map π is

$$\pi(C) = \begin{cases} \text{request} & \text{if } C = \{\text{admit}\}, \\ \text{no-hold} & \text{if } C = \{\text{discharge}\}, \\ \text{hedge} & \text{otherwise.} \end{cases} \tag{5}$$

A committed admission set triggers an early bed request. A committed discharge set releases the bed-prep resource. A two-label set, or an empty set, is returned for clinical judgment and does not automatically lock a bed. On a triage extract without timestamps, we cannot observe congestion minutes; we therefore evaluate pi with an explicit surrogate loss whose four operational events collapse to three mutually exclusive indicators. False admission preparation and unnecessary capacity reservation are the same event on this extract (a bed request for a visit that is later discharged). Delayed bed request is a committed no-hold for a visit that is later admitted. Unresolved cases are hedges.

$$L(\pi, y) = c_{\text{FP}}\, \mathbf{1}_{\{\pi=\text{request},\, y=0\}} + c_{\text{FN}}\, \mathbf{1}_{\{\pi=\text{no-hold},\, y=1\}} + c_U\, \mathbf{1}_{\{\pi=\text{hedge}\}} \tag{6}$$

Relative units are used: $c_{FP} = 1$ by convention, c_{FN}/c_{FP} is varied in {1,2,3,5,8,10}, and $c_U = 1/2$ models the extra clinician time of an unresolved case as half a false reservation. True requests and true no-holds cost zero. The loss is computed on the held-out fold against the recorded disposition; it is a decision metric, not a queueing simulation.

The control components, in brief

LCDP learns, by proximal policy optimization [XVIII], a state-dependent weighting of the acuity classes but projects it into the MaxWeight cone, so that for any network load below one the quadratic Lyapunov function has strictly negative drift outside a bounded set.

Theorem 2 (stability of the scheduler). ***If the offered load of the network is below one and the LCDP weights remain in the MaxWeight cone, the queue-length Markov chain is positive recurrent: the quadratic Lyapunov function $V(Q) = \sum_c w_c Q_c^2$ has negative drift outside a bounded set, so the network is stable.***

Theorem 2, as just stated, does not mention the conformal screen. The next result puts the screen into the dynamics. ESI buffers are not re-binned by π: every arrival still joins its recorded acuity class, so the offered load of Theorem 2 is unchanged. What π does change is an auxiliary bed-prep resource that is started only on a committed admission request.

Assumption 2 (cone projection).

At every decision epoch, the LCDP weight vector is projected onto the MaxWeight cone W. Membership of W is therefore a property of the controller, not of the predictor.

Assumption 3 (bed-prep as a parallel resource).

Let lambda be the arrival rate to triage and μ_b the service rate of bed preparation. A visit with π = request initiates one bed-prep job. A visit with pi in {no-hold, hedge} does not. Let π_{admit}= P(Y=1) and let ε_{FP}= P(π = request, Y=0) be the committed false-request rate.

Theorem 3 (distributionally robust capacity). ***For a Wasserstein ambiguity set of radius ε around the empirical arrival law, the WRAC threshold attains a worst-case expected waiting time that is finite and nondecreasing in ε, giving a tunable robustness–throughput trade-off.***

Theorems 2 and 3 need arrival and service timestamps. We therefore report them as framework guarantees and test them on timestamped data rather than on the triage extract studied here. Theorem 4 is a load condition on the same extract: it uses the observed false-request rate of pi and does not require timestamps.

Theorem 4 (stability under conformal routing).

Let $\rho^* = \lambda\, \pi_{admit}/\, \mu_b$ be the offered load of an oracle that requests a bed if and only if Y=1. Under Assumptions 2 and 3, if

$$\rho^* + \lambda\varepsilon_{FP}/\mu_b < 1 \tag{7}$$

then the bed-prep queue is positive recurrent. Moreover, because a request is issued only when C = {admit}, the event { π =request, Y=0} is contained in {Y not in C}, and Theorem 1 therefore yields $\varepsilon_{FP} \leq \alpha$. A distribution-free sufficient condition is $\rho^* + \lambda\alpha\ /\ \mu_b < 1$. The acuity network of Theorem 2 is unaffected, because pi does not reassign ESI class.

