Supplement 1: Sampling diagnostics (NUTS/HMC)

S1A — Energy Diagnostics

Hamiltonian Monte Carlo sampling was evaluated using the energy plot , which displays (i) the marginal distribution of the Hamiltonian energy E and (ii) the distribution of energy transitions ΔE between successive proposals. Adequate overlap and dispersion between these two distributions indicate efficient exploration of the typical set. As a scalar summary, we report the energy Bayesian fraction of missing information (E-BFMI); values ≥ 0.3 are generally considered adequate, whereas low E-BFMI suggests poor energy exploration and potential pathologies (e.g., sticky trajectories or a need for reparameterization).

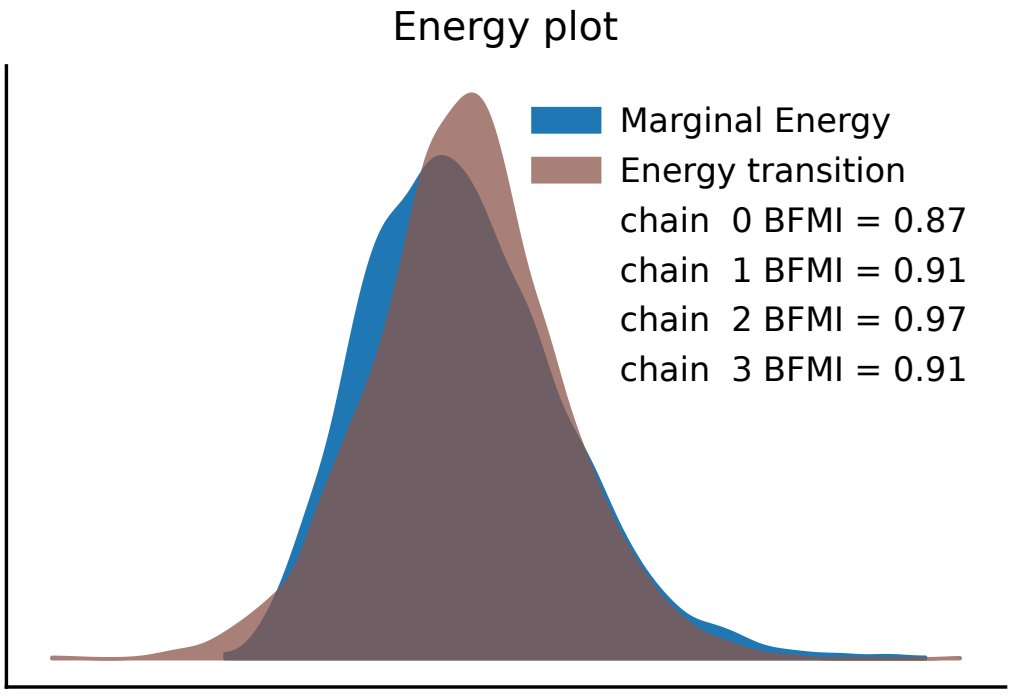


Figure S1A. Energy plot with E-BFMI by chain.

S1B — Trace Diagnostics (By Chain)

We inspected trace plots for representative parameters $(\rho_0, \rho_1, \mu_0, \mu_1, \sigma_0, \sigma_1, \boldsymbol{\beta})$ across chains to assess mixing and stationarity under NUTS/HMC. Well-behaved traces demonstrate rapid mixing, "fat-caterpillar" shapes without persistent trends, and stable running means after warm-up. Between-chain overlays were visually indistinguishable, and autocorrelation decayed rapidly, consistent with convergence and adequate effective sample sizes.

