Decision-Informed Online Conformal Prediction for ICU Resource Allocation

Chandra Sekhar Dronavajjala CHANDRAD@UCONN.EDU
Department of Computer Science and Engineering
University of Connecticut
Storrs, Connecticut, USA

Abstract

Hospitals allocate scarce ICU resources—nurses, beds, discharge slots—using predictions of patient length of stay (LOS), but prediction errors under shifting patient populations lead to costly misallocations. Online conformal prediction provides adaptive uncertainty sets that track a target coverage rate (e.g., 90%) without distributional assumptions, but the resulting robust decisions incur a *Price of Coverage*—the excess cost of hedging against uncertainty. We observe that standard conformal methods distribute uncertainty budgets uniformly, including on resources the optimizer assigns to minimum levels. We propose **Decision-Informed Conformal Adaptation (DICA)**, which uses the downstream optimization's resource allocation as feedback to reshape uncertainty margins: tighter where allocation is minimal (saving cost), wider where allocation is high (strengthening protection). DICA preserves the same adaptive quantile update and scalar coverage tracking as standard online conformal; the reshaped radii do not carry a formal per-component guarantee, but decision-level analysis shows that coverage misses are confined to lower-bound dimensions where their clinical cost is negligible. Across 328K patient stays from three real clinical datasets (MIMIC-IV, eICU, and a general hospital cohort), DICA reduces the Price of Coverage by 43–54% relative to uniform conformal while maintaining approximately 90% coverage. Under chronological patient ordering, static calibration methods degrade to 78–86% coverage, while online conformal methods remain near the target rate.

1. Introduction

Intensive care units (ICUs) must allocate scarce resources—nurses, beds, discharge slots—under uncertainty about patient length of stay (LOS). Machine learning models can predict LOS from electronic health records with moderate accuracy (Rajkomar et al., 2018), and these predictions feed downstream optimization models, specifically linear programs (LPs), that determine staffing schedules and bed assignments in the predict-then-optimize paradigm (Elmachtoub and Grigas, 2022). However, prediction errors can cause costly misallocations: understaffing a critical ward or discharging patients too early.

A key challenge in deploying such pipelines is *distribution shift*. Patient populations change with seasons, disease outbreaks, and evolving clinical protocols. Models trained on historical data degrade systematically when deployed prospectively (Nestor et al., 2019; Futoma et al., 2020), and static uncertainty estimates become miscalibrated. This is not a theoretical concern: Nestor et al. (2019) showed that standard clinical ML models suffer

Code available at https://github.com/chandrad/conformal-ops

significant performance degradation under temporal shift across multiple prediction tasks in MIMIC.

Conformal prediction provides a principled solution: construct prediction intervals that contain the true LOS with a target probability (e.g., 90%), then solve a *robust* optimization that hedges against all costs within these intervals (Johnstone and Cox, 2021; Patel et al., 2024). Unlike Bayesian or point-estimate uncertainty, conformal prediction provides distribution-free coverage properties without parametric assumptions. The Gibbs-Candès online update (Gibbs and Candès, 2021) extends this to the streaming setting, adapting the conformal quantile shift-by-shift so that empirical coverage tracks a target rate asymptotically, without requiring a static calibration set. This online approach is well-suited for hospitals, where the data distribution is inherently non-stationary.

This approach works: online conformal prediction empirically tracks 90% coverage where split conformal methods such as Conformal Predict-then-Optimize (CPO) degrade to 76–87% under random batching and as low as 78% under chronological patient arrival (Patel et al., 2024). But it comes at a price. The robust LP must plan for worst-case costs within the conformal radii, leading to systematically higher resource allocation—what we call the **Price of Coverage** (PoC). This cost parallels the classical "Price of Robustness" in robust optimization (Bertsimas and Sim, 2004). Across our experiments, Uniform Conformal Allocation (UCA)—standard online conformal (Gibbs and Candès, 2021) with uniform radii—incurs a PoC of +2% to +12% depending on the dataset and problem.

We observe that standard conformal prediction allocates uncertainty budgets *uniformly* across all dimensions. In a 20-bed staffing LP, the optimal solution might concentrate allocation on 2–3 critical beds while keeping 17 others at minimum staffing. Yet conformal radii are identical for all 20 beds. The radii on the 17 low-allocation beds inflate the robust cost without contributing meaningfully to the decision. Recent work on decision-aware conformal prediction (Angelopoulos et al., 2024; Lekeufack et al., 2024; Cortes-Gomez et al., 2025) has explored incorporating task structure into conformal methods. DICA takes a complementary approach: rather than modifying the score function or calibration procedure, it uses the optimization's own allocation output to reshape radii post-hoc.

Our contribution: DICA. We propose Decision-Informed Conformal Adaptation, which feeds the LP's allocation back into the conformal layer. After each LP solve, the allocation vector z^* reveals which dimensions are decision-relevant. DICA uses this signal (via exponential moving average) to reshape conformal radii: tighter where allocation is low, standard where allocation is high. The Gibbs-Candès quantile update remains standard, so the scalar coverage tracks the same target rate as uniform conformal. Specifically, our contributions are:

1. **DICA**, a method that uses LP allocation feedback to redistribute conformal radii, reducing the Price of Coverage by 43–54% across 328K patient stays from three real clinical datasets.

2. **An analysis** of when and why DICA's savings arise: in a d-dimensional LP with m binding constraints, DICA saves robustification cost on the $d - m$ non-basic variables, with net savings scaling with the sparsity ratio $(d - m)/d$.

3. **Empirical evidence** that online conformal prediction is viable for hospital resource allocation under temporal non-stationarity, while split conformal (CPO) degrades further under chronological patient ordering (Appendix D).

Generalizable Insights about Machine Learning in the Context of Healthcare

- **The cost of calibrated uncertainty depends on operational structure, not just model quality.** When ML predictions feed downstream resource allocation, the cost premium for maintaining calibrated prediction intervals depends on the optimization problem's constraint structure—specifically, how many resources the optimizer assigns to minimum levels. This generalizes to any ML-for-healthcare pipeline where predictions feed optimization with sparse solutions: staffing, scheduling, inventory, triage.
- **Online conformal prediction handles the non-stationarity inherent in hospital data.** Across 328K patient stays spanning pandemic-era temporal shifts and cross-site heterogeneity (208 hospitals), online conformal methods track 90% coverage where static calibration degrades to 78–86%. Patient populations shift with seasons, disease outbreaks, and evolving protocols; online methods adapt without retraining.
- **Before deploying uncertainty-aware decision support, estimate the cost of safety from the optimization's sensitivity structure.** This avoids engineering investment in settings where coverage is either free (the optimizer is insensitive to prediction errors) or prohibitively expensive (defer to clinical judgment). The assessment requires only a standard sensitivity report from the allocation model.
- **In healthcare, coverage has asymmetric value.** The cost of under-protection (understaffing, patient harm) typically exceeds the cost of over-protection (resource waste). This asymmetry should inform whether to adopt robust decision support, independent of the specific method used.

2. Related Work

Conformal Prediction for Optimization. Johnstone and Cox (2021) first proposed conformal uncertainty sets for robust optimization. Patel et al. (2024) developed CPO using split conformal with periodic recalibration, but coverage degrades under distribution shift. Yeh et al. (2025) learn the score function end-to-end to minimize task loss in batch settings. Cortes-Gomez et al. (2025) design utility-directed nonconformity scores. Binucci et al. (2025) combine online conformal risk control with Lyapunov optimization for networked resource allocation. DICA takes a complementary approach: it preserves the standard score and quantile but reshapes radii post-hoc using optimization feedback.

Online and Decision-Aware Conformal Prediction. Gibbs and Candès (2021) introduced adaptive conformal inference under distribution shift; Zaffran et al. (2022) extended it with expert aggregation, Bhatnagar et al. (2023) improved regret bounds, and Feldman et al. (2023) generalized to risk control. Hu et al. (2026) incorporate the estimated score distribution into the quantile update for tighter sets; their approach modifies *how*

the quantile adapts, while DICA modifies *how radii are distributed*. On the decision-aware side, Angelopoulos et al. (2024) introduced conformal risk control, Lekeufack et al. (2024) developed conformal decision theory, Kiyani et al. (2025) provide decision-theoretic foundations showing reshaped sets are optimal, and Zhou and Zhu (2026) trace miscoverage-regret Pareto frontiers. Locally adaptive methods (Romano et al., 2019; Lei et al., 2018) produce heterogeneous intervals adapted to prediction difficulty. All these modify the score, objective, or quantile update. DICA instead preserves the standard Gibbs-Candès machinery and reshapes only the radii allocation, connecting the Price of Robustness (Bertsimas and Sim, 2004) to online conformal prediction.

