APPENDIX
Safe Learning of Adaptive Control Policies for Remote Patient Monitoring

APPENDIX I
PROOF FOR THEOREM 1

Proof. Theorem 1 from [1] states that when $H \to \infty$, the optimal policy is π_o when

$$\gamma(\lambda_i - \lambda_o) \leq C_i/C_c. \tag{1}$$

Here π_o is the policy where the patient always stays in ordinary monitoring, i.e., threshold-based policy $\pi_{t,\bar{h}}$ with $\bar{h} = 0$. Theorem 2 from [1] states that the optimal policy is $\pi_{t,\bar{h}}$ with some $\bar{h}$ when the following are satisfied -

$$\gamma(\lambda_i - \lambda_o) > C_i/C_c, \tag{2}$$

and

$$\frac{\gamma\mu_o(1+\gamma\mu_o)}{1-\gamma^2\lambda_o\mu_o} \leq 1. \tag{3}$$

Note that (2) is the complement of (1) and hence one of them is always true. Hence we only need to verify that equation (3) is satisfied. Condition (3) is satisfied when

$$\gamma \leq \frac{-1+\sqrt{1+4/\mu_o}}{2},$$

which is satisfied for sufficiently large $1/\gamma$. This completes the proof for Theorem 1. □

APPENDIX II
CONVERGENCE GUARANTEES

We use the same model, notation, threshold policy convention, and epoch wise learning algorithm as defined in the main paper. In particular, $h^\star$ denotes the optimal threshold under the true parameters, $\hat{h}_k$ denotes the optimal threshold computed corresponding to the current parameter estimates at epoch k. We now state the additional assumptions needed for the safety and almost sure convergence arguments.

Assumption 1. *The health-state space is finite and is given by $\mathcal{H} = \{0, 1, \ldots, H\}$, where 0 is the critical state and H is the maximum health state. Each epoch consists of L patient-state transitions, where $L > H$. Moreover, the one-step upward and downward movement probabilities under both actions are positive, i.e., $p_{\min} = \min\{\lambda_i, \lambda_o, \mu_i, \mu_o\} > 0$.*

Assumption 2. *At epoch k, if the computed threshold $\hat{h}_k$ lies outside the interval $[h_{low}, h_{high}]$, the algorithm implements the threshold h_{low} or h_{high} with probability ρ_k. These probabilities satisfy $\sum_{k=1}^{\infty} \rho_k = \infty$.*

Assumption 3. *Let $\theta = (\lambda_i, \lambda_o, C_i, C_o)$ denote the true parameter vector. There exists $\varepsilon_{\text{pol}} > 0$ such that for every parameter vector θ' satisfying $\|\theta' - \theta\|_\infty < \varepsilon_{\text{pol}}$, the unique optimal threshold under θ' is $h^\star$.*

Assumption 4. *The cost noise under each action $a \in \{i, o\}$ is sub-Gaussian. That is, if $G_{a,t}$ denotes the cost observed on the t-th real sample of action a, then $G_{a,t} - C_a$ is sub-Gaussian. Let $\sigma^2 < \infty$ denote a common conditional sub-Gaussian variance proxy for the cost noises and the transition observation noises.*

We can now state the main result of this appendix.

Theorem 1. *Suppose $h_{low} \leq h^\star \leq h_{high}$. Under Assumptions 1, 2, 3, and 4, there exists an almost surely finite random epoch K_G such that, for all $k \geq K_G$, the threshold deployed by the algorithm at epoch k is equal to the optimal threshold $h^\star$.*

A. Proof

We first prove that each action is sampled infinitely often. Then we use this to prove convergence of the parameter estimates to the true parameters, eventual correctness of the computed threshold, and finally eventual correctness of the deployed threshold.

Lemma 1. *Let $X_{a,k}$ denote the number of real samples of action $a \in \{i, o\}$ collected during epoch k and $p_{\min} = \min\{\lambda_i, \lambda_o, \mu_i, \mu_o\} > 0$. Then, for each action $a \in \{i, o\}$, there exists $c_a > 0$ such that*

$$\mathbb{E}[X_{a,k} \mid \mathcal{F}_{k-1}] \geq c_a \rho_k. \tag{4}$$

Moreover, for each action $a \in \{i, o\}$, $N_{a,j} \to \infty$ a.s.