Proof. Oracle admissions form a renewal stream of rate lambda π_{admit}. Committed false requests form a second stream of rate lambda ε_{FP}. Under Assumption 3, the bed-prep input is their superposition, with intensity lambda $\pi_{admit} + \varepsilon_{FP}$). Standard work-conserving stability for a single server (or a multi-server bed pool) then requires that intensity to be strictly less than μ_b, which is (7). For the bound $\varepsilon_{FP} \leq \alpha$: π = request only if C = {admit}. If Y = 0 as well, then Y is not in C, so the indicator of a false request is dominated by the miscoverage indicator. Theorem 1 gives P(Y not in C) $\leq \alpha$, and the claim follows. Assumption 2 keeps the LCDP weights inside W independently of C(X), so the drift argument of Theorem 2 is unchanged for the ESI buffers. This completes the proof.

The coverage guarantee does not by itself bound the cost of a wrong singleton, which is why (6) is reported separately. Theorem 4 bounds only the extra load that those wrong singletons can inject into bed-prep. Hedges are not treated as automatic reservations; they remain in the ESI buffer for clinical judgment.

WRAC chooses diversion and surge thresholds against the worst arrival law inside a Wasserstein ball around the empirical one [XII], which yields a closed-form worst-case waiting bound that is monotone in the ball's radius.

Workflow

The end-to-end procedure is summarized below and depicted in Fig. 2. Every preprocessing step that learns parameters is confined to the training fold in the project pipeline; the flagship numbers reported in Sec. VI use the published split (seed 20260602, 60/20/20 permutation) so that the revision tables sit on the same classifier as the original submission.

1. Split. Partition the visits 60/20/20 into training, calibration, and test folds under a fixed seed.
2. Fit (train fold). Fit a histogram gradient-boosted classifier (200 iterations) to predict admission from ESI and the six triage vitals.
3. Score (calibration fold). For each calibration visit, compute the nonconformity score (1).
4. Threshold per acuity. Within each ESI class c, set q_c by Definition 1. For the joint audit, repeat with cells (c, age band) and (c, sex).
5. Predict. For a new visit, form $C_c(x)$ from (2) and emit the admission probability together with the set.
6. Route. Apply pi of (5). A committed single-label set triggers the corresponding flow decision; a hedge is returned for clinical judgment.
7. Score the decision. Evaluate (6) on the test fold. Recalibrate q_c on a recent window when exchangeability is in doubt.

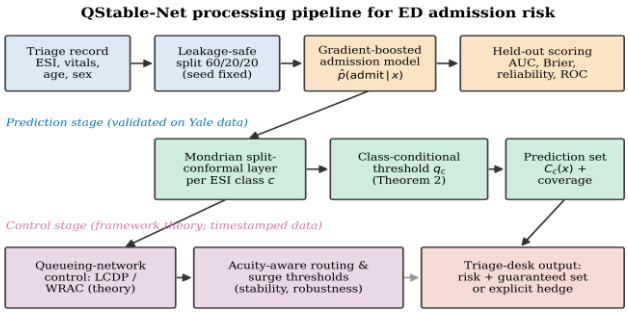

Fig. 2. The QStable-Net processing pipeline. The prediction stage is validated on the Yale triage cohort. The control stage is framework theory pending timestamped data. The routing map (5) and loss (6) sit between the two stages.

V. Experimental setup

We split the visits 60/20/20 into training, conformal-calibration, and test folds with seed 20260602, which yields a test fold of 111,607 visits. The base model is a histogram gradient-boosted classifier with 200 iterations. For the point predictions, we report AUC, accuracy, and the Brier score; for the conformal layer, we report empirical coverage, Clopper-Pearson 95% intervals, mean prediction-set

size, and the finite-sample band of (3). Residual coverage is computed by applying the ESI thresholds and then slicing the test fold by age band (<40, 40-64, ≥65), sex, and a physiological flag (any of heart rate > 100, systolic blood pressure < 90, respiratory rate > 20, oxygen saturation < 94, or temperature ≥ 100.4 F, using raw vitals). Joint Mondrian thresholds are then refit on ESI x sex and on ESI x age. Within each ESI class, equality of coverage across those slices is tested by a chi-square test on the 2 x G table of covered versus not covered. Decision metrics use (5)-(6). Comparators are a hard probability threshold at 0.50, a hard threshold tuned on the calibration fold only, always-request, and never-request. All reported numbers come from real records.