Conformal Prediction in Healthcare. Shahbazi et al. (2026) apply hierarchical conformal prediction for LOS prediction across 3,793 hospitals, using Bayesian uncertainty weighting to achieve 94.3% coverage—the closest healthcare application to ours, but operating in a batch (not online) setting without downstream resource optimization. Conformal methods have recently been applied to medication dosing (Diaz-Rincon et al., 2025), cardiac risk stratification (García et al., 2024), treatment triage in breast cancer (García-Galindo et al., 2023), pulmonary nodule detection (Hulsman et al., 2025), and survival screening (Sesia and Svetnik, 2025); all operate in batch settings. Group-conditional conformal prediction (Vovk et al., 2003; Jung et al., 2023) enables coverage guarantees conditional on subpopulations, with natural applications to clinical risk stratification. A key gap in the clinical conformal literature is the absence of *online, closed-loop* systems where conformal sets inform optimization and optimization feedback reshapes the conformal layer—precisely the setting DICA addresses.

3. Methods

3.1. Problem Setup

At each time step $t = 1, \ldots, T$:

1. A batch of d patients arrives with features $\{x_i\}_{i=1}^{d}$.
2. A predictor estimates LOS: $\hat{c}_i = f(x_i)$, yielding predicted cost vector $\hat{c} \in \mathbb{R}^d$.
3. The hospital solves a resource allocation LP:

$$z^* \in \arg\min_{z \in \mathcal{Z}} \hat{c}^\top z \tag{1}$$

   where $\mathcal{Z} = \{z : A_{\text{eq}} z = b_{\text{eq}},\ A_{\text{ub}} z \leq b_{\text{ub}},\ \ell \leq z \leq u\}$ encodes staffing constraints.
4. True costs c are revealed, and the actual cost is $c^\top z^*$.

The **nominal** approach solves (1) directly with $\hat{c}$. This ignores prediction uncertainty and performs poorly when $\hat{c} \neq c$.

3.2. Online Conformal Prediction

Following Gibbs and Candès (2021), we maintain a scalar quantile q_t that adapts online. At each step:

Score: The normalized max-residual score aggregates per-component errors into a single scalar:

$$s_t = \max_{j=1,\ldots,d} \frac{|c_j - \hat{c}_j|}{\hat{\sigma}_j} \tag{2}$$

where $\hat{\sigma}_j$ is an exponential moving average (EMA) of $|c_j - \hat{c}_j|$ over recent rounds.

Radii: The conformal radii are $r_j = q_t \cdot \hat{\sigma}_j$, giving the robust cost vector $\hat{c} + r$.

Coverage: The event $\{s_t \leq q_t\}$ is equivalent to $\{|c_j - \hat{c}_j| \leq r_j,\ \forall j\}$—joint coverage across all dimensions.

Update: The miscoverage rate α_t adapts via:

$$\alpha_{t+1} = \alpha_t + \eta\big(\alpha_{\text{target}} - \mathbf{1}[s_t > q_t]\big) \tag{3}$$

and q_t is the $(1 - \alpha_t)$-quantile of recent scores. This drives empirical marginal coverage toward $1 - \alpha_{\text{target}}$ asymptotically, even under distribution shift.

The **robust LP** replaces $\hat{c}$ with $\hat{c} + r$:

$$z^*_{\text{rob}} = \arg\min_{z \in \mathcal{Z}} (\hat{c} + r)^\top z \tag{4}$$

3.3. Price of Coverage

We define the **Price of Coverage** as the relative cost increase of the robust solution over the nominal:

$$\text{PoC} = \frac{\text{Cost}_{\text{robust}}}{\text{Cost}_{\text{nominal}}} - 1 \tag{5}$$

where costs are evaluated at the true cost vector c. A PoC of +5% means the robust (covered) solution costs 5% more than the uncalibrated nominal solution.

The robust LP's cost can be decomposed as:

$$(\hat{c} + r)^\top z^*_{\text{rob}} = \hat{c}^\top z^*_{\text{rob}} + r^\top z^*_{\text{rob}} \tag{6}$$

The second term, $r^\top z^*_{\text{rob}} = \sum_j r_j \cdot z^*_j$, is the *cost of robustification*. Standard conformal sets $r_j = q\hat{\sigma}_j$ uniformly, paying the robustification cost even on dimensions where z^*_j is at its lower bound.

3.4. DICA: Decision-Informed Conformal Adaptation

Key observation. In an LP with d variables and m equality constraints, the optimal solution has at most m basic variables (allocation beyond lower bounds). For nurse staffing ($d = 20$, budget constraint $\sum_j z_j = 12$, bounds $[0.1, 1]$), the LP solution has approximately 11 variables at the upper bound, 8 at the lower bound $\ell = 0.1$, and 1 fractional (40% sparsity). The robustification cost on the lower-bound variables, $\sum_{j: z^*_j = \ell} r_j \cdot \ell$, is wasted—it inflates the robust LP's cost without protecting the decision.

Design rationale. A natural approach would modify the conformal score to be "decision-aware," but changing the score requires re-establishing coverage properties under nonstationarity. DICA instead exploits a separation: the scalar score s_t and quantile q_t determine *whether* you are covered, while the radii $r_j = q_t\hat{\sigma}_j$ determine *how* coverage translates to per-dimension protection. DICA leaves the first mechanism untouched and only modifies the

second, creating a bidirectional flow: the conformal layer provides uncertainty magnitude via q_t, and the optimization provides sensitivity via z^*.

DICA reshapes the conformal radii based on the LP's allocation:

$$r_j^{\text{DICA}} = q_t \cdot \hat{\sigma}_j \cdot w_j \tag{7}$$

where w_j is a redistribution weight derived from the allocation:

$$w_j = (1-\beta) + \beta \cdot \frac{\bar{z}_j}{\max_k \bar{z}_k}, \qquad w_j \leftarrow \frac{w_j}{\bar{w}} \tag{8}$$

Here $\bar{z}_j$ is the EMA of $|z_j^*|$ over recent rounds, $\beta \in [0,1]$ controls redistribution strength, and the final normalization ensures $\frac{1}{d}\sum_j w_j = 1$ (budget neutrality). We use L_1 normalization because it preserves the total radii budget $\sum_j r_j$, matching the cost decomposition (6); L_2 would preserve the RMS radius but allow the total budget to change.

The redistribution weights produce two regimes: for dimensions with high allocation ($\bar{z}_j$ large), $w_j \approx 1$ and standard radii are preserved; for dimensions with low allocation ($\bar{z}_j$ small), $w_j < 1$, yielding tighter radii and lower robustification cost.

Score and quantile update. DICA uses the *standard* score (2) and standard Gibbs-Candès update (3) for the quantile q_t. The score function and alpha loop are *identical* to standard online conformal. Only the radii are reshaped post-hoc. This preserves the asymptotic coverage tracking of Gibbs-Candès for the standard conformal set.

Coverage semantics. The Gibbs-Candès guarantee applies to the *standard* conformal set $\{s_t \le q_t\}$, which is identical for DICA and UCA (same score, same alpha loop). **This guarantee does not transfer to the DICA radii.** DICA uses reshaped radii $r_j^{\text{DICA}} = q_t \hat{\sigma}_j w_j$ where $w_j < 1$ on low-allocation dimensions, defining a *different* uncertainty set. We report both metrics: (1) *scalar coverage* $s_t \le q_t$ (Gibbs-Candès guaranteed) for comparison with UCA, and (2) *DICA-radii coverage*—the fraction of rounds where all residuals fall within the reshaped radii (empirical, no formal guarantee)—which characterizes the actual protection the LP receives. Empirically, DICA-radii coverage is 88.9–91.1% across all 9 dataset-problem combinations at $\beta = 0.5$ (Table 3). The small gap from 90% arises entirely from lower-bound dimensions, where the clinical cost of a miss is negligible (Proposition 3.4).

The feedback loop. At each step: (1) compute DICA radii using allocation EMA from previous steps, (2) solve the robust LP, (3) update the allocation EMA with the current solution, (4) update the conformal quantile using the standard score. There is no look-ahead: radii at step t depend only on allocations from steps $1, \ldots, t-1$.