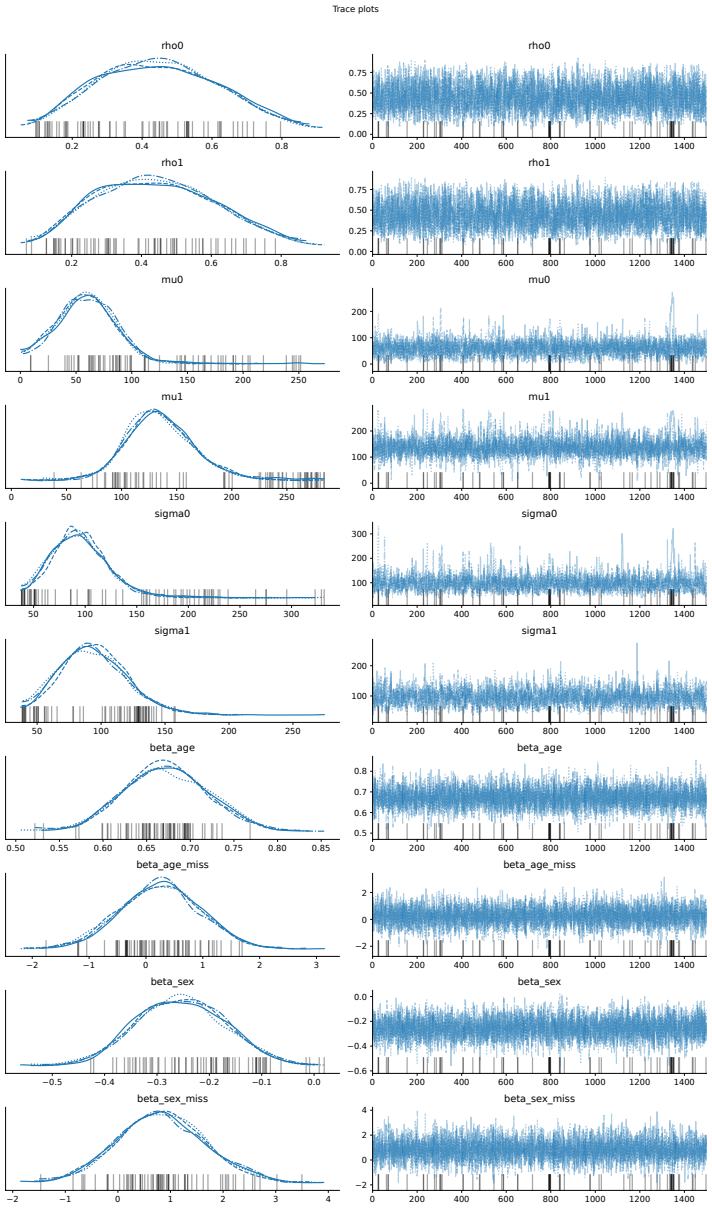


Figure S1B. Trace plots by chain with running means and autocorrelation panels.

S1C —Rank Diagnostics (Uniform Rank Histograms)

Convergence was further evaluated with rank plots (uniform rank histograms). For each parameter, draws from each chain are ranked within the pooled posterior; under good mixing, per-chain rank frequencies are approximately uniform. Systematic deviations (e.g., U- or ∩-shapes or edge spikes) indicate poor between-chain overlap or adaptation issues. The observed histograms were approximately flat with only random fluctuation, supporting convergence.

