PULMOWORLD: AN ASYNCHRONOUS MULTIMODAL MEDICAL WORLD MODEL FOR LONGITUDINAL PNEUMONIA DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pneumonia changes across respiration, inflammation, and organ function, while the measurements that describe it arrive at different times and are often incomplete. We introduce PULMOWORLD, a world model for longitudinal pneumonia dynamics that connects asynchronous state estimation, recursive simulation, and evidence-constrained planning. History-aware filtering separates state propagation from observation-driven correction, so a measurement updates the shared belief only when it is observed. That belief supports recursive forecasts of clinical values and modality representations, and summaries of future prognosis. Cap-MBRL searches imagined continuations inside an authorized action set and replaces a frozen fallback only when a candidate meets the risk budget and a predicted-cost margin. On the pneumonia cohort MASS, recursive forecasts are competitive with task-trained sequence models, and simulated future states retain prognostic information. Trained separately on public MIMIC-IV-CXR chest radiographs and clinical trajectories, this model improves prognostic discrimination over latest-value fusion in development and is competitive on long-horizon error on the locked pneumonia panel. In five known-mechanism MASS-Syn environments, Cap-MBRL lowers selection regret relative to direct fallback execution.

1 INTRODUCTION

Pneumonia evolves across respiratory function, inflammation, renal function, and hemodynamics (Metlay et al., 2019). A current risk score captures one aspect of this process; longitudinal care also requires anticipating how these measurements may change over the coming hours or days. A patient simulator must infer the current physiological state and evolve it through time, including trajectories conditioned on recorded or candidate care.

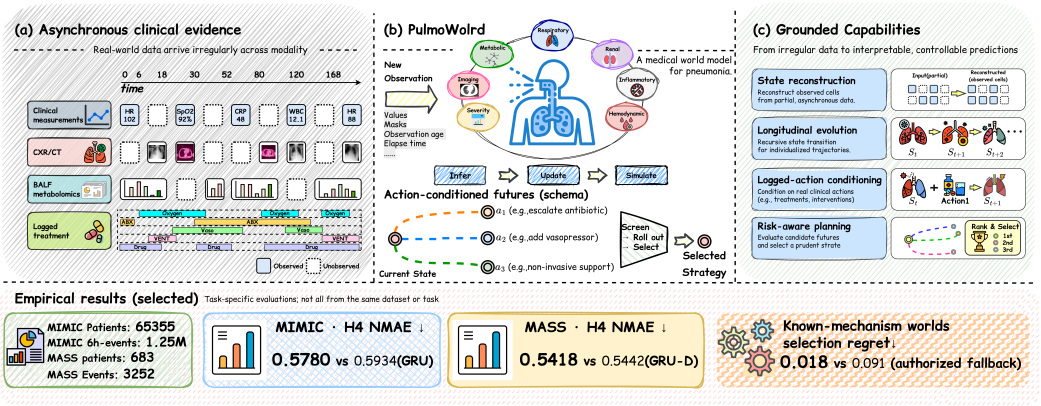


Figure 1: **A shared patient world for longitudinal simulation and evidence-constrained planning.** (a) Asynchronous multimodal observations inform (b) an evolving patient belief and (c) simulation and planning interfaces. The lower strip summarizes cohort forecasting and MASS-Syn selection. Treatment branches show imagined continuations scored in MASS-Syn.

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

The evidence needed to infer this state arrives unevenly. Clinical measurements, chest imaging, and bronchoalveolar lavage fluid (BALF) metabolomics describe complementary aspects of a patient, but differ in observation time and availability (Figure 1(a)). Combining the latest value from each modality can mix evidence of different ages. Conversely, requiring complete multimodal visits discards partially observed histories. The modeling problem is to maintain a patient state between observations, incorporate each available measurement when it arrives, and continue simulating when no further measurements are supplied (Ha & Schmidhuber, 2018; LeCun et al., 2022; Hafner et al., 2019b).

Existing medical world models address important parts of this problem. EHRWorld learns from clinical notes and structured events, while SepsisAgent couples a numerical clinical world model with an LLM treatment policy (Mu et al., 2026; Wu et al., 2026). ChronoMedicalWorld models trajectories from longitudinal care records (Wang et al., 2026). Imaging-centered models include MEWM and ORBIT (Yang et al., 2025; Yan et al., 2026); CLARITY conditions imaging-derived dynamics on clinical context and irregular follow-up intervals (Ding et al., 2025). Our focus is the observation-to-state interface for incomplete clinical, imaging, and molecular histories: these streams supply evidence about an evolving patient, with available modalities changing over time.

PULMOWORLD separates state propagation from observation-driven correction. Elapsed time advances a prior patient belief; an observed measurement corrects that belief through a masked innovation update. A missing measurement supplies no correction. The model tracks evidence freshness separately from state uncertainty, and decodes future clinical values and modality representations from recursively propagated states. This connects asynchronous evidence assimilation to simulation through a shared state interface.

Using those simulations for planning requires an additional decision about which action comparisons have evidential support. Recorded care conditions the simulated trajectory. Cap-MBRL restricts action comparisons to a fixed authorized set and evaluates candidate continuations inside the learned world model. A candidate replaces the frozen fallback only if it meets the risk budget and improves predicted cost by a specified margin; otherwise, the fallback is retained.