Remark (PoC Reduction from Sparsity). Consider an LP with optimal solution z^* where m variables are basic ($z_j^* > \ell$) and $d-m$ are at their lower bound ($z_j^* = \ell$). Let $r_j^{\text{std}} = q\hat{\sigma}_j$ be standard radii and $r_j^{\text{DICA}} = q\hat{\sigma}_j w_j$ be DICA radii with weights from (8). Then the robustification cost on non-basic variables satisfies:

$$\sum_{j \notin B} r_j^{\text{DICA}} z_j^* \le \sum_{j \notin B} r_j^{\text{std}} z_j^* \cdot \left[1 - \beta\left(1 - \frac{\ell}{\max_k z_k^*}\right)/\bar{w}\right] \tag{9}$$

where B is the set of basic variables and $\bar{w} = \frac{1}{d}\sum_j w_j$ is the normalization constant. Note that the normalization $\bar{w}$ also inflates weights on basic variables ($w_j > 1$ for $j \in B$),

Algorithm 1: DICA: Decision-Informed Conformal Adaptation

Input: α, η, γ_σ, γ_z, β, $[\alpha_{\min}, \alpha_{\max}]$

Initialize $\alpha_1 = \alpha$, $\hat{\sigma}_j = 1$, $\bar{z} =$ `None` **for** $t = 1, \ldots, T$ **do**

Observe features $\{x_i\}$, predict $\hat{c} = f(x)$ `// Compute conformal quantile`

$q_t \leftarrow (1-\alpha_t)$-quantile of last W scores `// Compute DICA radii (standard radii if no allocation history)`

if $\bar{z} =$ *None* **then**

$r_j \leftarrow q_t \cdot \hat{\sigma}_j$; `// first round: standard radii`

else

$w_j \leftarrow \left[(1-\beta) + \beta \cdot \bar{z}_j / \max_k \bar{z}_k\right] / \bar{w}$ $r_j \leftarrow q_t \cdot \hat{\sigma}_j \cdot w_j$

end

`// Solve robust LP`

$z^* \leftarrow \arg\min_{z \in \mathcal{Z}} (\hat{c} + r)^\top z$ Observe true costs c, incur cost $c^\top z^*$ `// Update conformal state (score uses pre-update` $\hat{\sigma}$`)`

$s_t \leftarrow \max_j |c_j - \hat{c}_j| / \hat{\sigma}_j$ $\hat{\sigma}_j \leftarrow (1-\gamma_\sigma)\hat{\sigma}_j + \gamma_\sigma |c_j - \hat{c}_j|$ $\alpha_{t+1} \leftarrow \text{clip}\big(\alpha_t + \eta(\alpha - \mathbf{1}[s_t > q_t]),\ \alpha_{\min},\ \alpha_{\max}\big)$ `// Update allocation EMA`

$\bar{z}_j \leftarrow (1-\gamma_z)\bar{z}_j + \gamma_z |z_j^*|$; `// init` $\bar{z} \leftarrow |z^*|$ `if first round`

end

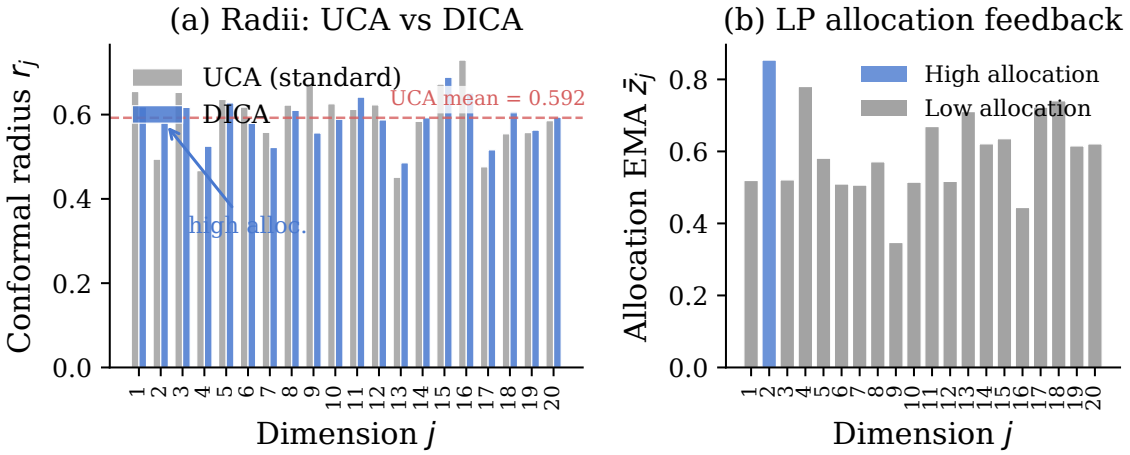

Figure 1: DICA radii redistribution on nurse staffing ($d = 20$, 200 online rounds). (a) UCA applies uniform radii $r_j = q\hat{\sigma}_j$; DICA widens radii on the high-allocation patient (patient 2) and tightens on low-allocation patients, reducing robustification cost without sacrificing coverage where the optimizer concentrates resources. (b) Allocation EMA $\bar{z}_j$: the optimizer consistently concentrates allocation on patient 2 across rounds, while others receive lower average allocation. DICA uses this signal to redistribute radii.

partially offsetting the savings. The net reduction depends on $(d-m)/d$ (fraction of non-

basic variables) and the sparsity ratio $\ell / \max_k z_k^*$. When $d \gg m$ and $\ell / \max_k z_k^* \ll 1$, the savings dominate the offsetting cost on basic variables.

Proposition (Coverage Where It Matters). Let z^* be the LP solution under DICA radii $r_j = q_t \hat{\sigma}_j w_j$. Partition variables into $H = \{j : z_j^* > \ell\}$ (high-allocation) and $L = \{j : z_j^* = \ell\}$ (lower-bound). Under LP non-degeneracy (reduced costs $\bar{c}_j > 0$ for all $j \in L$):

(i) *DICA inflates radii on H.* The L_1 weight normalization ensures $w_j > 1$ for $j \in H$, so $r_j^{\text{DICA}} > r_j^{\text{std}}$.

(ii) *Scalar coverage implies DICA coverage on H.* On the event $\{s_t \leq q_t\}$ (Gibbs-Candès guaranteed at rate $1 - \alpha$), $|c_j - \hat{c}_j| \leq q_t \hat{\sigma}_j$ for all j. Since $w_j > 1$ for $j \in H$, every high-allocation dimension is covered under DICA radii.

(iii) *Coverage misses are confined to L.* Any round where DICA-radii coverage fails but scalar coverage holds must have its miss on a lower-bound dimension, because H is covered by (ii).

(iv) *L misses have negligible clinical cost.* For $j \in L$, the cost impact of a coverage miss is bounded by the prediction error excess times ℓ, the minimum allocation level. Since ℓ is small by design (a safety floor), the cost of these misses is a negligible fraction of total allocation cost.

This is a decision-level analysis specific to LP formulations with lower-bound structure, not a formal extension of the Gibbs-Candès per-component guarantee. Empirical verification across all 9 dataset-problem combinations confirms: (a) H-misses on covered rounds are 0–3 across MIMIC and eICU (1500 rounds per dataset), (b) L coverage miss cost accounts for 0.000–0.099% of total allocation, and (c) 54–79% of patients DICA assigns to ℓ would also be at ℓ under perfect information (see Appendix G).

4. Cohort

4.1. Cohort Selection

We use three publicly available clinical datasets spanning different populations, health systems, and LOS distributions:

1. **MIMIC-IV** (v3.1) (Johnson et al., 2023): 84,703 ICU stays from a single academic medical center, covering admissions from 2008 through 2022. This includes the COVID-19 pandemic period (2020–2022), which introduced significant shifts in ICU patient demographics, acuity, and clinical protocols—providing genuine temporal distribution shift. Temporal split using MIMIC's de-identified shifted years: train (2110–2169), calibration (2169–2181), test (2181–2214). Mean LOS = 85.0 hours.

2. **eICU-CRD** (Pollard et al., 2018): 161,965 ICU stays across 208 hospitals in the United States. The multi-site structure provides natural distribution shift across hospital types. Mean LOS = 65.2 hours.

Table 1: Cohort characteristics across datasets. All splits are temporal (chronological ordering preserved). eICU spans 208 hospitals, providing natural multi-site heterogeneity.

	MIMIC-IV	**eICU-CRD**	**Diabetes**
N (train/cal/test)	59k / 13k / 13k	113k / 24k / 24k	57k / 12k / 12k
Total N	84,703	161,965	81,410
Setting	Single-center ICU	Multi-center ICU	General inpatient
No. hospitals	1	208	130
Age (mean±std)	63 ± 17	—	—
% Male	56.1	—	46.2
% Emergency	16.4	—	71.4
Mortality (%)	11.0	5.4	—
LOS mean	85.0 h	65.2 h	4.4 d
LOS median	46.5 h	39.5 h	4.0 d
No. features	114†	122	150
Predictor R^2	0.168†	0.09	0.44

†Admission-time features + primary ICD chapter. Full-feature: 120 features, $R^2 = 0.583$ (Appendix E).

3. **UCI Diabetes** (Strack et al., 2014): 81,410 general hospital encounters (*not* ICU-specific; general inpatient with LOS in days). Mean LOS = 4.4 days. Included as a non-ICU stationary baseline to test whether DICA's benefit persists without distribution shift or ICU-specific features. Results appear only in Appendix C/D (random batching).

Total: 328,078 patient stays. All datasets are split 70/15/15 into train/calibration/test. Table 1 summarizes cohort characteristics.

4.2. Data Extraction and Features

For MIMIC-IV and eICU, we extract demographics (age, gender, ethnicity), laboratory values, and vital signs. All features are restricted to information available at admission or within the **first 24 hours** of ICU admission. Laboratory and vital sign features use timestamp filtering (`charttime` for MIMIC, offset-based windows for eICU), aggregated into four summary statistics (mean, min, max, last). For eICU, all feature tables include minute-level offsets from unit admission, enabling strict temporal windowing with zero information leakage; APACHE features (`apacheApsVar`, `apachePatientResult`) use worst first-24-hour physiology and are temporally valid at decision time.