Proof. Fix an action $a \in \{i, o\}$ and an epoch k. With a probability at least ρ_k, the deployed threshold has at least one health state s_a at which action a is used. This is possible because $h_{\text{low}} \geq 1$ and $h_{\text{high}} \leq H - 1$, so the deployed thresholds are such that both actions have nonempty state regions. Now fix any starting health state $s \in \{1, \ldots, H\}$. Since the state space has size H, the target state s_a can be reached from s in at most $H - 1$ one-step moves by moving monotonically toward s_a. Each required one-step move has probability at least $p_{\min}$. Since $L > H$, this path fits inside one epoch. Hence, conditional on using such a threshold, the probability of visiting a state where action a is used during the epoch is at least $p_{\min}^H$. Therefore, $\mathbb{P}(X_{a,k} \geq 1 \mid \mathcal{F}_{k-1}) \geq \rho_k p_{\min}^H$. Since $X_{a,k}$ is a nonnegative integer-valued random variable,

$$\begin{aligned}\mathbb{E}[X_{a,k} \mid \mathcal{F}_{k-1}] &\geq \mathbb{P}(X_{a,k} \geq 1 \mid \mathcal{F}_{k-1}) \\ &\geq \rho_k p_{\min}^H.\end{aligned}$$

Thus first part of the lemma holds with $c_a = p^H_{\min} > 0$. Since each epoch has L patient-state transitions, $0 \le X_{a,k} \le L$. Since $X_{a,k} \le L\mathbf{1}_{\{X_{a,k}\ge 1\}}$, we have

$$\mathbb{E}[X_{a,k} \mid \mathcal{F}_{k-1}] \le L\mathbb{P}(\{X_{a,k} \ge 1\} \mid \mathcal{F}_{k-1}).$$

Therefore, by (4),

$$\mathbb{P}(\{X_{a,k} \ge 1\} \mid \mathcal{F}_{k-1}) \ge \frac{1}{L}\mathbb{E}[X_{a,k} \mid \mathcal{F}_{k-1}] \ge \frac{c_a}{L}\rho_k.$$

Since $\sum_{k=1}^{\infty} \rho_k = \infty$, it follows that

$$\sum_{k=1}^{\infty} \mathbb{P}(\{X_{a,k} \ge 1\} \mid \mathcal{F}_{k-1}) = \infty \text{ a.s.}$$

By the extended Borel-Cantelli lemma [2, Corollary 5.29], the events $\{X_{a,k} \ge 1\}$ occur infinitely often almost surely. Hence $\sum_{k=1}^{j} X_{a,k} \to \infty$ a.s. Since $N_{a,j} = n_0 + \sum_{k=1}^{j} X_{a,k}$, we get $N_{a,j} \to \infty$ a.s. This proves the lemma. □

Lemma 2. *The parameter estimates converge to the true parameter vector almost surely, i.e., $\hat{\theta}_k \to \theta$ a.s. Moreover, there exists an almost surely finite random epoch K_G such that $\hat{h}_k = h^\star$ for all $k \ge K_G$.*

Proof. Fix an action $a \in \{i, o\}$. Let $M_{a,k} = N_{a,k} - n_0$ denote the number of real samples of action a collected up to the end of epoch k. By Lemma 1, $M_{a,k} \to \infty$ a.s. Let $Y_{a,t}$ be the upward-movement indicator observed on the t-th real sample of action a. $\mathbb{E}[Y_{a,t} \mid \mathcal{F}_{a,t-1}] = \lambda_a$, where $\mathcal{F}_{a,t-1}$ denotes the information available before the t-th real sample of action a. Therefore $Y_{a,t} - \lambda_a$ is a bounded martingale difference sequence. By the martingale strong law of large numbers [3],

$$\frac{1}{n}\sum_{t=1}^{n}(Y_{a,t} - \lambda_a) \to 0 \qquad \text{a.s.}$$

Since $M_{a,k} \to \infty$ almost surely, the same convergence holds along the random subsequence $n = M_{a,k}$,

$$\frac{1}{M_{a,k}}\sum_{t=1}^{M_{a,k}} Y_{a,t} \to \lambda_a \qquad \text{a.s.}$$

The prior-weighted transition estimate can be written as

$$\hat{\lambda}_{a,k} = \frac{n_0\lambda_{a,p} + \sum_{t=1}^{M_{a,k}} Y_{a,t}}{n_0 + M_{a,k}}.$$