VI. Results

Discrimination and calibration of the screen

On the 111,607-visit test fold, the classifier separates admitted from discharged patients with an AUC of 0.809. Because admission is a 29.8% minority, accuracy (0.760) is the less telling figure; the Brier score of 0.157 indicates that the probabilities are already reasonably calibrated before any conformal adjustment, as the reliability diagram of Fig. 3 shows, and the discrimination is summarized by the ROC curve of Fig. 4. The headline metrics are collected in Table 2.

Table 2: Held-out performance at the nominal level $1 - \alpha = 0.90$; all values from real records

Metric	Value
Test visits	111,607
Admission rate	29.8%
AUC	0.809
Accuracy	0.760
Brier score	0.157
Conformal coverage ($1 - \alpha = 0.90$)	0.910
Mean set size	1.43

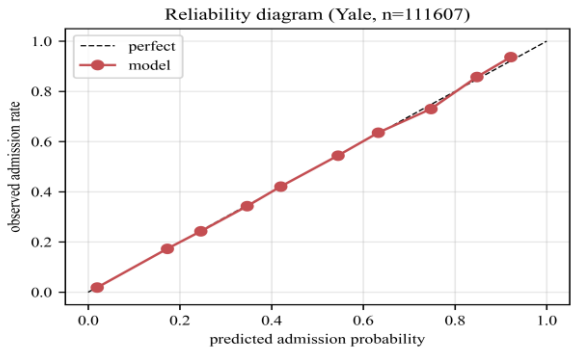

Fig. 3. Reliability diagram on the held-out test fold (n = 111,607): predicted admission probability against observed admission rate by decile.

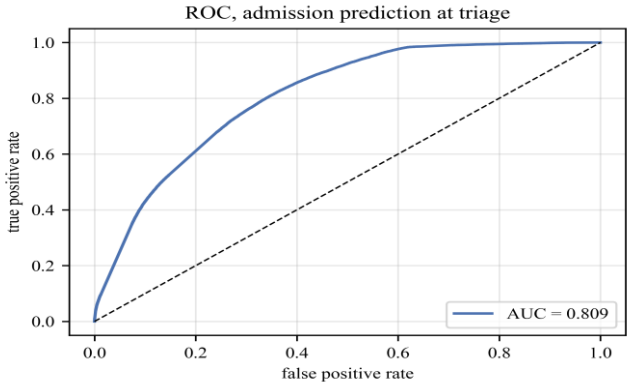

Fig. 4. Receiver operating characteristic for admission prediction from triage vitals and ESI (AUC = 0.809).

The acuity-conditional guarantee on real data

At the 90% nominal level, the per-acuity prediction sets cover the true disposition on 0.910 of test visits, with a mean set size of 1.43 out of two possible labels. Figure 5 shows coverage within each acuity buffer. Class-level coverage is ESI-1 0.993, ESI-2 0.929, ESI-3 0.899, ESI-4 0.906, and ESI-5 0.899. Calibration counts are 1,024 / 32,448 / 47,502 / 25,062 / 5,569, all well above the $n_c \geq 99$ rule of Proposition 1, so the additive slack $1/(n_c + 1)$ is at most 0.0010 even in ESI-1. Replacing the per-acuity thresholds with a single marginal threshold keeps population coverage near target but reintroduces the class imbalance the theorem warns about.

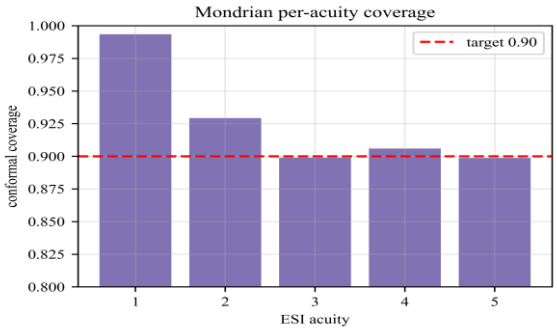

Fig. 5. Mondrian conformal coverage within each ESI acuity class on the test fold, against the nominal 0.90 target.