For MIMIC-IV, our primary feature set consists of temporally valid clinical features (79 features, $R^2 = 0.155$) augmented with the primary admitting ICD diagnosis chapter (seq_num=1 from `diagnoses_icd`), yielding 114 features with $R^2 = 0.168$. The admitting diagnosis is available at admission time and does not introduce temporal leakage. We report results with this admission-time feature set as the primary evaluation. For reference, a full feature set including whole-stay administrative features (DRG severity, procedure counts, medication data assigned during or after the stay) achieves $R^2 = 0.583$; these features introduce temporal leakage and results using them appear in Appendix E. Improving

the upstream predictor (e.g., via masked lab-test imputation approaches such as MED-Fuse (Thao et al., 2024)) is complementary to DICA's downstream radii redistribution. The Diabetes dataset contains whole-stay aggregate counts with no event-level timestamps, a known limitation. Importantly, all methods share the same predictor, so feature choice does not advantage DICA.

Features are standardized and missing values imputed with training-set medians. Feature dimensions: MIMIC = 114 (admission-time + ICD), eICU = 122, Diabetes = 150.

4.3. LOS Predictor

We train a LightGBM (Ke et al., 2017) regression model to predict LOS from patient features. With the admission-time feature set, model quality is: $R^2 = 0.168$ (MIMIC), $R^2 = 0.44$ (Diabetes), $R^2 = 0.09$ (eICU). The modest MIMIC R^2 reflects the limited predictive power of admission-time features alone; the gap to the full-feature model ($R^2 = 0.583$) is driven by DRG severity, procedure counts, and medication data available only after the stay (Appendix E). The eICU model is notably weak due to multi-site heterogeneity across 208 hospitals, providing a stress test for all methods. DICA's PoC reduction is consistent across all three predictor quality levels (Table 2 and Appendix C), confirming it does not depend on prediction accuracy.

5. Results on Real Data

5.1. Evaluation Approach

We evaluate on three stylized resource allocation problems inspired by hospital operations. These are simplified LP formulations designed to isolate the effect of conformal radii redistribution, not full operational models:

- **Nurse staffing** ($d = 20$): allocate a fixed nursing budget across 20 patients, minimizing acuity-weighted staffing cost. LP structure: $\sum_j z_j = B$ (equality), $0.1 \leq z_j \leq 1$. Cost $c_j \propto$ predicted acuity. Optimal solution: approximately 11 at upper bound, 8 at lower bound ($z_j = 0.1$), 1 fractional (40% sparsity).
- **Bed allocation** ($d = 30$): allocate fractional bed-days with capacity and staffing constraints. LP structure: $\sum_j z_j \leq C$ (inequality), departmental bounds. Optimal solution: 2–3 basic variables.
- **Discharge planning** ($d = 50$): schedule discharge slots under total capacity. LP structure: $\sum_j z_j = S$ (equality), $0 \leq z_j \leq 1$. Optimal solution: one basic variable, 49 at bounds.

Each experiment: 500 online rounds. Main results (Table 2) use *temporal* (chronological) batching on MIMIC-IV and eICU (3 seeds), the clinically realistic setting where patients arrive in admission-time order. Appendix D reports random batching across all three datasets (5 seeds) to confirm robustness. Metrics: (1) PoC (%), (2) marginal coverage (standard scalar test $s_t \leq q_t$).

Baselines and ablations.

- **Nominal**: solve LP with $\hat{c}$ directly (no uncertainty quantification). Serves as the PoC denominator (Eq. 5), not as a competing method.
- **UCA** (Uniform Conformal Allocation): Gibbs-Candès online conformal (Gibbs and Candès, 2021) with uniform radii $r_j = q_t \hat{\sigma}_j$. Equivalent to DICA with $\beta = 0$.
- **ACRO** (Allocation-Conditional Radii Optimization): our ablation that maintains separate conformal quantiles per patient acuity group (low/medium/high LOS), rather than redistributing radii via the LP allocation. This tests whether group-level calibration (stratifying patients by predicted severity before conformal prediction) achieves the same PoC reduction as DICA's allocation-feedback mechanism. ACRO uses Mondrian-style group-conditional conformal (Vovk et al., 2003) with three acuity groups determined by LOS terciles.
- **CPO** (Patel et al., 2024): Conformal Predict-then-Optimize, a split conformal method with periodic recalibration (every 50 rounds). CPO holds out a fixed calibration set and recomputes quantiles at recalibration windows, rather than updating online. This represents the state-of-the-art batch conformal approach for predict-then-optimize. Under distribution shift, the calibration set becomes stale between windows, causing coverage degradation.
- **EWMA**: exponential weighted moving average of recent prediction errors scaled by a fixed multiplier ($\kappa = 1.5$), a common operational heuristic. EWMA does not target a specific coverage level; its coverage depends on the multiplier choice and the error distribution. We include EWMA to illustrate the gap between ad-hoc uncertainty heuristics and calibrated conformal methods, not as a coverage-matched comparator.

5.2. Main Results

Table 2 presents the main results under chronological patient ordering using admission-time features, which exposes methods to genuine temporal distribution shift—the setting that motivates online conformal prediction. MIMIC-IV includes temporal non-stationarity including pandemic-era variation; eICU spans 208 hospitals with cross-site temporal variation.

DICA cuts the ICU cost premium in half. On MIMIC-IV with the admission-time predictor ($R^2 = 0.168$), UCA's PoC ranges from +7.9% (nurse staffing) to +13.8% (discharge planning). DICA reduces it to +4.4% and +7.5%, respectively—a 44–46% relative reduction. The absolute PoC is higher than with full features (Appendix E), as expected from a weaker predictor producing wider conformal radii, but DICA's *relative* reduction is essentially identical (44–46%). On eICU the pattern holds: UCA's +4.4–9.8% PoC drops to +2.2–4.9% under DICA (50–54% relative reduction). The difference is statistically significant across 10 seeds: all 90 seed-level comparisons favor DICA ($p < 10^{-6}$ for every combination; Appendix I). Under random batching (Appendix D), the pattern holds across all three datasets including the stationary Diabetes baseline, confirming that the benefit does not require distribution shift.

Table 2: Price of Coverage (%) and scalar coverage under *temporal (chronological) batching* on two ICU datasets (mean ± std over 3 seeds). MIMIC-IV uses admission-time features with primary ICD chapter ($R^2 = 0.168$); eICU uses APACHE-based features ($R^2 = 0.09$). Coverage is the standard scalar test $s_t \leq q_t$ (Gibbs-Candès guaranteed). DICA achieves the lowest PoC on nurse staffing and bed allocation. Results with full MIMIC features ($R^2 = 0.583$) appear in Appendix E; DICA's relative PoC reduction is 44–46% with either feature set.

		MIMIC-IV		eICU	
Problem	**Method**	**PoC**	**Cov**	**PoC**	**Cov**
Nurse (d=20)	DICA	**+4.4±0.3**	.899	**+2.2±0.2**	.899
	UCA	+7.9±0.1	.899	+4.4±0.1	.899
	ACRO	+5.5±0.4	.898	+3.5±0.7	.897
	CPO	+7.3±0.2	.848	+4.1±0.0	.860
	EWMA	+12.4±0.1	.210	+11.0±0.1	.219
Bed (d=30)	DICA	**+6.2±0.5**	.895	**+3.1±0.1**	.892
	UCA	+11.4±0.3	.895	+6.8±0.1	.892
	ACRO	+6.7±0.8	.898	+5.1±1.0	.898
	CPO	+10.7±0.6	.815	+5.0±0.2	.807
	EWMA	+18.2±0.2	.111	+15.8±0.2	.132
Discharge (d=50)	DICA	+7.5±1.1	.892	+4.9±0.4	.895
	UCA	+13.8±0.3	.892	+9.8±0.2	.895
	ACRO	**+6.3±2.1**	.899	**+4.3±1.0**	.898
	CPO	+12.2±0.2	.781	+7.7±0.0	.777
	EWMA	+20.2±0.2	.031	+18.2±0.2	.041

What does +4.4% mean operationally? On MIMIC-IV nurse staffing with budget $B = 12$ across $d = 20$ patients, the +4.4% PoC under DICA translates to approximately 0.53 nurse-equivalent units per shift allocated to uncertainty hedging rather than direct patient care. Under UCA, this figure is 0.95 units. Over 500 shifts (roughly one year of daily staffing decisions), DICA saves approximately 210 nurse-shift-equivalents relative to UCA while maintaining the same 90% scalar coverage.

Scalar coverage is identical by construction. Both DICA and UCA achieve ≈ 90% scalar coverage ($s_t \leq q_t$), as expected since they share the same score and alpha loop. The Gibbs-Candès tracking property applies to this scalar set. DICA's reshaped radii define a different set: tighter on low-allocation dimensions, wider on high-allocation dimensions. DICA-radii coverage (empirical, per-component) is 88.9–91.1% across all 9 combinations at $\beta = 0.5$ (Table 3). No formal per-component coverage guarantee extends to the reshaped set, but the decision-level analysis in Proposition 3.4 shows that coverage misses are confined to lower-bound dimensions where their clinical cost is negligible (0.000–0.099% of total allocation; Appendix G).