Since $n_0 < \infty$, $\lambda_{a,p}$ is finite, and $M_{a,k} \to \infty$, the prior contribution vanishes. Therefore $\hat{\lambda}_{a,k} \to \lambda_a$ a.s. The same argument applies to the cost estimate. Let $G_{a,t}$ be the cost observed on the t-th real sample of action a. By Assumption 4, $\mathbb{E}[G_{a,t} \mid \mathcal{F}_{a,t-1}] = C_a$, and $G_{a,t} - C_a$ is sub-Gaussian. In particular, it has uniformly bounded second moment. Hence the martingale strong law gives

$$\frac{1}{n}\sum_{t=1}^{n}(G_{a,t} - C_a) \to 0 \qquad \text{a.s.}$$

As before, evaluating this convergence along $n = M_{a,k}$ and as the prior weight is finite, $\hat{C}_{a,k} \to C_a$ a.s. This holds for both $a = i$ and $a = o$. Therefore $\hat{\theta}_k \to \theta$ a.s. As $\hat{\theta}_k \to \theta$ a.s., there exists a finite random epoch K_G such that for every $k \ge K_G$, $\|\hat{\theta}_k - \theta\|_\infty < \varepsilon_{\text{pol}}$. By Assumption 3, whenever this inequality holds, the unique optimal threshold under $\hat{\theta}_k$ is $h^\star$. So, it follows that $\hat{h}_k = h^\star$ for all $k \ge K_G$. This proves the lemma. □

Proof of Theorem 1. By Lemma 2, there exists an almost surely finite random epoch K_G such that $\hat{h}_k = h^\star$ for all $k \ge K_G$. Since $h_{\text{low}} \le h^\star \le h_{\text{high}}$, the optimal threshold lies in the region where the algorithm does not change the threshold. Hence, for all $k \ge K_G$, the threshold deployed by the algorithm at epoch k is equal to the computed threshold $\hat{h}_k$, and both are equal to the optimal threshold $h^\star$. □

APPENDIX III
SAFETY GUARANTEES

The safety guarantee requires two additional assumptions stated below, beyond those stated in Appendix.

Assumption 5. *The optimal integer threshold $h^\star$ is non-decreasing in λ_i and C_o, and non-increasing in λ_o and C_i.*

Assumption 6. *The prior parameter vector $\theta_p = (\lambda_{i,p}, \lambda_{o,p}, C_{i,p}, C_{o,p}, C_c)$ is biased by $\Delta > 0$ such that $\lambda_{i,p} \ge \lambda_i + \Delta, \lambda_{o,p} \le \lambda_o - \Delta, C_{i,p} \le C_i - \Delta$ and $C_{o,p} \ge C_o + \Delta$.*

We can now state the main result of this appendix.

Theorem 2. *Suppose Assumptions 3, 4, 5, and 6 hold, and suppose $h_{low} \le h^\star \le h_{high}$. Fix any $0 < \varepsilon_{\text{safe}} < \varepsilon_{\text{pol}}$. If*

$$n_0 \ge \frac{\sigma^2 \log(4/\delta)}{2\varepsilon_{\text{safe}}(\Delta + \varepsilon_{\text{safe}})},$$

then, with probability at least $1 - \delta$, the threshold deployed by the algorithm at epoch k is at least $h^\star$ for all $k \ge 1$.

A. Proof

We first show that the local uniqueness assumption, together with monotonicity, implies a one-sided safety region around the true parameter vector.

Lemma 3. *Fix any $0 < \varepsilon_{\text{safe}} < \varepsilon_{\text{pol}}$. If a parameter vector $\theta' = (\lambda_i', \lambda_o', C_i', C_o')$ satisfies*

$$\lambda_i' \ge \lambda_i - \varepsilon_{\text{safe}}, \qquad \lambda_o' \le \lambda_o + \varepsilon_{\text{safe}},$$

$$C_i' \le C_i + \varepsilon_{\text{safe}}, \qquad C_o' \ge C_o - \varepsilon_{\text{safe}},$$

then the computed threshold under θ' is safe, i.e., $h(\theta') \ge h^\star$.