Our contributions are threefold:

- **Asynchronous multimodal state estimation.** We develop a history-aware patient belief that integrates incomplete multimodal observations through state propagation and observation-driven correction. On MIMIC-IV-CXR development data, it improves prognostic discrimination relative to latest-value fusion.
- **Recursive patient simulation and predictive utility.** We model clinical and modality-specific evolution within a shared belief space, including treatment-conditioned trajectories. On MASS and on the locked MIMIC pneumonia panel, these forecasts and simulated states retain clinical and prognostic information.
- **Evidence-constrained planning.** We introduce Cap-MBRL, which compares authorized actions through model rollouts and replaces a frozen fallback only under a risk budget and a predicted-cost margin. In known-mechanism MASS-Syn, this selection lowers regret relative to executing the frozen fallback under the same authorization.

2 RELATED WORK

2.1 MEDICAL WORLD MODELS

EHR-based world models learn longitudinal patient dynamics from clinical records. EHRWorld combines note-derived profiles with structured clinical events (Mu et al., 2026); SepsisAgent uses a numerical physiological world model for an LLM's propose–simulate–refine treatment workflow (Wu et al., 2026). ChronoMedicalWorld learns action-conditioned latent transitions from clinical measurements and care communications (Wang et al., 2026). Imaging-centered methods include MEWM and ORBIT (Yang et al., 2025; Yan et al., 2026). CLARITY combines imaging-derived states, clinical and genomic context, and irregular follow-up intervals for trajectory simulation and treatment planning (Ding et al., 2025). Latent imagination, learned-model search, and uncertainty-penalized offline control supply the planning background (Hafner et al., 2019a; 2020; 2025; Schrittwieser et al., 2020; Chua et al., 2018; Yu et al., 2020; 2021). Sepsis-3 supplies the clinical definition, and

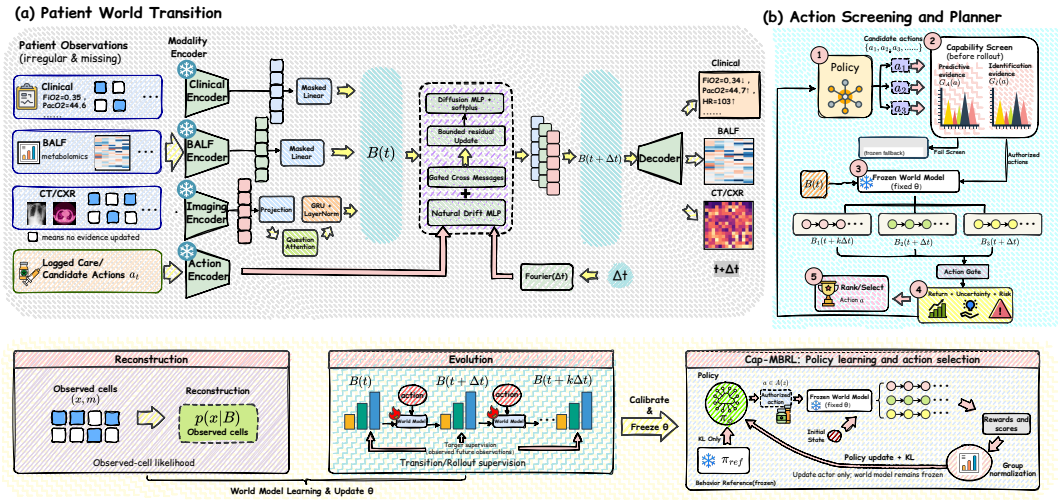


Figure 2: **Pneumonia state learning and evidence-constrained planning.** (a) Irregular clinical, imaging, and BALF observations update one patient belief, which a shared transition then propagates. (b) In MASS-Syn, a capability screen authorizes actions before frozen-world rollouts; predicted return, uncertainty, and risk are used when ranking candidates against the fallback. Below, reconstruction and rollout losses fit the world model.

reporting guidelines describe how such policies should be evaluated (Singer et al., 2016; Gottesman et al., 2019). Logged intensive-care studies then learn treatment policies from recorded trajectories, including model-based sepsis treatment (Komorowski et al., 2018; Raghu et al., 2017; 2018; Xu et al., 2025). BEHRT and Med-BERT pretrain on structured records, and LLaVA-Med aligns biomedical images with text (Li et al., 2020; Rasmy et al., 2021; Li et al., 2023). Our task keeps incomplete clinical, imaging, and molecular evidence in one belief, then compares authorized actions in a known-mechanism simulator.

2.2 Multimodal patient-state modeling

Missingness and irregular sampling are established problems in clinical time-series modeling. GRU-D models missingness in recurrent prediction, BRITS learns recurrent imputation, and latent ODEs describe continuous-time latent evolution (Che et al., 2018; Cao et al., 2018; Rubanova et al., 2019). Continuous-time encoders and chest-radiograph datasets extend this setting (Chen et al., 2018; Shukla & Marlin, 2021; Irvin et al., 2019). General-purpose and medical VLMs provide complementary image–text representations (Bai et al., 2025; Chen et al., 2024; Xu et al., 2026; Jiang et al., 2025). The filter updates this belief when evidence arrives and continues the forecast after observations stop. Experiments separate history retention, direct VLM forecasts, and task-trained recursive dynamics.