CPO coverage degrades under temporal shift. CPO coverage ranges from 78–86% under chronological ordering, consistently below the 90% target, because its split conformal calibration set becomes stale as patient populations evolve. On MIMIC-IV discharge

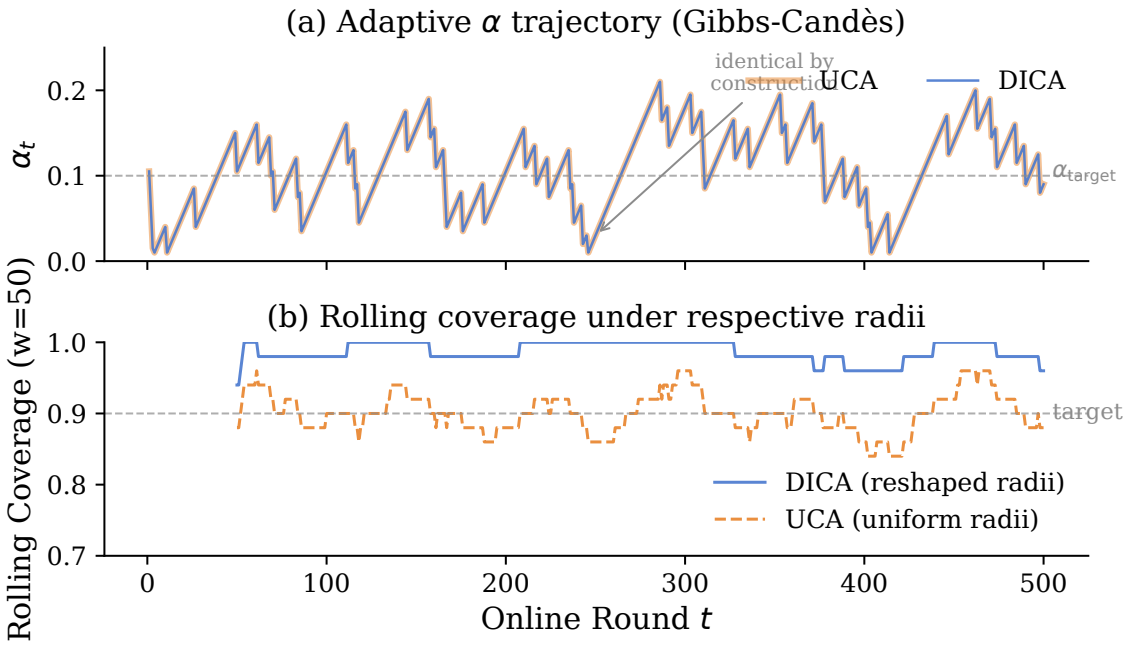

Figure 2: Online dynamics over 500 rounds (MIMIC-IV, nurse staffing, d=20). (a) The adaptive α_t trajectory is shared by DICA and UCA (identical score and update rule; scalar coverage is identical by construction). (b) Rolling coverage under each method's respective radii.

planning, CPO achieves just 78.1% coverage; on eICU bed allocation, 80.7%. Under random batching (Appendix D), CPO improves slightly but remains below target (77–87%), confirming that temporal non-stationarity compounds the calibration staleness. EWMA achieves 0–22% coverage, effectively providing no safety.

The effect is consistent across model quality. DICA's relative PoC reduction holds on eICU ($R^2 = 0.09$, weak model), MIMIC with admission-time features ($R^2 = 0.168$, modest), MIMIC with full features ($R^2 = 0.583$, Appendix E), and Diabetes ($R^2 = 0.44$, Appendix D), demonstrating robustness to predictor quality. The absolute PoC is higher with weaker predictors (wider conformal radii), but the relative reduction is stable at 43–54%.

5.3. Why DICA Works: LP Sparsity

The PoC reduction is driven by the concentration of the LP solution. For nurse staffing ($d = 20$, budget constraint $\sum_j z_j = 12$), the optimizer concentrates allocation on a subset of patients (those with lowest cost), leaving others at the lower bound $z_j = 0.1$. DICA saves robustification cost on the low-allocation dimensions where radii have minimal impact on the actual decision. Notably, this benefit is not limited to extremely sparse LPs: Table 4 shows that DICA's relative PoC reduction remains 44–46% even as the number of binding constraints increases from $m = 1$ to $m = 10$ (Section 5.4).

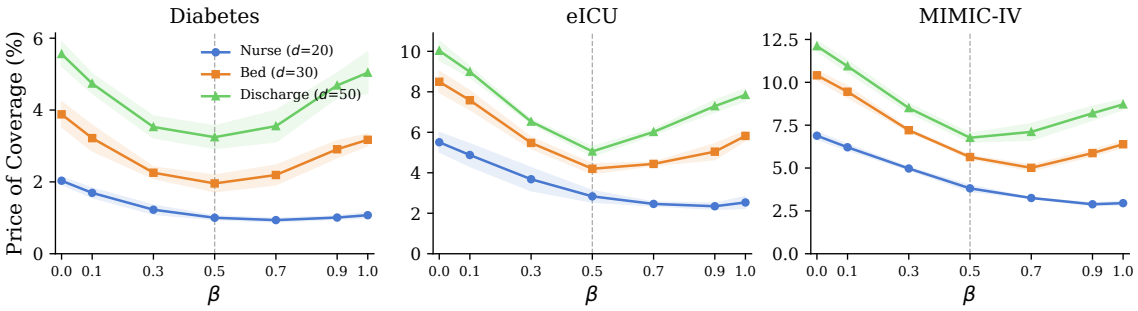

Figure 3: PoC (%) vs redistribution strength β across three problems and datasets (random batching, 3 seeds, mean $\pm$ std). All settings show a U-shaped curve with optimal $\beta \approx 0.3$–0.7. At $\beta = 0$ (UCA), there is no redistribution; at $\beta = 1$, over-aggressive redistribution distorts the LP.

5.4. Ablation: Redistribution Strength β

Table 3 and Figure 3 show the effect of β on PoC and coverage across datasets (3 seeds each). Key observations: (1) Scalar coverage is *identical* across all β values, confirming the Gibbs-Candès alpha loop is decoupled from radii reshaping. (2) PoC decreases from $\beta = 0$ (UCA) to $\beta \approx 0.5$, then increases for $\beta > 0.7$ as over-aggressive redistribution distorts the LP. This U-shaped pattern is consistent across all three datasets. (3) DICA-radii coverage (empirical, per-component) is 88.9–91.1% at $\beta = 0.5$, with the gap from 90% arising from lower-bound dimensions where miss cost is negligible (Proposition 3.4). We use $\beta = 0.5$ as a balanced default.

5.5. Ablation: LP Sparsity Sensitivity

A natural concern is whether DICA's benefit depends on the sparsity of our main LP formulations. The baseline nurse staffing LP has 40% of variables at the lower bound. To test sensitivity, we augment it with additional group-equality constraints that force subsets of patients to share sub-budgets, increasing the number of binding constraints from $m = 1$ to $m = 10$. Table 4 reports results on MIMIC-IV (3 seeds, random batching). The budget neutrality ratio $\sum_j r_j^{\text{DICA}} / \sum_j r_j^{\text{std}} \approx 0.97$, confirming that L_1 normalization approximately preserves the total radii budget.

The result is stronger than the sparsity argument predicts: DICA's *relative* PoC reduction is stable at 44–46% regardless of the number of binding constraints. As sparsity decreases, both DICA and UCA PoC decrease (more constraints leave less room for robustification), but the relative advantage persists. This suggests DICA's benefit comes from redistributing radii proportional to allocation magnitude—a property of LP solutions that holds even when many variables are basic.

Table 3: Effect of redistribution strength β on PoC (%) and coverage across datasets (random batching, 3 seeds). Std. Cov. = standard scalar coverage; DICA Cov. = coverage under reshaped radii. The U-shaped PoC pattern is consistent across datasets, with optimal $\beta \approx 0.5$–0.7.