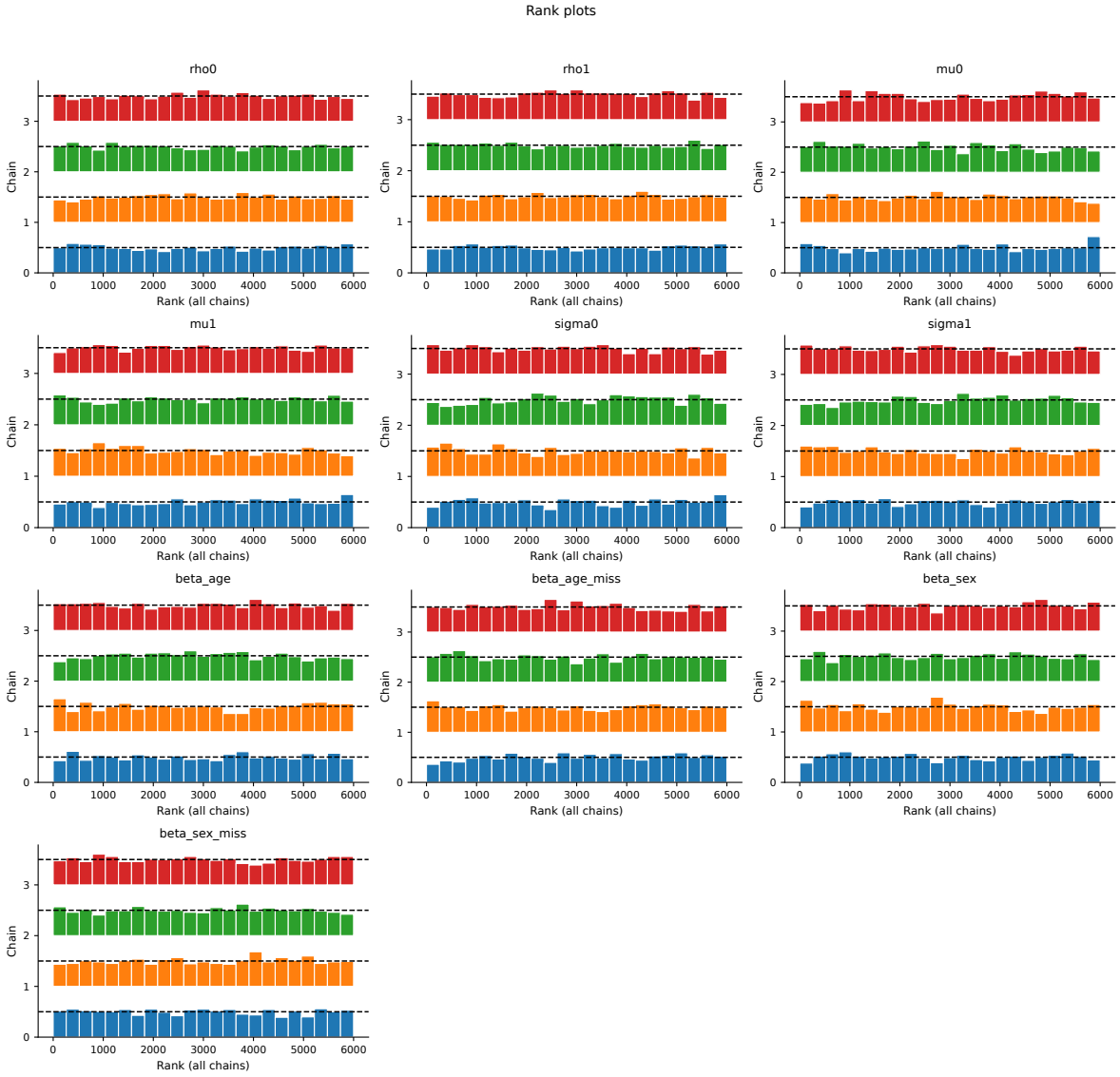
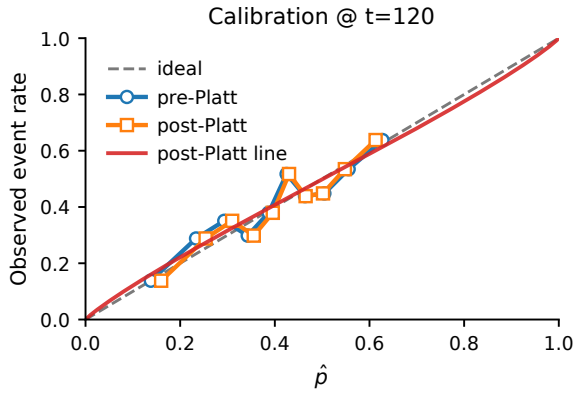


Figure S1C. Uniform rank histograms by chain.

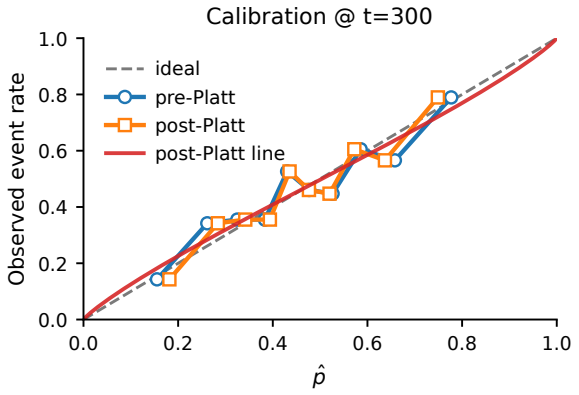
Supplement 2: Recalibration, Decision Curves, and Robustness