3 Method

PulmoWorld maintains one patient belief that integrates available observations, evolves when measurements stop, and branches under supplied care. The framework connects asynchronous estimation, shared-state simulation, and evidence-constrained planning (Figure 2).

3.1 Patient state and multimodal observations

We infer patient state from cohort-specific observations, $\mathcal{M}_{\text{MASS}} = \{\text{clinical}, \text{CT}, \text{BALF}\}$ and $\mathcal{M}_{\text{MIMIC}} = \{\text{clinical}, \text{CXR}\}$. Clinical inputs are numerical measurements; imaging and BALF metabolomics use precomputed representations. Let $\mathcal{H}_t = ((o_j, m_j, \tau_j, a_{j-1}))_{j=1}^{t}$ denote observed history, including masks, timestamps, and preceding care when available. The belief $B_t = q_\theta(Z_t \mid \mathcal{H}_t)$ represents a single latent patient state Z_t. Masks distinguish missing entries from measured zeros

without requiring all modalities to be observed together. The belief-transition model $\mathcal{T}_\theta(B' \mid B, \Delta t, a)$ advances this state over elapsed time, optionally conditioned on care a. An observation model decodes values, modality representations, and availability, supporting evidence assimilation and simulation from the same state.

3.2 Asynchronous multimodal belief learning

History-aware filtering combines the state predicted from past evidence with newly available observations. It first propagates B_{t-1} to the current timestamp, yielding $\widetilde{B}_t$, and then corrects this prior. For structured filtering, the patient belief has Gaussian components with means μ_t^c and diagonal variances. Here c indexes internal state components, not observation modalities; dataset-specific partitions are specified in Section B.1. Within each modality, observation groups v are encoded as $e_t^v = E_v(m_t^v \odot o_t^v, m_t^v)$ and update their assigned component $c(v)$:

$$\nu_t^v = e_t^v - D_v(\widetilde{\mu}_t^{c(v)}), \qquad \mu_t^c = \widetilde{\mu}_t^c + \sum_{v:c(v)=c} K_t^v \odot P_v(\nu_t^v). \tag{1}$$

Here ν_t^v is the observation innovation, and D_v and P_v map between state and token spaces. The learned correction gain K_t^v is zero when the group has no observed evidence. Missing observations therefore leave the propagated state intact, while available measurements update its mean and uncertainty. Belief variance increases through learned diffusion during propagation and contracts under measurement correction. Evidence reliability instead records freshness: observed evidence resets it to one, and it decays between updates. Thus, state uncertainty and evidence freshness serve distinct roles. Full variance, gain, and recency equations appear in Section B.3.

3.3 Shared-state dynamics and world-model fitting

During recursive simulation, each component combines local dynamics with messages from connected components. Supplied care branches this same transition through a state-dependent residual:

$$\widetilde{\mu}_{t+1}^c = \mu_t^c + \eta_c(\Delta t_t)\left[F_{\mathrm{loc}}^c(\mu_t^c, T(\Delta t_t)) + \sum_{j\to c} g_{j\to c,t} M_{j\to c}(\mu_t^j) + R_a^c(\mu_t^c, a_t, \Delta t_t)\right]. \tag{2}$$

T encodes $\Delta t_t = \tau_{t+1} - \tau_t$; η_c is a bounded time gain with $\eta_c(0) = 0$. Each $M_{j\to c}$ maps a source component to a message, weighted by g using state and evidence reliability. The care residual R_a^c is initialized to zero. Removing it recovers state-only prediction. After initialization, predicted beliefs feed subsequent transitions without assimilating factual intermediate observations. Decoded imaging outputs are representation forecasts rather than pixels.

World fitting grounds the source belief in observations and supervises its recursive evolution through

$$\mathcal{L}_{\mathrm{world}} = \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{dyn}}\mathcal{L}_{\mathrm{dyn}} + \lambda_{\mathrm{avail}}\mathcal{L}_{\mathrm{avail}} + \mathcal{L}_{\mathrm{aux}}. \tag{3}$$

Reconstruction and forecasting use Gaussian negative log-likelihoods on observed targets; availability uses binary cross-entropy. Auxiliary terms provide alignment and consistency supervision where enabled. Future observations supply targets, not intermediate rollout inputs. The curriculum trains reconstruction before rollouts; the history ablation also evaluates auxiliary reconstruction. MASS-Syn fits its transition ensemble using one-step latent squared error. After fitting and calibration, θ is frozen; weights and schedules appear in Sections B.2 and B.3.

3.4 Evidence-constrained planning

Cap-MBRL first determines which action comparisons have evidential support, then learns and selects within that authorized set. Predictive evidence G_A requires action-conditioned improvement with shuffling and support checks. Identification evidence G_I additionally checks action-effect accuracy and decision regret against a known mechanism. The authorized set $\mathcal{A}_{\mathrm{auth}}$ contains actions passing both checks; the actor is normalized only over this set. In MASS-Syn, one fixed action-channel certificate instantiates this set at every simulated state and retains the no-intervention reference. That certificate is an input to policy optimization, not an output learned by the actor. Audit tests and thresholds are specified in Section B.3.