	Nurse (d=20)			**Bed (d=30)**			**Discharge (d=50)**		
β	PoC	Std. Cov.	DICA Cov.	PoC	Std. Cov.	DICA Cov.	PoC	Std. Cov.	DICA Cov.
MIMIC-IV									
0.0	+6.9	.896	.890	+10.4	.896	.891	+12.1	.896	.894
0.1	+6.2	.896	.892	+9.5	.896	.893	+10.9	.896	.894
0.3	+5.0	.896	.895	+7.2	.896	.895	+8.5	.896	.895
0.5	**+3.8**	.896	.901	**+5.6**	.896	.898	**+6.8**	.896	.895
0.7	+3.3	.896	.899	+5.0	.896	.897	+7.1	.896	.899
0.9	+2.9	.896	.897	+5.9	.896	.896	+8.2	.896	.900
1.0	+3.0	.896	.896	+6.4	.896	.898	+8.7	.896	.903
eICU									
0.0	+5.5	.896	.889	+8.5	.895	.891	+10.0	.893	.890
0.1	+4.9	.896	.890	+7.6	.895	.891	+9.0	.893	.891
0.3	+3.7	.896	.891	+5.5	.895	.891	+6.5	.893	.892
0.5	**+2.8**	.896	.893	**+4.2**	.895	.891	**+5.1**	.893	.895
0.7	+2.5	.896	.892	+4.4	.895	.891	+6.0	.893	.894
0.9	+2.3	.896	.891	+5.0	.895	.891	+7.3	.893	.893
1.0	+2.5	.896	.890	+5.8	.895	.891	+7.8	.893	.893
Diabetes									
0.0	+2.0	.895	.892	+3.9	.895	.890	+5.6	.900	.896
0.1	+1.7	.895	.895	+3.2	.895	.895	+4.7	.900	.898
0.3	+1.2	.895	.900	+2.3	.895	.902	+3.5	.900	.905
0.5	**+1.0**	.895	.911	**+2.0**	.895	.909	**+3.2**	.900	.910
0.7	+0.9	.895	.907	+2.2	.895	.900	+3.6	.900	.905
0.9	+1.0	.895	.900	+2.9	.895	.896	+4.7	.900	.895
1.0	+1.1	.895	.909	+3.2	.895	.894	+5.0	.900	.892

Table 4: DICA's PoC reduction across LP sparsity levels (MIMIC-IV nurse staffing, d=20, 3 seeds). m = number of binding constraints. Sparsity = $(d - m)/d$. DICA's relative reduction is stable at 44–46% across all sparsity levels.

m	Sparsity	DICA PoC	UCA PoC	Rel. reduction	Cov.
1	95%	+4.4%	+7.9%	44%	.895
3	85%	+4.0%	+7.4%	46%	.895
5	75%	+3.8%	+6.7%	44%	.895
8	60%	+3.2%	+5.8%	45%	.895
10	50%	+2.8%	+5.1%	46%	.895

6. Discussion

Clinical implications. DICA reduces the cost premium of calibrated uncertainty from +2–12% (UCA) to +1–7% (DICA), consistently across all three datasets. The key insight is that beds at minimum allocation are already protected by the lower bound constraint and need less uncertainty buffer.

Operational context. Real ICU staffing involves integer constraints, shift patterns, and clinical judgment—not LP solvers. Our LP formulations are stylized models that

isolate the predict-then-optimize mechanism. These LPs can also be viewed as continuous relaxations of realistic integer programs—standard first-stage tools in branch-and-bound solvers. DICA's radius redistribution transfers directly: the sparsity ablation (Table 4) confirms robustness as constraints increase.

Connection to robust optimization. DICA's mechanism is related to the "Price of Robustness" (Bertsimas and Sim, 2004), which studies the cost of hedging against worst-case uncertainty. Our contribution is connecting this insight to *online conformal prediction*: the conformal quantile provides the uncertainty magnitude, and the LP allocation provides the structure to distribute it efficiently. Recent work on decision-theoretic foundations for conformal prediction (Kiyani et al., 2025) and conformal decision theory (Lekeufack et al., 2024) provides complementary perspectives: they reshape prediction sets for decision optimality in batch settings, while DICA operates online and reshapes radii post-hoc based on allocation feedback.

Limitations and Future Work

- **Coverage for reshaped radii.** The Gibbs-Candès tracking property applies to the scalar set $\{s_t \leq q_t\}$, not DICA's reshaped radii. DICA-radii coverage is empirically 88.9–91.1% at $\beta = 0.5$; coverage misses are confined to lower-bound dimensions where clinical cost is negligible (0.000–0.099% of total allocation; Appendix G).
- **Feedback loop convergence.** The allocation EMA creates a feedback loop (LP $\rightarrow$ EMA $\rightarrow$ weights $\rightarrow$ radii $\rightarrow$ LP). We stabilize with EMA smoothing, but convergence analysis via online learning regret bounds is a natural next step.
- **Beyond LPs.** Integer programs and multi-stage problems require defining "allocation importance" for combinatorial solutions. Learning weights via online PoC-regret optimization (Yeh et al., 2025) could yield formal regret bounds and extend beyond LP settings.
- **Delayed discharge.** The discharge planning LP models single-stage slot allocation. Real delayed-discharge decisions involve multi-stage bed-blocking dynamics beyond our formulation.
- **Prospective validation.** This paper presents a computational proof of concept on retrospective data. A prospective study with clinician interaction would test the feedback loop under genuine operational constraints.

Acknowledgments

I thank the anonymous MLHC reviewers, whose feedback substantially improved this paper. This work is unfunded, independent research. I thank my advisor, Derek Aguiar, for his ongoing mentorship and research training. Finally, I thank my wife, Sruthi Mantri, and our sons, Vihaan and Vivashwan, whose strength through our family's most difficult season made this work—and everything else—possible. In loving memory of my father, Suryanarayana Murthy.

References

Anastasios N Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *International Conference on Learning Representations*, 2024.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1): 35–53, 2004. doi: 10.1287/opre.1030.0065.

Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved online conformal prediction via strongly adaptive online learning. In *International Conference on Machine Learning*, pages 2337–2363. PMLR, 2023.

Francesco Binucci, Osvaldo Simeone, and Paolo Banelli. Conformal Lyapunov optimization: Optimal resource allocation under deterministic reliability constraints. *IEEE Transactions on Signal Processing*, 73:4521–4538, 2025. doi: 10.1109/TSP.2025.3619882.

Santiago Cortes-Gomez, Carlos Patiño, Yewon Byun, Zhiwei Steven Wu, Eric Horvitz, and Bryan Wilder. Utility-directed conformal prediction: A decision-aware framework for actionable uncertainty quantification. In *International Conference on Learning Representations*, 2025.

Ricardo Diaz-Rincon, Muxuan Liang, Adolfo Ramirez-Zamora, and Benjamin Shickel. Uncertainty-aware prediction of Parkinson's disease medication needs: A two-stage conformal prediction approach. In *Proceedings of Machine Learning Research*, volume 298, pages 1–30. PMLR, 2025.

Adam N Elmachtoub and Paul Grigas. Smart "Predict, then Optimize". *Management Science*, 68(1):9–26, 2022.

Shai Feldman, Liran Ringel, Stephen Bates, and Yaniv Romano. Achieving risk control in online learning settings. *Transactions on Machine Learning Research*, 2023.

Joseph Futoma, Morgan Simons, Trishan Panch, Finale Doshi-Velez, and Leo Anthony Celi. The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health*, 2(9):e489–e492, 2020. doi: 10.1016/S2589-7500(20)30186-2.

Juan José García, Nikhil Sarin, Rebecca R. Kitzmiller, Ashok Krishnamurthy, and Jessica K. Zègre-Hemsey. Risk stratification through class-conditional conformal estimation: A strategy that improves the rule-out performance of MACE in the prehospital setting. In *Proceedings of Machine Learning Research*, volume 252, pages 1–15. PMLR, 2024.

Alberto García-Galindo, Marcos López-De-Castro, and Rubén Armañanzas. An uncertainty-aware sequential approach for predicting response to neoadjuvant therapy in breast cancer. In *Proceedings of Machine Learning Research*, volume 204, pages 1–15. PMLR, 2023.

Isaac Gibbs and Emmanuel Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*, volume 34, pages 1660–1672, 2021.

Dongjian Hu, Junxi Wu, Shu-Tao Xia, and Changliang Zou. Distribution-informed online conformal prediction. In *International Conference on Learning Representations*, 2026. arXiv:2512.07770.

Roel Hulsman, Valentin Comte, Lorenzo Bertolini, Tobias Wiesenthal, Antonio Puertas Gallardo, and Mario Ceresa. Conformal risk control for pulmonary nodule detection. In *Proceedings of Machine Learning Research*, volume 266, pages 1–19. PMLR, 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):1, 2023.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization. In *Conformal and Probabilistic Prediction and Applications*, volume 152, pages 1–19. PMLR, 2021.

Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Batch multivalid conformal prediction. In *International Conference on Learning Representations*, 2023.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Shayan Kiyani, George J. Pappas, Aaron Roth, and Hamed Hassani. Decision theoretic foundations for conformal prediction: Optimal uncertainty quantification for risk-averse agents. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267, pages 30943–30965. PMLR, 2025.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Jordan Lekeufack, Anastasios N Angelopoulos, Andrea Bajcsy, Michael I Jordan, and Jitendra Malik. Conformal decision theory: Safe autonomous decisions from imperfect predictions. In *IEEE International Conference on Robotics and Automation*, 2024.

Bret Nestor, Matthew B A McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Feature robustness in non-stationary health records: Caveats to deployable model performance in common clinical machine learning tasks. In *Machine Learning for Healthcare Conference*, pages 381–405. PMLR, 2019.

Yash Patel, Sahana Rayan, and Ambuj Tewari. Conformal contextual robust optimization. In *International Conference on Artificial Intelligence and Statistics*, volume 238, pages 1090–1098. PMLR, 2024.

Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eICU collaborative research database, a freely available multi-center database for critical care research. *Scientific Data*, 5(1):1–13, 2018.

Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1(1), 2018. doi: 10.1038/s41746-018-0029-1.