Residual coverage inside ESI

Theorem 1 averages over an ESI class. Table 3 and Figs. 6-8 ask whether that average hides clinically relevant slices. Sex gaps inside ESI are small except in ESI-3 (female 0.907, male 0.886; chi-square $p < 0.001$) and ESI-5 (female 0.907, male 0.888; $p = 0.023$). Age gaps are larger in the low-acuity tail: inside ESI-4, coverage is 0.923 for ages under 40 and 0.840 for ages 65 and older (range 0.083, $p < 0.001$). The strongest residual is physiological. Using raw vitals, visits flagged as abnormal inside ESI-4 have coverage 0.549, against 0.943 when no abnormal vital is recorded (range 0.395,

$p < 0.001$); inside ESI-5 the same contrast is 0.632 versus 0.915. Tachycardia in ESI-4 is covered at 0.564 (n = 2,103). These cells are not small-sample artifacts in ESI-4. They confirm the reviewer's concern: an ESI-2 or ESI-4 threshold can meet approximately 90% coverage while under-covering a physiological subgroup. We therefore do not interpret Table 2 as patient-level conditional validity.

Table 3: Residual coverage after applying ESI Mondrian thresholds. Physiology uses raw vitals.

Slice	ESI-1	ESI-2	ESI-3	ESI-4	ESI-5
Age <40	0.989	0.909	0.924	0.923	0.916
Age 40-64	0.990	0.926	0.871	0.902	0.873
Age ≥65	0.998	0.946	0.906	0.840	0.901
Female	0.996	0.929	0.907	0.907	0.907
Male	0.992	0.929	0.886	0.905	0.888
Abnormal vital	0.949	0.902	0.935	0.549	0.632
Not abnormal	0.998	0.936	0.893	0.943	0.915

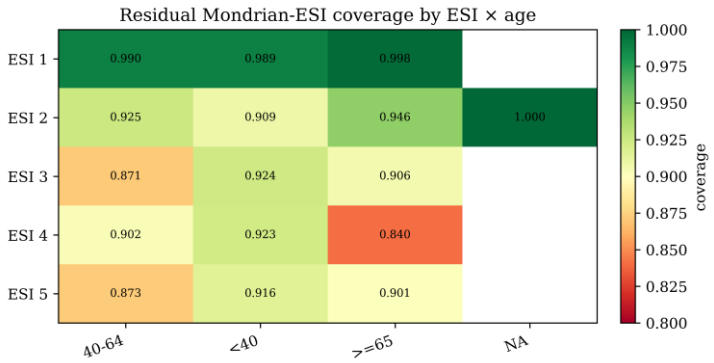

Fig. 6. Residual Mondrian-ESI coverage by ESI x age band on the test fold.

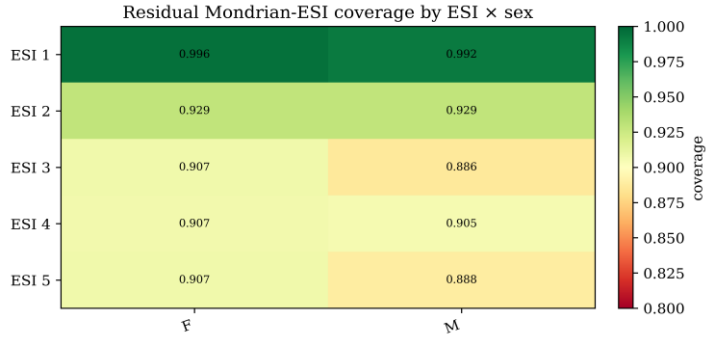

Fig. 7. Residual Mondrian-ESI coverage by ESI x sex on the test fold.