Within this set, policy π_ϕ generates alternative continuations from the same initial state through a frozen three-member world ensemble. Its mean prediction advances the state; normalized disagreement supplies uncertainty u, distinct from the Gaussian belief variance. The score for continuation k is $S_{ik} = \sum_h \gamma^{h-1}(r_{ikh} - \kappa u_{ikh})$. Here i indexes initial states, h indexes rollout steps, γ discounts future rewards, and κ weights uncertainty. Reward is negative latent severity. Scores are centered and standardized within each initial-state group to form advantages $\widehat{A}_{ik}$. The actor update uses

$$\mathcal{L}_{\text{actor}} = -\mathbb{E}_{i,k}\left[\text{sg}(\widehat{A}_{ik})\sum_h \gamma^{h-1}\log\pi_\phi(a_{ikh} \mid B_{ikh})\right] + \lambda_{\text{KL}}\widehat{D}_{\text{KL}}(\pi_\phi, \mu_b) + \lambda_\rho \mathbb{E}_{i,k}\left[(\bar{\rho}_{ik} - \bar{b}_{ik})_+\right], \tag{4}$$

with a frozen behavior policy μ_b. The stop-gradient applies only to the advantage. The actor is categorical, so this term is a score-function gradient through $\log\pi_\phi$. $\bar{\rho}$ and $\bar{b}$ are discount-normalized risk and budget; their positive gap enters the same sampled objective, and the same budget screens the executed action.

At decision time, candidate first actions use the learned policy's highest-probability authorized continuations; the frozen fallback follows its own policy. For discounted severity-plus-risk cost $\widehat{C}$, discount-normalized risk $\widehat{\rho}$, calibrated budget $\widehat{b}$, and fixed margin $\delta \geq 0$, eligible replacements are $\mathcal{F}(B) = \{a \in \mathcal{A}_{\text{auth}} : \widehat{\rho}(B,a) \leq \widehat{b}(B),\ \widehat{C}(B,a) < \widehat{C}_{\text{fb}}(B) - \delta\}$. Given an authorized fallback first action a_{fb} with trajectory cost $\widehat{C}_{\text{fb}}$, execute

$$a_{\text{exec}}(B) = \begin{cases} \arg\min_{a\in\mathcal{F}(B)} \widehat{C}(B,a), & \mathcal{F}(B) \neq \varnothing, \\ a_{\text{fb}}, & \text{otherwise}, \end{cases} \tag{5}$$

If no authorized fallback is available, selection abstains. Risk combines ensemble disagreement, behavior surprisal, recursive–direct discrepancy, and capability diagnostics. The budget constrains replacements, not fallback risk. Only the first action is executed before replanning in MASS-Syn; full definitions appear in Sections B.3 and B.6.

4 Experiments

4.1 Experimental setup

Data and evaluation settings. We use two real-world cohorts and a separate known-mechanism simulator. The two cohorts are trained and evaluated separately. Public MIMIC-IV-CXR uses clinical measurements and chest radiographs from 65,355 patients (Johnson et al., 2023; 2019). Private MASS is a pneumonia cohort with clinical measurements, CT, and BALF metabolomics. Locked forecasting evaluates 300 pneumonia patients drawn from the 1,178 pneumonia cases in the 6,535-patient MIMIC test split, and 40 MASS patients. MIMIC uses 13 clinical variables and CXR representations. MASS uses Clinical19, with seven compartments for the clinical groups, CT, and BALF, for development and for the 40-patient locked forecast. Both locked forecasters use the history-aware belief. The MIMIC development comparison is against latest-value fusion (Section 4.2). MASS-Syn contains no patient records and tests planning in five known-mechanism environments. Splits are patient-disjoint and transformations are fitted on training data; cohort-specific protocols are in Sections A and B.1.

Comparators. Forecasting controls are persistence and the training-set median. Task-trained comparators include GRU, a causal Transformer, GRU-D, independently adapted CLARITY, and a separately trained Qwen2.5-VL-7B SFT model. General-purpose and medical Direct VLMs receive the cohort's numerical input and output format. Table 2 reports the full MIMIC comparison; the complete MASS roster is in Table 10. Policy comparisons separately test unrestricted optimization, uncertainty/support penalties, authorization alone, and risk-budgeted improvement.

Evaluation. History retention is assessed under a frozen mortality-risk probe using AUROC, AUPRC, and Brier score, with accuracy and balanced accuracy as descriptive metrics. Recursive forecasts advance H1–H4 without factual intermediate observations. Macro NMAE averages errors normalized by training IQR; Δ-NMAE measures error in consecutive changes, exposing trajectories that miss temporal variation despite plausible endpoint values. We also report H4 direction agreement. Patient-level bootstrap 95% intervals quantify uncertainty. Newly reported paired history, MIMIC

forecast, and prognosis contrasts use 5,000 resamples, with outcome stratification for classification and fixed fitted models. History metrics are computed per seed and then averaged. Planning regret uses simulator-scored discounted cost relative to the best authorized action sequence. That sequence is used only to score regret. MASS-Syn certificates use the known mechanism as the action-effect reference and are not a result about real MASS or MIMIC. Metric definitions and task-specific uncertainty protocols are in Section C.1.

4.2 Patient-state estimation from incomplete histories

We first compare asynchronous history filtering with latest-value fusion on 3,758 MIMIC development patients (524 in-hospital deaths), using three seeds and one frozen window per patient. Training both inputs with and without auxiliary masked-source reconstruction tests whether history contributes information beyond the training objective.