Proof. Define the boundary parameter vector

$$\theta^- = (\lambda_i - \varepsilon_{\text{safe}}, \lambda_o + \varepsilon_{\text{safe}}, C_i + \varepsilon_{\text{safe}}, C_o - \varepsilon_{\text{safe}}).$$

Since $\varepsilon_{\text{safe}} < \varepsilon_{\text{pol}}$, $\|\theta^- - \theta\|_\infty = \varepsilon_{\text{safe}} < \varepsilon_{\text{pol}}$. By Assumption 3, the unique optimal threshold under θ^- is $h^\star$. Hence $h(\theta^-) = h^\star$. Now let θ' satisfy the one-sided inequalities in the statement of the lemma. Compared with θ^-, the parameter vector θ' has larger λ_i, smaller λ_o, smaller

C_i, and larger C_o. By Assumption 5, these changes can only increase the optimal threshold. Therefore $h(\theta') \geq h(\theta^-) = h^\star$. This proves the lemma. □

Proof of Theorem 2. We prove that, with probability at least $1-\delta$, the parameter estimates remain in the one-sided safe region of Lemma 3 for all epochs. We first consider the estimate of λ_i. Let Y_t denote the upward movement indicator on the t-th real sample collected under action i. We know, $\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = \lambda_i$. Define $S_n = \sum_{t=1}^{n}(Y_t - \lambda_i)$. Then $(S_n)_{n\geq 0}$ is a martingale. The prior-weighted estimate after n real samples of action i is

$$\hat{\lambda}_i(n) = \frac{n_0\lambda_{i,p} + \sum_{t=1}^{n} Y_t}{n_0 + n}.$$

By Assumption 6, $\lambda_{i,p} \geq \lambda_i + \Delta$. Suppose that for some n, $\hat{\lambda}_i(n) < \lambda_i - \varepsilon_{\text{safe}}$. Then

$$n_0\lambda_{i,p} + \sum_{t=1}^{n} Y_t < (n_0 + n)(\lambda_i - \varepsilon_{\text{safe}}).$$

Using $\lambda_{i,p} \geq \lambda_i + \Delta$, this implies

$$S_n = \sum_{t=1}^{n}(Y_t - \lambda_i) < -n_0(\Delta + \varepsilon_{\text{safe}}) - n\varepsilon_{\text{safe}}.$$

Since $Y_t - \lambda_i$ is conditionally sub-Gaussian with proxy σ^2, for every $\eta > 0$,

$$\mathbb{E}[\exp(-\eta(Y_t - \lambda_i)) \mid \mathcal{F}_{t-1}] \leq \exp\left(\frac{\eta^2\sigma^2}{2}\right).$$

Hence

$$Z_n = \exp\left(-\eta S_n - \frac{\eta^2\sigma^2 n}{2}\right)$$

is a nonnegative supermartingale with $\mathbb{E}[Z_0] = 1$. Choose $\eta = \frac{2\varepsilon_{\text{safe}}}{\sigma^2}$. On the event

$$S_n < -n_0(\Delta + \varepsilon_{\text{safe}}) - n\varepsilon_{\text{safe}},$$

we have

$$-\eta S_n - \frac{\eta^2\sigma^2 n}{2} > \eta n_0(\Delta + \varepsilon_{\text{safe}}) + \left(\eta\varepsilon_{\text{safe}} - \frac{\eta^2\sigma^2}{2}\right) n.$$

By the chosen value of η, $\eta\varepsilon_{\text{safe}} - \frac{\eta^2\sigma^2}{2} = 0$. Therefore, on the event that the estimate crosses the unsafe boundary, we have

$$Z_n > \exp\left(\eta n_0(\Delta + \varepsilon_{\text{safe}})\right).$$

By Ville's inequality for nonnegative supermartingales [4],

$$\mathbb{P}\left(\exists n \geq 1 : \hat{\lambda}_i(n) < \lambda_i - \varepsilon_{\text{safe}}\right) \leq \exp\left(-\eta n_0(\Delta + \varepsilon_{\text{safe}})\right).$$

The same argument applies to the other three estimated coordinates, with the unsafe direction changed according to Assumption 6. By the assumed lower bound on n_0, each of the four probabilities is at most $\delta/4$. Therefore, by a union bound, with probability at least $1-\delta$, none of the four bad events occurs. On this good event, none of the four unsafe crossings occurs at any number of real observations and hence for all epochs. Thus $\hat{\theta}_k$ satisfies the one-sided safe-region inequalities in Lemma 3. Hence $\hat{h}_k \geq h^\star$ for all $k \geq 1$. Since $h_{\text{low}} \leq h^\star \leq h_{\text{high}}$, the algorithm never changes a computed threshold that is at least $h^\star$ to a threshold below $h^\star$. Therefore, the threshold deployed by the algorithm at epoch k is at least $h^\star$ for all $k \geq 1$. This proves the theorem. □

APPENDIX IV
EXPERIMENT DETAILS

In this section, we present complete experimental details for the plots presented in the main paper. We set $h_{\text{low}} = 4$, $h_{\text{high}} = 17$ across all experiments in this section.