S2.A — Logistic/Platt Recalibration

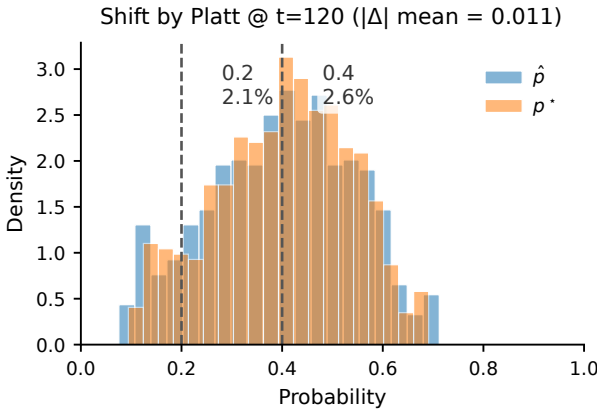
Model. $\text{logit}(p^\star) = \alpha(t) + \beta(t)\,\text{logit}(\hat{p})$ at each landmark t.
Panel A. Calibration at $t = 120$ and $t = 300$: observed event rate (deciles) vs. $\hat{p}$; overlay pre-/post-Platt lines (ideal: slope $= 1$, intercept $= 0$).
Panel B. Distribution shift: histogram/density of $\hat{p}$ vs. $p^\star$; annotate mean absolute change and % crossing thresholds (e.g., 0.2/0.4).



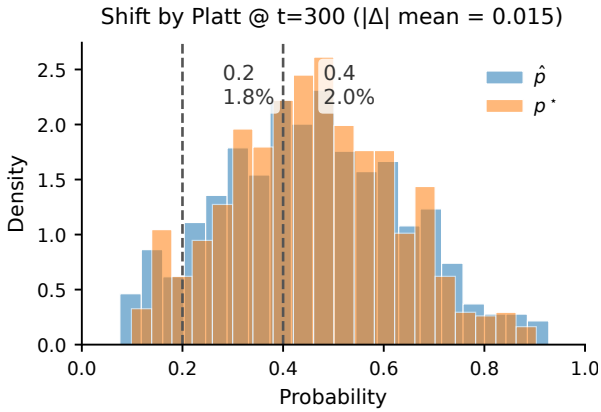
S2.A — Panel A (t=120)



S2.A — Panel A (t=300)



S2.A — Panel B (t=120)



S2.A — Panel B (t=300)

Figure S2A. Platt recalibration at selected landmarks ($t = 120, 300$).

Findings from our run. Platt recalibration at $t = 120/300$ improved calibration toward the 45° line ($\alpha_{120} = -0.010$, $\beta_{120} = 0.899$; $\alpha_{300} = -0.013$, $\beta_{300} = 0.879$; 95% CrIs: add). Discrimination was unchanged (AUC_{120} 0.655→0.655; AUC_{300} 0.681→0.681). The mean absolute shift $|\Delta|$ was 0.011 ($t = 120$) and 0.015 ($t = 300$); only 2.6% and 2.0% crossed $p_t = 0.4$, respectively—indicating improved probability interpretability with minimal decision impact.

S2.B — Decision-Curve Analysis

Purpose. Quantify clinical net benefit across threshold probabilities p_t.
Panel. Curves for Baseline (time only), Covariate (time+age+sex), *Admit all*, *Admit none* over $p_t \in [0.1, 0.6]$.
Formula. $\mathrm{NB}(p_t) = \frac{\mathrm{TP}}{N} - \frac{\mathrm{FP}}{N}\frac{p_t}{1-p_t}$.