Table 1: **Asynchronous history filtering improves prognostic prediction.** MIMIC development, three-seed means on 3,758 patients (524 deaths). The asynchronous filter retains ordered history; fusion uses latest values. Bold marks the better configuration within each reconstruction setting.

State estimation	AUROC ↑	AUPRC ↑	ACC ↑	Bal. ACC ↑	Brier ↓
Without auxiliary reconstruction					
Asynchronous filtering	**0.7565**	**0.3737**	**0.8481**	**0.6092**	**0.1064**
Latest-value fusion	0.7352	0.3538	0.8454	0.6061	0.1079
With auxiliary reconstruction					
Asynchronous filtering	**0.7748**	**0.4326**	**0.8673**	**0.6311**	**0.0965**
Latest-value fusion	0.7367	0.3548	0.8427	0.6082	0.1075

ACC uses the frozen probe threshold of 0.23; Bal. ACC averages sensitivity and specificity. Majority-class ACC is 0.8606, so thresholded accuracy is descriptive, not the primary evidence of prognostic utility.

Asynchronous filtering improved AUROC over latest-value fusion by 0.0213 (95% paired CI $[0.0132, 0.0295]$) without reconstruction (Table 1). With reconstruction, the AUROC, AUPRC, and Brier changes against latest-value fusion are $+0.0381$, $+0.0778$, and -0.0110. The combined comparison performed best among these four cells. The 3,758 patients are the MIMIC development set: 13 clinical variables and CXR representations, one frozen window per patient. The contrast is latest-value fusion. MASS is trained separately, so its clinical, CT, and BALF forecasts are outside the comparison reported here.

We also evaluate source completion under missing evidence (Table 12). This frozen history-aware MIMIC checkpoint fills 718 masked values across 271 requests (macro NMAE 0.834); the locked forecast tables are separate three-seed means. In development, clinical RMSE rose from 0.7688 to 0.9504 at 75% source dropout (Tables 11 and 14).

4.3 Recursive and treatment-conditioned simulation

We next evaluate recursive forecasts initialized from observed evidence. Table 2 reports the full MIMIC comparison, and Figure 3a,b summarizes the MASS clinical rollouts. On MIMIC, PULMOWORLD reached H4 NMAE 0.5780, versus 0.5934 for GRU and 0.5954 for the causal Transformer, although the Transformer and GRU had lower H1 and H2 errors, respectively. Relative to adapted CLARITY, H4 NMAE decreased by 0.0548 (95% paired CI $[0.0331, 0.0757]$). On MASS, mean H1–H4 NMAE was 0.557, compared with 0.570 for GRU-D, with the remaining task-trained comparisons shown in Figure 3a. GRU-D remained close at H4 (0.5442 versus 0.5418 for PULMOWORLD). MASS consecutive-change error was 0.498, compared with 0.503 for GRU-D (Figure 3b).

Multimodal evolution. Beyond clinical measurements, MASS development rollouts predicted CT and BALF representations (Figure 3c,d). H1 MSE was 0.836 for both modalities, versus 0.868 for CT and 0.889 for BALF with GRU (Table 21). Together with clinical rollouts, these results show that the model simulates multiple aspects of the patient state. Imaging outputs are representation forecasts; the observed images in the case displays provide visual context.

Table 2: **MIMIC-IV-CXR locked-test longitudinal simulation.** H1–H4: macro NMAE; Δ-NMAE: consecutive-change error with patient-bootstrap 95% CI; H4 dir.: direction agreement. Bold/underline denote best and second-best task-trained result in each column.

MIMIC-IV-CXR pneumonia locked test (300 patients)

Method	H1↓	H2↓	H3↓	H4↓	H4 EHR-S@15↑	Δ-NMAE↓	H4 dir.↑
Controls							
Persistence	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
Train median	0.7188	0.7009	0.6839	0.7090	48.8%	0.471 [0.454,0.489]	59.6%
General-purpose Direct VLMs							
Qwen2.5-VL-3B†	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
Qwen2.5-VL-7B	2.0176	2.8403	2.1843	1.8233	44.7%	0.682 [0.572,0.795]	56.7%
Qwen2.5-VL-72B	0.9069	2.2291	0.8848	3.0042	44.0%	0.774 [0.446,1.410]	57.0%
Qwen3-VL-8B-Instruct	4.5230	5.5157	6.1501	6.2845	46.0%	0.467 [0.449,0.486]	57.0%
Qwen3-VL-30B-A3B	33.2483	35.0294	40.3677	25.2325	45.2%	0.581 [0.469,0.754]	56.1%
InternVL3-8B†	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
LLaVA-OV-1.5-8B	2.4461	1.0740	3.4511	1.2832	38.3%	0.644 [0.564,0.742]	55.5%
Medical Direct VLMs							
Lingshu-7B	1.0245	1.1025	1.1310	1.1103	46.9%	0.458 [0.430,0.500]	56.8%
Hulu-Med-7B	24.1641	11.6647	14.1120	9.8788	47.0%	0.897 [0.564,1.500]	57.7%
Hulu-Med-32B	54.5972	34.4492	45.6867	22.0711	47.8%	0.588 [0.447,0.845]	57.6%
Task-trained models: temporal, adapted VLM, and world model							
GRU	0.5068	**0.5483**	0.5818	0.5934	52.1%	0.412 [0.396,0.428]	60.4%
Causal Transformer	**0.5005**	0.5527	0.5877	0.5954	52.5%	0.411 [0.395,0.428]	60.1%
GRU-D	0.5238	0.6815	0.8418	0.9600	38.3%	0.444 [0.429,0.461]	58.1%
Qwen2.5-VL-7B SFT	0.6484	1.6620	0.7182	0.7221	49.5%	0.875 [0.589,1.415]	56.4%
CLARITY (adapted)¶	0.5110	0.5746	0.6129	0.6328	50.8%	0.531 [0.513,0.550]	61.4%
Ours							
PULMOWORLD	0.5034	0.5489	**0.5722**	**0.5780**	**54.9%**	**0.407 [0.392,0.423]**	**61.7%**