a) **Figure 2a** - *Percentage of converged runs after each epoch for different initial thresholds*
- Each epoch comprised of 75 timesteps.
- True Parameters: $H = 20, \lambda_o = 0.2, \lambda_i = 0.55, C_o = 0, C_i = 25, C_c = 50000, \gamma = 0.9$. The optimal threshold is 12.
- Priors: $\lambda_{o,p} = 0.15, \lambda_{i,p} = 0.6, C_{o,p} = 0, C_{i,p} \in \{0.005, 0.1, 0.5, 2.0\}$. The initial thresholds are $20, 18, 16, 14$ for $C_{i,p} = 0.005, 0.1, 0.5, 2.0$, respectively.
- Prior Strength: $n_0 = 1000$
- Probability Sequence: $\rho_k = 0.95k^{-1/3}$, where k denotes the epoch number.

b) **Figure 2b** - *Percentage of converged runs after each epoch for different prior strengths*
- Each epoch comprised of 75 timesteps.
- True Parameters: $H = 20, \lambda_o = 0.2, \lambda_i = 0.55, C_o = 0, C_i = 25, C_c = 50000, \gamma = 0.9$. The optimal threshold is 12.
- Priors: $\lambda_{o,p} = 0.15, \lambda_{i,p} = 0.6, C_{o,p} = 0, C_{i,p} = 0.5$. The initial threshold is 16.
- Prior Strength: $n_0 \in \{10, 100, 200, 500, 1000\}$.
- Probability Sequence: $\rho_k = 0.95k^{-1/3}$, where k denotes the epoch number.

c) **Figure 3a** - *Average time to reach critical health for different initial thresholds*
- Each epoch comprised of 75 timesteps.
- True Parameters: $H = 20, \lambda_o = 0.2, \lambda_i = 0.55, C_o = 0, C_i = 25, C_c = 50000, \gamma = 0.9$. The optimal threshold is 12.
- Priors: $\lambda_{o,p} = 0.15, \lambda_{i,p} = 0.6, C_{o,p} = 0, C_{i,p} \in \{0.005, 0.1, 0.5, 2.0\}$. The initial thresholds are $20, 18, 16, 14$ for $C_{i,p} = 0.005, 0.1, 0.5, 2.0$, respectively.
- Prior Strength: $n_0 = 1000$
- Probability Sequence: $\rho_k = 0.95k^{-1/3}$, where k denotes the epoch number.

d) **Figure 3b** - *Average time to reach critical health for different prior strengths*
- Each epoch comprised of 75 timesteps.
- True Parameters: $H = 20, \lambda_o = 0.2, \lambda_i = 0.55, C_o = 0, C_i = 25, C_c = 50000, \gamma = 0.9$. The optimal threshold is 12.

- Priors: $\lambda_{o,p} = 0.15, \lambda_{i,p} = 0.6, C_{o,p} = 0, C_{i,p} = 0.5$. The initial threshold is 16.
- Prior Strength: $n_0 \in \{10, 100, 200, 500, 1000\}$.
- Probability Sequence: $\rho_k = 0.95k^{-1/3}$, where k denotes the epoch number.

e) **Table 1** - *Safety of deployed threshold for different prior strengths*

- Each epoch comprised of 75 timesteps.
- True Parameters: $H = 20, \lambda_o = 0.2, \lambda_i = 0.55, C_o = 0, C_i = 25, C_c = 50000, \gamma = 0.9$. The optimal threshold is 12.
- Priors: $\lambda_{o,p} = 0.15, \lambda_{i,p} = 0.6, C_{o,p} = 0, C_{i,p} = 0.5$. The initial threshold is 16.
- Prior Strength: $n_0 \in \{10, 100, 200, 500, 1000\}$.
- Probability Sequence: $\rho_k = 0.95k^{-1/3}$, where k denotes the epoch number.

REFERENCES

[1] S. Chandak, I. Thapa, N. Bambos, and D. Scheinker, "Tiered service architecture for remote patient monitoring," in *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, 2024, pp. 1–7.

[2] L. Breiman, *Probability*. Society for Industrial and Applied Mathematics, 1992, vol. 7.

[3] Y. F. Atchadé, "A strong law of large numbers for martingale arrays," 2009, arXiv:0905.2761.

[4] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon, "Time-uniform chernoff bounds via nonnegative supermartingales," *Probability Surveys*, vol. 17, pp. 257–317, 2020.