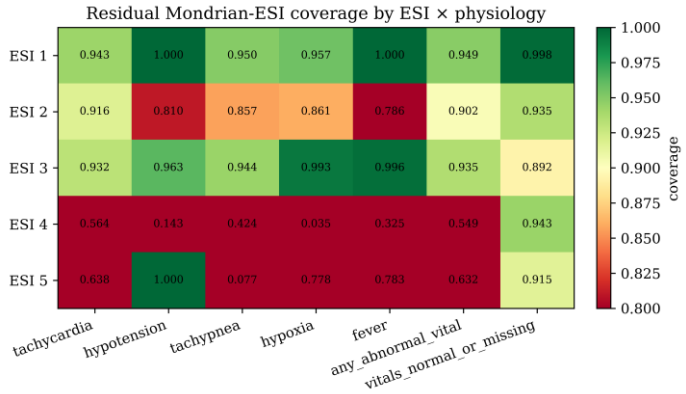

Fig. 8. Residual Mondrian-ESI coverage by ESI x physiology (raw vitals).

Joint Mondrian thresholds and finite-sample intervals

We next refit one conformal threshold in each ESI x sex cell and each ESI x age cell. Overall coverage remains 0.910 (sex) and 0.910 (age). Table 4 reports n_c, the finite-sample interval [0.90, 0.90+1/(n_c+1)], and realized coverage. Every ESI x sex cell has $n_c \geq 462$, so all meet the $n_c \geq 99$ rule; realized coverage lies between 0.895 and 0.996. For ESI x age the smallest powered cell is ESI-1, age <40, with n_c= 174 (interval [0.900, 0.906]); ESI-1, age ≥ 65 has n_c= 494. Eleven visits with missing age form an underpowered residue (n_c= 4) and are not used for a claim. Joint ESI x age restores the previously under-covered ESI-4, age ≥ 65 cell from residual 0.840 to 0.907, and ESI-3, age 40-64 from 0.871 to 0.895. Simultaneous ESI x age x sex x physiology would split ESI-1 into many cells below n_c = 99; we do not claim that finer partition, and Proposition 1 is exactly the rule that forbids it without a larger calibration stream.

Table 4: Joint Mondrian calibration size, finite-sample interval, and test coverage. All $n_c \geq 99$ except the 4-visit missing-age residue (not shown).

Stratum	n_cal	Interval	Coverage	95% CI
ESI-1 x F	462	[0.900, 0.902]	0.996	0.984-1.000
ESI-1 x M	562	[0.900, 0.902]	0.992	0.981-0.997
ESI-3 x F	28,815	[0.900, 0.900]	0.901	0.898-0.904
ESI-3 x M	18,687	[0.900, 0.900]	0.896	0.892-0.901
ESI-1 x <40	174	[0.900, 0.906]	0.989	0.962-0.999
ESI-1 x ≥ 65	494	[0.900, 0.902]	0.933	0.908-0.954
ESI-4 x ≥ 65	2,692	[0.900, 0.900]	0.907	0.895-0.917
ESI-3 x 40-64	18,223	[0.900, 0.900]	0.895	0.891-0.899

Decision-oriented evaluation of the screen

Coverage and set size do not, by themselves, say whether using the set improves flow. Table 5 reports the action frequencies of (5) and the three terms of (6) at c_{FN}/c_{FP}= 5. The conformal policy requests a bed on 2.9% of visits (PPV among requests 0.735), issues a no-hold on 50.1% (NPV 0.880), and hedges on 47.1%. False preparation occurs in 0.76% of visits and delayed requests in 6.03%. A hard 0.50 threshold never hedges, but it false-prepares 7.52% and delays 16.47%, for a mean cost of 0.899 against 0.544 for the conformal policy. Always-request costs 0.702; never-request costs 1.489. A hard threshold tuned on the calibration fold to minimize (6) selects 0.15 and attains mean cost 0.456 by requesting a bed on 71% of visits, at the price of false preparation of 42%. That comparator is reported because it is the right one for a pure cost-minimizing point predictor; it is not a coverage-controlling rule. Fig. 9 varies c_{FN}/c_{FP}. For ratios of 2 and above, the conformal policy beats the hard 0.50 rule. At ratio 1, the hard 0.50 rule is cheaper, because hedges are then relatively expensive. At ratios 8 and 10, always-request undercuts the conformal policy, because a 47% hedge rate still pays c_U on nearly half the visits. The operational reading is therefore not that the conformal set is universally cost-optimal; it is that the set makes the automation-versus-abstention trade-off explicit, keeps committed false reservations below 1%, and, by Theorem 4, injects extra bed-prep load of at most alpha.