¶Independently adapted CLARITY. †This Direct VLM row repeats the persistence control on every cell. EHR-S@15: feature-macro success within 15% relative error, with denominator $\max(|y|, 0.1\,\mathrm{IQR}_{\mathrm{train}})$; a compatibility score. Detailed handling: Appendix C.2.

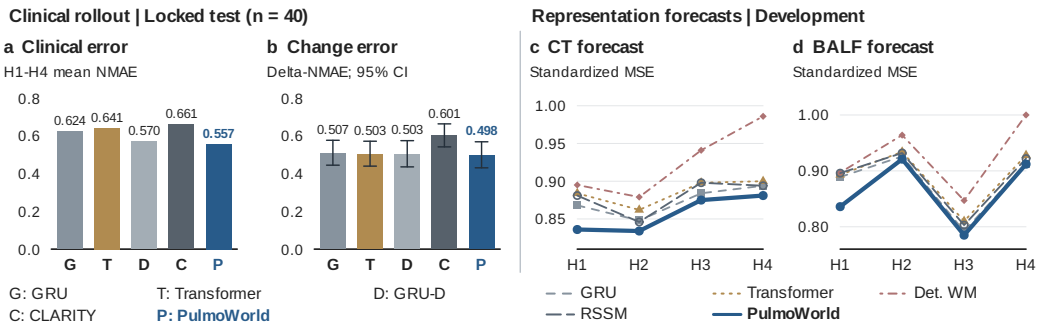


Figure 3: **Recursive forecasts of clinical measurements and multimodal representations on MASS.** (a,b) Clinical19 locked-test mean H1–H4 NMAE and change error ($n = 40$; patient-bootstrap 95% CIs in b). (c,d) Clinical19 development CT/BALF representation MSE, not the 40-patient locked forecast (three-seed means; non-zero axis origins). Blue denotes PULMOWORLD; lower is better. Complete scores are in Tables 10 and 21.

Evolution in individual cases. Figure 4 complements aggregate errors with observed CXR sequences, representation forecasts, and separate clinical rollouts. The examples include declining measurements and late changes missed by rollouts.

Reconstruction-to-rollout training. A matched MIMIC development study replaced the first three of twelve rollout-training epochs with reconstruction, keeping the optimizer-step budget fixed. The paired RMSE reduction was 0.00371 [0.00196, 0.00535] (Table 13).

Information in simulated future states. One frozen outcome probe per cohort tests whether simulated future states retain prognostic information about in-hospital death on MIMIC and 28-day death on MASS. Compared with current-state inputs, future-state summaries give AUROC 0.7135 versus 0.7106 on MIMIC and 0.7186 versus 0.6856 on MASS; AUPRC is 0.3237 versus 0.3022 and

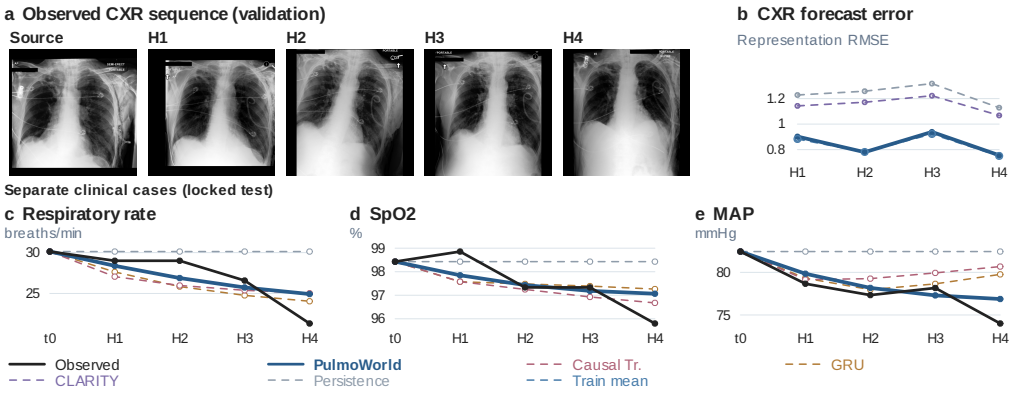


Figure 4: **Observed imaging evolution and recursive clinical forecasts.** (a) Observed MIMIC validation CXRs, not generated images or future rollout inputs. H1–H4 denote six-hour windows, not acquisition times. (b) Representation RMSE for the same sequence. (c–e) Separate, outcome-selected locked-test clinical rollouts. Selection and image timing are in Section D.6 and Figure 12.