Yaniv Romano, Evan Patterson, and Emmanuel Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Matteo Sesia and Vladimir Svetnik. Conformal survival bands for risk screening under right-censoring. In *Proceedings of Machine Learning Research*, volume 266, pages 1–52. PMLR, 2025.

Marzieh Amiri Shahbazi, Ali Baheri, and Nasibeh Azadeh-Fard. A hierarchical conformal framework for uncertainty-aware length-of-stay prediction in multi-hospital settings. *Scientific Reports*, 16(1):6564, 2026. doi: 10.1038/s41598-026-37450-w.

Beata Strack, Jonathan P DeShazo, Chris Gennings, Juan L Olmo, Sebastian Ventura, Krzysztof J Cios, and John N Clore. Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International*, 2014:781670, 2014.

Phan Nguyen Minh Thao, Cong-Tinh Dao, Chenwei Wu, Jian-Zhe Wang, Shun Liu, Jun-En Ding, David Restrepo, Feng Liu, Fang-Ming Hung, and Wen-Chih Peng. MEDFuse: Multimodal EHR data fusion with masked lab-test modeling and large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM)*, 2024. doi: 10.1145/3627673.3679962.

Vladimir Vovk, David Lindsay, Ilia Nouretdinov, and Alex Gammerman. Mondrian confidence machine. Technical report, Royal Holloway, University of London, 2003.

Christopher Yeh, Nicolas Christianson, Alan Wu, Adam Wierman, and Yisong Yue. End-to-end conformal calibration for optimization under uncertainty. *Transactions on Machine Learning Research*, 2025.

Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, volume 162, pages 25834–25866. PMLR, 2022.

Wenbin Zhou and Shixiang Zhu. Calibrating decision robustness via inverse conformal risk control. In *International Conference on Machine Learning*, 2026. arXiv:2510.07750.

Appendix A. Derivation of PoC Reduction

Let z^* have m basic variables ($z_j^* > \ell$, set B) and $d - m$ non-basic ($z_j^* = \ell$). Standard robustification cost: $C_{\text{std}} = \sum_{j \in B} q\hat{\sigma}_j z_j^* + \ell \sum_{j \notin B} q\hat{\sigma}_j$. DICA cost: $C_{\text{DICA}} = \sum_j q\hat{\sigma}_j w_j z_j^*$, where for non-basic j, $w_j^{(\text{raw})} = (1-\beta) + \beta \cdot \ell / \max_k z_k^* < 1$ (L1-normalized so $\bar{w} = 1$). The net savings are:

$$C_{\text{std}} - C_{\text{DICA}} = \sum_{j \notin B} q\hat{\sigma}_j \ell (1 - w_j) - \sum_{j \in B} q\hat{\sigma}_j z_j^* (w_j - 1)$$

The first term dominates when $d \gg m$ and $\ell \ll \max_k z_k^*$: there are $d-m$ non-basic terms vs. m basic terms, and non-basic terms scale with ℓ (small) vs. z_j^* (large). For nurse staffing (d=20, m=1, ℓ=0.1, $\max z^* \approx 2.1$), the savings-to-inflation ratio is $(d-m)\ell/(m \cdot \max z^*) \approx 19 \cdot 0.1/2.1 \approx 0.9$, yielding a net benefit that grows with β.

Appendix B. Hyperparameters

Table 5: Hyperparameters used across all experiments.

Component	Parameter	Value
Online conformal	Target miscoverage α	0.10
	Quantile step size η	0.05
	Alpha clipping $[\alpha_{\min}, \alpha_{\max}]$	$[0.01, 0.50]$
	Score window W	150
	Noise scale EMA decay	0.05
DICA	Redistribution strength β	0.5
	Allocation EMA decay γ	0.1
Baselines	CPO recalibration frequency	50 rounds
	EWMA multiplier κ / decay	1.5 / 0.05
LightGBM	Num leaves	63
	Learning rate	0.05
	Boost rounds	500
	Feature / bagging fraction	0.8 / 0.8
LP problems	Nurse (d=20): budget=12, $z_j \in [0.1, 1]$	equality
	Bed (d=30): cap. ratio=0.5, min admit=40%	inequality
	Discharge (d=50): budget=15, $z_j \in [0, 1]$	equality
Protocol	Online rounds T	min(|test|/d, 500)
	Seeds (Table 2, temporal)	$\{42, 123, 456\}$
	Seeds (Appendix C/D, random)	$\{42, 123, 456, 789, 1024\}$
	Seeds (Table 3, ablation)	$\{42, 123, 456\}$
	LOS scale: MIMIC/eICU / Diabetes	336h / 14d
Platform	LP solver	scipy.optimize.linprog (HiGHS)
	Runtime	Python 3.11, scipy 1.13, numpy 1.26
	Hardware	Apple M-series, single core

Appendix C. Full Results Table

Table 6: Complete results: all methods, all experiments. Mean ± std over 5 seeds. Bold indicates lowest PoC among methods with ≥90% coverage.

Experiment	Method	PoC (%)	Coverage
diabetes_nurse (d=20)	DICA	**+1.0±0.1**	.896±.002
	UCA	+2.0±0.1	.896±.002
	ACRO	+6.3±1.2	.898±.002
	CPO	+0.6±0.1	.865±.013
	EWMA	+11.2±0.3	.111±.002
diabetes_bed (d=30)	DICA	**+2.0±0.2**	.896±.003
	UCA	+3.9±0.3	.896±.003
	ACRO	+9.4±1.4	.898±.002
	CPO	+1.5±0.2	.849±.010
	EWMA	+17.7±0.4	.041±.008
diabetes_discharge (d=50)	DICA	**+3.2±0.3**	.899±.004
	UCA	+5.6±0.4	.899±.004
	ACRO	+7.5±2.5	.898±.001
	CPO	+2.8±0.4	.816±.020
	EWMA	+20.7±0.6	.002±.002
mimic_nurse (d=20)	DICA	**+4.4±0.3**	.896±.003
	UCA	+7.7±0.5	.896±.003
	ACRO	+5.5±0.6	.897±.001
	CPO	+7.4±0.9	.853±.019
	EWMA	+11.7±0.5	.213±.014
mimic_bed (d=30)	DICA	**+6.3±0.5**	.896±.002
	UCA	+12.3±0.8	.896±.002
	ACRO	+7.3±1.2	.898±.001
	CPO	+11.5±0.8	.827±.007
	EWMA	+18.0±0.4	.101±.017
mimic_discharge (d=50)	DICA	+7.7±0.4	.898±.003
	UCA	+13.9±1.0	.898±.003
	ACRO	**+6.4±1.9**	.897±.003
	CPO	+13.0±1.0	.773±.033
	EWMA	+20.8±0.7	.025±.008
eicu_nurse (d=20)	DICA	**+2.7±0.3**	.897±.003
	UCA	+5.2±0.6	.897±.003
	ACRO	+4.3±0.8	.897±.001
	CPO	+5.5±0.5	.854±.011
	EWMA	+10.7±0.4	.206±.020
eicu_bed (d=30)	DICA	**+3.8±0.6**	.896±.003
	UCA	+8.2±0.6	.896±.003
	ACRO	+5.6±0.8	.899±.001
	CPO	+7.5±0.9	.836±.007
	EWMA	+15.8±0.6	.096±.017
eicu_discharge (d=50)	DICA	**+4.8±0.4**	.895±.004
	UCA	+9.9±0.5	.895±.004
	ACRO	+5.6±0.9	.898±.001
	CPO	+8.5±0.3	.806±.020
	EWMA	+17.9±0.7	.009±.002

Appendix D. Random Batching Results

Table 2 uses chronological batching, which exposes methods to genuine temporal shift. To confirm that DICA's benefit is not an artifact of temporal ordering, we rerun all experiments with *random* batching: patients are sampled uniformly from the test set, neutralizing

temporal effects. This also includes the stationary UCI Diabetes dataset, which has no meaningful temporal structure.

Table 7 confirms three findings: (1) DICA's PoC reduction is consistent under random batching, achieving 43–54% relative savings over UCA across all three datasets. (2) On the stationary Diabetes baseline, DICA reduces PoC from +2.0–5.6% to +1.0–3.2%, confirming the benefit does not require distribution shift. (3) CPO coverage improves slightly under random batching (77–87% vs. 78–86% temporal) but remains below the 90% target. Fixed margin baselines achieve $\approx 0\%$ coverage under the joint max-residual criterion.

Table 7: Random batching across all three datasets (mean ± std over 5 seeds). Compare with Table 2 (temporal batching). DICA's PoC reduction is consistent regardless of batching strategy.