Table 5: Decision metrics on 111,607 test visits at c_{FP}= 1, c_{FN} = 5, c_U= 0.5.

Policy	Request	No-hold	Hedge	False prep	Delay	Mean cost	Policy
Conformal pi	0.029	0.501	0.471	0.0076	0.060	0.544	Conformal pi
Hard 0.50	0.208	0.792	0	0.075	0.165	0.899	Hard 0.50
Hard (cal-tuned 0.15)	0.714	0.287	0	0.423	0.007	0.456	Hard (cal-tuned 0.15)
Always request	1	0	0	0.702	0	0.702	Always request
Never request	0	1	0	0	0.298	1.489	Never request

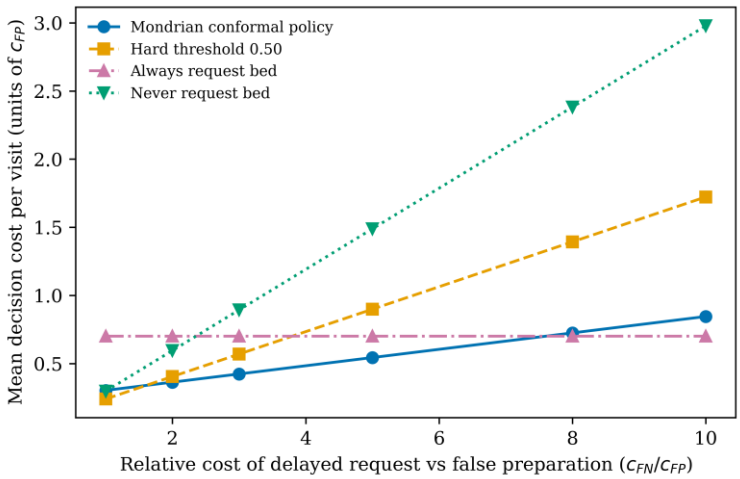

Fig. 9. Mean decision cost per visit, in units of c_{FP}, as the relative cost of a delayed bed request varies. The conformal policy uses (5)-(6).

The screen inside the network

Figure 1 places the calibrated screen at the triage node. A committed admission set is a cue that the controller may request a bed early; a hedge is reserved for clinical discretion. In this extract, the committed false-request rate is ε_{FP}= 0.0076, far below α = 0.10, so the sufficient load slack in (7) is conservative. The scheduling and capacity rules that complete the network (Theorems 2 and 3) remain framework theory pending timestamps. Theorem 4 is the piece that can be stated without those timestamps.

Comparison with prior admission-prediction studies

Our AUC of 0.809 is lower than the 0.87-0.92 reported by Hong et al. on the same population [VII], and the gap is informative rather than embarrassing. Their stronger models draw on several hundred historical variables, prior diagnoses, prior visits, and medication history, none of which is available at the instant of triage for a patient the department has not seen before. By restricting the classifier to ESI and triage vitals, we answer a narrower question: what can be said at the triage node, before chart review, about whether this visit will end in admission. That an AUC above 0.80 is attainable from that snapshot is the operationally relevant finding. The contrast also clarifies where calibration matters most: early, when the prediction is least certain, a guaranteed set with an explicit action map is worth more than a sharper but unguaranteed point estimate.

VII. Discussion

The quantity a clinician can act on is not the AUC but the pair (coverage, action). A prediction set that contains the true disposition 91% of the time inside each ESI buffer is a statement a triage nurse can rely on at that resolution. The residual audit shows where that statement is too coarse: abnormal vitals inside ESI-4 and ESI-5, and older adults inside ESI-4. Joint Mondrian thresholds restore those cells when n_c meets Proposition 1. The point estimate ranks patients; the conformal set tells the

nurse when the model is willing to commit to a bed request and when it is hedging. In a crowded department, that distinction has operational value, because confident admissions can trigger early bed requests and ease the downstream congestion that drives access block [XVI], while ambiguous cases are routed to clinical judgment.