0.5468 versus 0.4263, respectively. MIMIC uses these 300 pneumonia test patients at H4; MASS uses a 99-patient validation panel at available H2–H4 endpoints (Table 3). The shared probe scores source observations together with endpoint changes.

Table 3: **Cross-dataset prognosis metrics from predicted future states.** One frozen probe per cohort: MIMIC pneumonia locked test ($n = 300$) and MASS frozen validation ($n = 99$). MIMIC uses H4; MASS uses available endpoints through H4. ACC uses frozen calibration-selected thresholds. Bold and underline mark the best and second-best task-trained result in each column.

	MIMIC locked test: in-hospital death				MASS frozen validation: 28-day death			
Method	ACC↑	AUROC↑	AUPRC↑	Brier↓	ACC↑	AUROC↑	AUPRC↑	Brier↓
Controls								
Current state	0.8400	0.7106	0.3022	0.0899	0.6162	0.6856	0.4263	0.2051
Persistence	0.5500	0.6244	0.2139	0.1299	0.5960	0.6841	0.4254	0.2100
Train median	0.4200	0.5635	0.1947	0.1220	0.6263	0.6408	0.3776	0.2070
General-purpose Direct VLMs								
Qwen2.5-VL-3B†	0.8400	0.7106	0.3022	0.0899	0.6162	0.6856	0.4263	0.2051
Qwen2.5-VL-7B	0.6000	0.6179	0.1740	0.1765	0.6162	0.6851	0.4237	0.2052
Qwen2.5-VL-72B	0.6700	0.6817	0.2457	0.1229	0.6162	0.6987	0.4463	0.2056
Qwen3-VL-8B-Instruct	0.6133	0.6343	0.1597	0.1778	0.5960	0.6585	0.3613	0.2425
Qwen3-VL-30B-A3B	0.7400	0.6032	0.2429	0.1214	0.6061	0.6796	0.4149	0.2129
InternVL3-8B†	0.8400	0.7106	0.3022	0.0899	0.6162	0.6856	0.4263	0.2051
LLaVA-OV-1.5-8B	0.4667	0.6249	0.1988	0.2069	0.6566	0.7057	0.4471	0.2088
Medical Direct VLMs								
Lingshu-7B‡	0.8400	0.7106	0.3022	0.0899	0.6364	0.6650	0.3822	0.2129
Hulu-Med-7B	0.7567	0.6582	0.1651	0.1631	0.5960	0.6730	0.4216	0.2073
Hulu-Med-32B	0.7600	0.6587	0.2294	0.1113	0.6364	0.5986	0.3790	0.2368
Task-trained models: temporal, adapted VLM, and world model								
GRU	0.5767	0.5977	0.2381	0.1201	0.6229	0.6442	0.3931	0.2385
Causal Transformer	0.5967	0.6059	0.2494	0.1147	0.5993	0.6524	0.3992	0.2439
GRU-D	0.6200	0.6382	0.2285	0.1342	0.6061	0.6715	0.3828	0.2146
Qwen2.5-VL-7B SFT	0.8467	0.7003	0.2861	0.0906	0.6162	0.6856	0.4263	0.2051
CLARITY (adapted)	0.5333	0.6040	0.2433	0.1273	0.6465	0.6932	0.4636	0.1907
Ours								
PULMOWORLD	**0.8500**	**0.7135**	**0.3237**	**0.0867**	**0.6768**	**0.7186**	**0.5468**	**0.1361**

†This Direct VLM row repeats the current-state control on every cell. ‡The MIMIC cells repeat the current-state control; the MASS cells do not. Cohort-specific fitting details: Appendices C.2 and C.6.

Conditioning on recorded care. Logged fluid–vasopressor inputs reduced MIMIC validation RMSE by 0.61% [0.45, 0.80] on action-exposed windows and 0.20% overall. Four train-derived care templates produce distinct MAP, lactate, and urine-output trajectories from the same state (Figure 5a–c). Selection among those continuations is then tested in MASS-Syn.

4.4 EVIDENCE-CONSTRAINED PLANNING

In five known-mechanism MASS-Syn environments, the fallback, the soft-penalty policy, and Cap-MBRL share one authorized action set (Table 4). That set is one action-channel certificate and stays fixed at every simulated state. Ungated controls still select actions outside that authorized set. The unauthorized-action rate in Table 4 counts executed actions outside the permitted set, so a zero rate means the mask was respected. Selection regret is the gap to the best authorized action. Cap-MBRL applies the risk budget and the predicted-cost margin, and this replacement of direct fallback execution moves regret from 0.091 to 0.018. The authorized soft-penalty policy is a separate objective inside the same set and remains at 0.3636 (Figure 5d,e). The upper rows of Table 4 execute outside that certificate, so their selection-regret cells are blank. The lower block keeps unauthorized rates at zero, with regrets 0.3636, 0.0910, and 0.0180.

Table 4: **Policy controls for evidence-constrained planning.** Five-environment MASS-Syn means (lower is better). Unauthorized-action rate is the rate of executed actions outside the permitted set. Selection regret is the gap to the best action inside that set.