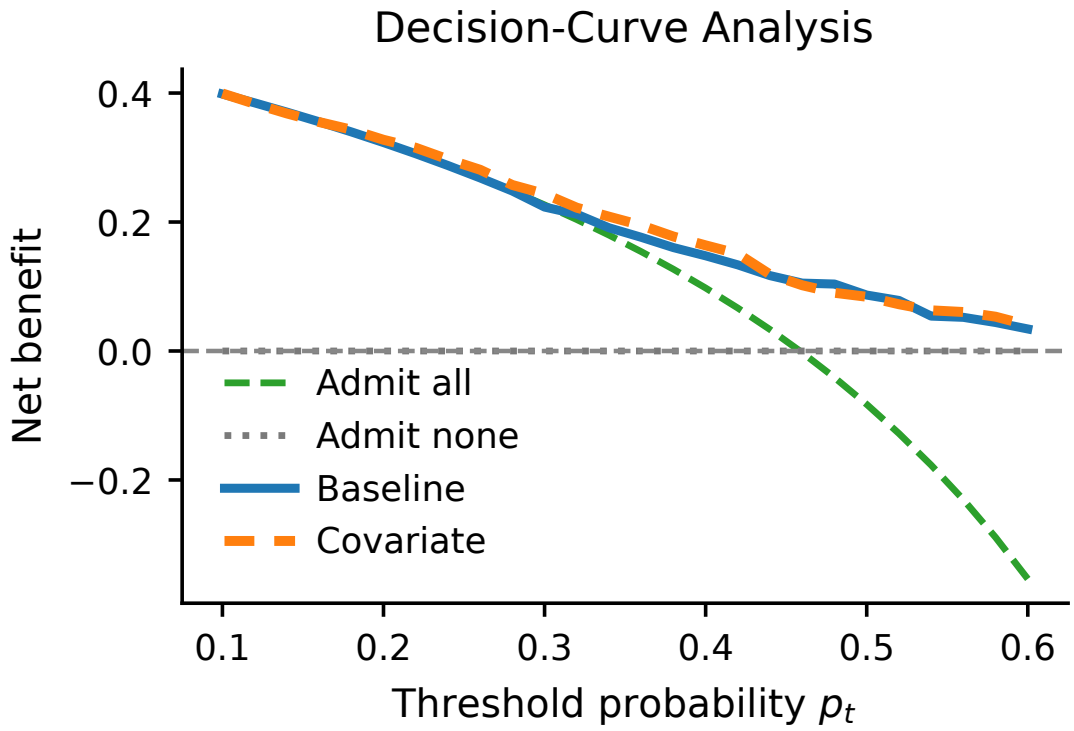


Figure S2B. Decision-curve analysis comparing Baseline vs. Covariate with *Admit all*/*Admit none* references.

Reporting. ΔNB (Covariate − Baseline): p_t = 0.2:0.0043 (95% CI −0.0848,0.0887); p_t = 0.3:0.0205 (95% CI −0.0779,0.1151); p_t = 0.4:0.0162 (95% CI −0.0767,0.1060). Interpretation: Covariate shows small, directionally favorable net benefit for $p_t = 0.2$–0.4, but CIs include zero.

S2.C — Robustness to TTU Measurement Error (One Table)

Add jitter to t_i^{raw}: $\tilde{t}_i = t_i^{\text{raw}} + \epsilon$, with $\epsilon \sim \text{Uniform}(-\delta, \delta)$, $\delta \in \{5, 10\}$ min; then recompute landmark metrics from posterior predictive draws.

Table S2C. Robustness of time-dependent metrics to TTU jitter (landmarks in minutes).

Jitter (±min)	AUC					Brier					Cal. Intercept		Cal. Slope	
	60	120	180	240	300	60	120	180	240	300	120	300	120	300
0.0	0.642	0.655	0.651	0.662	0.681	0.211	0.223	0.228	0.228	0.223	-0.010	-0.010	0.899	0.884
5.0	0.642	0.650	0.648	0.660	0.680	0.212	0.224	0.229	0.228	0.223	-0.036	-0.011	0.861	0.876
10.0	0.642	0.650	0.651	0.662	0.682	0.210	0.224	0.228	0.227	0.222	-0.045	-0.009	0.870	0.891

One-liner for Results. Sensitivity with $\pm 5/\pm 10$-minute TTU jitter yielded $\Delta\text{AUC}(t) \leq 0.005$, $\Delta\text{Brier}(t) \leq 0.001$, and ΔCalibration slope ≤ 0.038 across landmarks; qualitative conclusions were preserved.