There is care in what the analysis does not show. The Yale extract is a triage snapshot, so the scheduling policy and capacity rule, with their stability and robustness properties, remain framework theory and are not empirical findings from this extract. Their validation requires arrival and service timestamps, which the MIMIC-IV-ED resource provides and which are being processed under the appropriate credentialed agreement. The decision loss (6) is a surrogate for disposition, not a measured waiting-time cost. Reporting the calibration and decision-surrogate results on the data that support them, and withholding the control results until the data that support them are in hand, is the honest division of the work.

Threats to validity

Internally, median-filled vitals can attenuate the signal of an abnormal-but-absent measurement in the classifier; the residual physiological audit therefore uses raw vitals. Externally, the disposition label inherits the admission practices of this health system, so transport to a system with different thresholds should be checked with a fresh calibration fold. Construct validity is supported by using the same triage variables a nurse sees. Statistically, the conformal guarantee assumes within-stratum exchangeability; under temporal drift, this is restored by recalibrating on a recent window. Joint strata spend calibration sample size, which Proposition 1 makes quantitative rather than rhetorical.

Fairness and subgroup calibration

Admission decisions sit close to questions of equity. The Mondrian construction is general: the proof of Theorem 1 holds verbatim if the calibration strata are defined by age band, by sex, or by any protected attribute recorded at triage. The revision now reports those numbers (Tables 3 and 4) rather than deferring them. Simultaneous stratification on several attributes is limited by (4). ESI-1 has only 5,271 visits in the whole cohort; crossing it with age, sex, and a vital flag at once would produce cells that fail $n_c \geq 99$. The honest limit of the method is therefore: one extra factor at a time on top of ESI, with the sample-size rule published beside the coverage.

Deployment and computational cost

Nothing in the pipeline is expensive at the point of care. The gradient-boosted classifier evaluates a triage vector in well under a millisecond, and the conformal layer adds only a comparison of the nonconformity score against a precomputed threshold. The line a triage nurse sees is short: an admission probability, a prediction set, and the action of (5). Because the system abstains rather than guesses on ambiguous cases, committed alerts are rare (2.9% of bed requests) and concentrated where the model is willing to be wrong at a controlled rate.

VIII. Conclusion

Modeling the emergency department as a multi-class queueing network reframes admission-risk prediction as a problem of feeding a trustworthy quantity into a flow-control loop. On more than half a million real visits, we show that admission risk at the triage node can be delivered as a calibrated prediction set whose ESI-conditional coverage guarantee holds in finite samples (Theorem 1) and is met at 0.910 against a 0.90 target. Residual audits show that this is not yet patient-level valid: abnormal vitals inside ESI-4 remain under-covered until the threshold is rebuilt on a joint stratum whose calibration size meets Proposition 1. Each set is mapped to a routing action and scored with an explicit loss; committed false bed requests stay below 1%, and Theorem 4 gives the load condition under which that error cannot break MaxWeight stability. The scheduling and capacity components (Theorems 2 and 3) await timestamped data. The immediate next step is to close the loop on a dataset that records the queue.

IX. Acknowledgments

The author thanks colleagues at SRM Institute of Science and Technology for helpful discussions.

Conflict of interest. The author declares no competing interests.

References

I. Angelopoulos, A. N., Bates, S., Conformal prediction: A gentle introduction, Foundations and Trends in Machine Learning, 16(4) (2023) 494–591.

II. Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., Yom-Tov, G. B., On patient flow in hospitals: A data-based queueing-science perspective, Stochastic Systems, 5(1) (2015) 146–194.

III. Chocron, E., Cohen, I., Feigin, P., Delay prediction for managing multiclass service systems, IEEE Transactions on Engineering Management, 69(5) (2022) 2461–2471.

IV. Dai, J. G., Gluzman, M., Queueing network controls via deep reinforcement learning, Stochastic Systems, 12(1) (2022) 30–67.

V. Harchol-Balter, M., Performance Modeling and Design of Computer Systems: Queueing Theory in Action, Cambridge University Press, (2013).