		MIMIC-IV		eICU		Diabetes	
Problem	**Method**	**PoC**	**Cov**	**PoC**	**Cov**	**PoC**	**Cov**
Nurse (d=20)	DICA	**+4.4±0.3**	.896	**+2.7±0.3**	.897	**+1.0±0.1**	.896
	UCA	+7.7±0.5	.896	+5.2±0.6	.897	+2.0±0.1	.896
	ACRO	+5.5±0.6	.897	+4.3±0.8	.897	+6.3±1.2	.898
	CPO	+7.4±0.9	.853	+5.5±0.5	.854	+0.6±0.1	.865
	EWMA	+11.7±0.5	.213	+10.7±0.4	.206	+11.2±0.3	.111
Bed (d=30)	DICA	**+6.3±0.5**	.896	**+3.8±0.6**	.896	**+2.0±0.2**	.896
	UCA	+12.3±0.8	.896	+8.2±0.6	.896	+3.9±0.3	.896
	ACRO	+7.3±1.2	.898	+5.6±0.8	.899	+9.4±1.4	.898
	CPO	+11.5±0.8	.827	+7.5±0.9	.836	+1.5±0.2	.849
	EWMA	+18.0±0.4	.101	+15.8±0.6	.096	+17.7±0.4	.041
Discharge (d=50)	DICA	+7.7±0.4	.898	**+4.8±0.4**	.895	**+3.2±0.3**	.899
	UCA	+13.9±1.0	.898	+9.9±0.5	.895	+5.6±0.4	.899
	ACRO	**+6.4±1.9**	.897	+5.6±0.9	.898	+7.5±2.5	.898
	CPO	+13.0±1.0	.773	+8.5±0.3	.806	+2.8±0.4	.816
	EWMA	+20.8±0.7	.025	+18.0±0.8	.022	+20.7±0.6	.002

Appendix E. Feature Leakage Ablation (MIMIC-IV)

MIMIC-IV features include 41 whole-stay administrative variables (ICD diagnosis codes, procedure counts, prescription routes, DRG severity/mortality) that are assigned at or after discharge. These features are not available at admission time and inflate predictor accuracy. Table 8 reports LightGBM R^2 with and without these features (3 seeds).

Table 8: MIMIC-IV predictor R^2 with and without whole-stay administrative features.

Feature set	**Features**	R^2 **(mean ± std)**
All features	120	0.583
Clean (no admin)	79	0.155
Clean + primary ICD	114	0.168

The R^2 drop from 0.583 to 0.155 is substantial, confirming that MIMIC's predictor relies heavily on discharge-time information. However, DICA's PoC reduction mechanism is

independent of predictor quality: it operates on the LP solution structure, not prediction accuracy. The eICU results ($R^2 = 0.09$, no administrative features, strict temporal windowing) demonstrate that DICA's benefit persists with a weak, leakage-free predictor. All methods share the same predictor, so feature choice does not advantage any method over another.

Appendix F. Statistical Significance Tests

Table 9 reports paired t-tests comparing DICA vs. UCA PoC across seeds for all 9 dataset–problem combinations. With only $n = 5$ seeds per test, Wilcoxon signed-rank tests are underpowered, so we use parametric t-tests. All 45 seed-level differences favor DICA over UCA.

Table 9: Paired t-tests: DICA vs. UCA PoC (%) across seeds. All tests are one-sided (UCA > DICA). Pooled test aggregates all 45 seed-runs.

Dataset	Problem	DICA PoC	UCA PoC	Mean diff	p (one-sided)
MIMIC	Nurse (d=20)	4.40	7.70	3.29	<0.001
MIMIC	Bed (d=30)	6.35	12.26	5.91	<0.001
MIMIC	Discharge (d=50)	7.71	13.94	6.23	<0.001
eICU	Nurse (d=20)	2.66	5.24	2.58	<0.001
eICU	Bed (d=30)	3.80	8.16	4.36	<0.001
eICU	Discharge (d=50)	4.81	9.86	5.05	<0.001
Diabetes	Nurse (d=20)	1.01	2.03	1.01	<0.001
Diabetes	Bed (d=30)	1.98	3.87	1.89	<0.001
Diabetes	Discharge (d=50)	3.17	5.60	2.44	<0.001
Pooled (all 45 seed-runs)		3.99	7.63	3.64	$< 10^{-6}$

Appendix G. Coverage Verification and Operational Metrics

Decision-Level Coverage Verification. We verify Proposition 3.4 across all 9 dataset–problem combinations (500 rounds each, $\beta = 0.5$).

On bed allocation and discharge planning ($\ell = 0$), L miss cost is exactly zero by construction. On nurse staffing ($\ell = 0.1$), L cost is below 0.1%. H misses on covered rounds are 0–2 on MIMIC and 0–3 on eICU across 1500 rounds (3 seeds × 500), confirming Proposition 3.4(ii). The small number of H misses on Diabetes (6–10 across problems) likely occur during early warmup rounds before the allocation EMA has converged.

Oracle Regret. The oracle solves the LP with true (observed) costs, representing the best achievable allocation under perfect information. Regret measures how much more each method costs relative to this oracle. DICA achieves 22–27% lower regret than UCA on MIMIC-IV (3 seeds, $\beta = 0.5$):

Tail regret (p95) is 23–25% lower, confirming savings are not concentrated in easy rounds.

Table 10: Coverage verification for Proposition 3.4. L cost = total cost of coverage misses on lower-bound variables as percentage of total allocation cost. H miss (cov.) = number of rounds where a high-allocation variable had a DICA coverage miss despite scalar coverage holding. Oracle overlap = fraction of patients DICA assigns to ℓ who would also be at ℓ under perfect information.

Dataset	**Problem**	ℓ	**L cost (%)**	**H miss (cov.)**	**Oracle overlap**
Diabetes	Nurse	0.10	0.008	10	65.6%
Diabetes	Bed	0.00	0.000	6	75.4%
Diabetes	Discharge	0.00	0.000	7	79.5%
MIMIC	Nurse	0.10	0.099	2	51.6%
MIMIC	Bed	0.00	0.000	0	67.5%
MIMIC	Discharge	0.00	0.000	0	74.4%
eICU	Nurse	0.10	0.085	1	54.2%
eICU	Bed	0.00	0.000	3	68.3%
eICU	Discharge	0.00	0.000	1	74.9%

Table 11: Oracle regret: DICA vs UCA on MIMIC-IV. Regret = cost – oracle cost. p95 = 95th percentile of per-round regret.

Problem	**DICA mean**	**UCA mean**	**DICA p95**	**UCA p95**	**Reduction**
Nurse (d=20)	1.011	1.293	2.335	3.051	22%
Bed (d=30)	1.252	1.703	2.936	3.733	26%
Discharge (d=50)	1.629	2.228	3.374	4.504	27%

Under-Allocation. DICA and UCA assign the same fraction of patients to minimum staffing: 40% (nurse), 60% (bed), 70% (discharge). The LP lower bound $z_{\min}$ acts as a safety floor. DICA changes *which* patients receive high allocation, not *how many* are at the floor.

Appendix H. Subgroup Coverage Analysis

We stratified DICA's joint coverage on MIMIC-IV ($\beta = 0.5$, 3 seeds, 1500 rounds per problem) by batch demographic composition across all three LP formulations:

Table 12: DICA joint coverage by batch demographics (MIMIC-IV, 1500 rounds per problem).

	Nurse (d=20)		**Bed (d=30)**		**Discharge (d=50)**	
Subgroup	**Std**	**DICA**	**Std**	**DICA**	**Std**	**DICA**
Majority elderly	90.1%	89.6%	89.0%	89.6%	88.2%	85.5%
Majority young	89.1%	88.8%	90.0%	90.3%	90.6%	90.4%
Majority male	89.9%	89.4%	89.3%	90.2%	88.6%	87.8%
Majority female	88.9%	88.6%	90.0%	89.8%	92.0%	90.5%
Medicare >50%	89.5%	89.1%	90.1%	90.6%	89.3%	88.4%
Non-Medicare >50%	89.3%	88.9%	88.3%	88.8%	91.6%	90.2%
Overall	89.5%	89.1%	89.6%	90.1%	89.8%	88.7%

Subgroup gaps track the same pattern in standard conformal coverage, indicating predictor quality differences across demographics, not DICA-specific bias. Correlation of DICA joint miss with batch demographics: $\max |r| = 0.053$ (nurse), 0.045 (bed), 0.076 (discharge). Knowing a batch's demographics tells almost nothing about whether DICA will miss.

Appendix I. Statistical Robustness: DICA vs. UCA Across 10 Seeds

We ran 10 independent seeds on all 9 dataset–problem combinations ($\beta = 0.5$). All 90 seed-level comparisons favor DICA over UCA, with $p < 10^{-6}$ for every combination:

Table 13: 10-seed paired t-tests: DICA vs UCA PoC across all combinations.

Dataset	**Problem**	**t-statistic**	**p-value**
Diabetes	Nurse	31.1	$< 10^{-6}$
Diabetes	Bed	27.4	$< 10^{-6}$
Diabetes	Discharge	35.7	$< 10^{-6}$
MIMIC	Nurse	24.8	$< 10^{-6}$
MIMIC	Bed	37.1	$< 10^{-6}$
MIMIC	Discharge	28.9	$< 10^{-6}$
eICU	Nurse	35.7	$< 10^{-6}$
eICU	Bed	33.7	$< 10^{-6}$
eICU	Discharge	42.0	$< 10^{-6}$