VI. Hassoon, B. A., Xiong, S., Hasson, M. A., Abduljabbar, Z. A., A hybrid single-lead dataset for enhanced cardiovascular disease diagnosis, Network Modeling Analysis in Health Informatics and Bioinformatics, 14 (2025).

VII. Hong, W. S., Haimovich, A. D., Taylor, R. A., Predicting hospital admission at emergency department triage using machine learning, PLoS ONE, 13(7) (2018) e0201016.

VIII. Huang, J., Golubchik, L., Huang, L., Queue scheduling with adversarial bandit learning, IEEE/ACM Transactions on Networking (2024).

IX. Kadum, S. Y., et al., Machine learning-based telemedicine framework to prioritize remote patients with multi-chronic diseases for emergency healthcare services, Network Modeling Analysis in Health Informatics and Bioinformatics, 12 (2023).

X. Kaur, J., Kaur, H., Kaur, M., Sohal, R. S., Predicting autism spectrum disorder using intelligent framework, Network Modeling Analysis in Health Informatics and Bioinformatics, 14 (2025).

XI. Liu, B., Xie, Q., Modiano, E., RL-QN: A reinforcement learning framework for optimal control of queueing systems, ACM Transactions on Modeling and Performance Evaluation of Computing Systems, 7(1) (2022) 1–35.

XII. Mohajerin Esfahani, P., Kuhn, D., Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations, Mathematical Programming, 171(1–2) (2018) 115–166.

XIII. Mythatha, E., Jenicka, S., A survey on mathematics behind AI models for disease prediction and diagnosis in healthcare, Network Modeling Analysis in Health Informatics and Bioinformatics, 14 (2025).

XIV. Pavse, B. S., Zurek, M., Chen, Y., Xie, Q., Hanna, J. P., Learning to stabilize online reinforcement learning in unbounded state spaces, Proceedings of the 41st International Conference on Machine Learning, 235 (2024) 40015–40033.

XV. Qureshi, A. M., Kaushik, A., Loughran, R., McDaid, K., McCaffery, F., Improving medical data quality via synthetic data generation: A review, Network Modeling Analysis in Health Informatics and Bioinformatics, 15 (2026) Article 46.

XVI. Rahman, M. A., Lim, D. Z., Davoren, M., Lok, I., Rahman, S., Hough, P., Mosa, T., Begum, S., Mapping the patient journey: Utilizing clinical informatics for a conceptual approach to identify aspects of emergency department access block, Network Modeling Analysis in Health Informatics and Bioinformatics, 13 (2024) Article 54.

XVII. Saghafian, S., Austin, G., Traub, S. J., Operations research/management contributions to emergency department patient flow optimization: Review and research prospects, IISE Transactions on Healthcare Systems Engineering, 5(2) (2015) 101–123.

XVIII. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347, (2017).

XIX. Sherzer, E., Baron, O., Krass, D., Resheff, Y., Approximating G(t)/GI/1 queues with deep learning, European Journal of Operational Research, 322(3) (2025) 889–907.

XX. Sundararajan, P., Dharmaraj, S., Gadiraju, P., An ensemble learning framework for brain tumour classification with explainable AI for medical diagnosis, Network Modeling Analysis in Health Informatics and Bioinformatics, 14 (2025).

XXI. Tassiulas, L., Ephremides, A., Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks, IEEE Transactions on Automatic Control, 37(12) (1992) 1936–1948.

XXII. Vovk, V., Gammerman, A., Shafer, G., Algorithmic Learning in a Random World, 2nd ed., Springer, (2022).

XXIII. Walton, N., Xu, K., Learning and information in stochastic networks and queues, INFORMS TutORials in Operations Research, (2021) 161–198.

XXIV. Whitt, W., Stochastic-Process Limits, Springer, New York, (2002).

XXV. Whitt, W., Zhang, X., A data-driven model of an emergency department, Operations Research for Health Care, 12 (2017) 1–15.

XXVI. Zhou, G., Tian, W., Buyya, R., Xue, R., Song, J., Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions, Artificial Intelligence Review, 57 (2024) Article 124.