Policy control	Unauth. (%)	Sel. regret
Unrestricted WM-RL	32.19	–
Uncertainty-aware	32.26	–
Support-aware	33.71	–
Path-consistency weighting	31.58	–
Same authorized action set		
Authorized + soft penalties	0.00	0.3636
Authorized policy (fallback)	0.00	0.0910
Cap-MBRL	0.00	**0.0180**

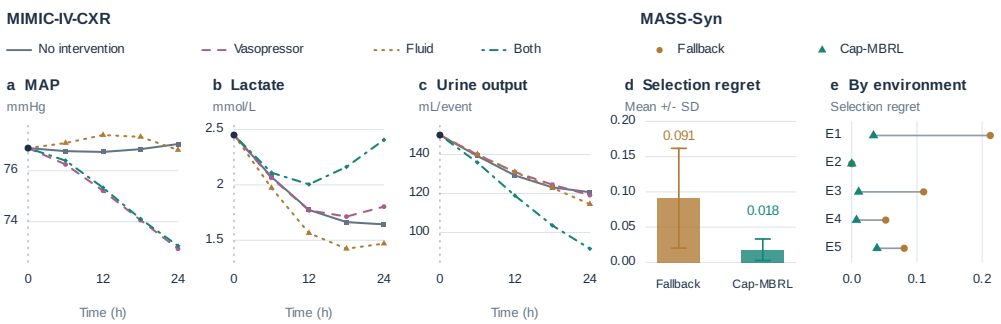


Figure 5: **Conditional trajectories and authorized-set selection.** (a–c) Conditional MIMIC validation forecasts under four six-hour care templates; zero fluid and vasopressor inputs denote "No intervention" (Section C.9). (d,e) MASS-Syn selection regret: five-environment mean ± population SD and paired environment results.

5 CONCLUSION

PULMOWORLD is a world model for longitudinal pneumonia dynamics. It keeps incomplete clinical, imaging, and BALF evidence in one belief, rolls that belief forward, and compares actions only inside an authorized set. The two cohorts are trained and evaluated separately. On MASS, recursive forecasts are competitive with task-trained sequence models, and simulated future states retain prognostic information about 28-day death. On MIMIC-IV-CXR, history-aware filtering improves prognostic discrimination over latest-value fusion in development. Locked evaluation uses 300 pneumonia patients drawn from the 1,178 pneumonia cases in the 6,535-patient test split. On that panel, the forecaster is competitive with task-trained models on long-horizon error, and simulated states retain information about in-hospital death. In five known-mechanism MASS-Syn environments, Cap-MBRL lowers selection regret from 0.091 under direct execution of the frozen fallback to 0.018.

AI USE STATEMENT

AI assistants supported manuscript drafting and language polishing, research feedback, and code development and checking. The authors reviewed the assisted work and take full responsibility for the accuracy, originality, and integrity of the manuscript.

ETHICS STATEMENT

This study uses retrospective data and involves no intervention in patient care. MASS records were used with permission after personal identifying information was removed. MIMIC-IV and MIMIC-CXR were accessed under the PhysioNet data-use agreement (Johnson et al., 2023; 2019).

REPRODUCIBILITY STATEMENT

We will open-source the training and evaluation code, preprocessing procedures, experiment configurations, and synthetic environments to support reproducibility. The appendix details data construction, model specifications, and experimental protocols. Patient-level MASS records are excluded from the release, and MIMIC data remain subject to their original access conditions.

REFERENCES

Shuai Bai, Keqin Chen, Xuejing Liu, et al. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31, 2018.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. volume 31, 2018.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24185–24198, 2024.

Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. volume 31, 2018.

Tianxingjian Ding, Yuanhao Zou, Chen Chen, Mubarak Shah, and Yu Tian. Clarity: Medical world model for guiding treatment decisions by modeling context-aware disease trajectories in latent space. *arXiv preprint arXiv:2512.08029*, 2025.

Omer Gottesman, Fredrik Johansson, Matthieu Komorowski, Aldo Faisal, David Sontag, Finale Doshi-Velez, and Leo Anthony Celi. Guidelines for reinforcement learning in healthcare. *Nature medicine*, 25(1):16–18, 2019.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2(3):440, 2018.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019a.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International conference on machine learning*, pp. 2555–2565. PMLR, 2019b.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, 2025.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *AAAI*, 2019.

Songtao Jiang, Yuan Wang, Sibo Song, Tianxiang Hu, Chenyi Zhou, Bin Pu, Yan Zhang, Zhibo Yang, Yang Feng, Joey Tianyi Zhou, et al. Hulu-med: A transparent generalist model towards holistic medical vision-language understanding. *arXiv preprint arXiv:2510.08668*, 2025.

Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317, 2019.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1, 2023.

Matthieu Komorowski, Leo A Celi, Omar Badawi, Anthony C Gordon, and A Aldo Faisal. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature medicine*, 24(11):1716–1720, 2018.

Yann LeCun et al. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1):1–62, 2022.

Chunyuan Li, Cliff Wong, Sheng Zhang, et al. LLaVA-Med: Training a large language-and-vision assistant for biomedicine in one day. In *NeurIPS*, 2023.

Yikuan Li, Shishir Rao, Jose Roberto Ayala Solares, et al. BEHRT: Transformer for electronic health records. *Scientific Reports*, 10, 2020. doi: 10.1038/s41598-020-62922-y.

Joshua P. Metlay, Grant W. Waterer, Ann C. Long, et al. Diagnosis and treatment of adults with community-acquired pneumonia. an official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *American Journal of Respiratory and Critical Care Medicine*, 200:e45–e67, 2019. doi: 10.1164/rccm.201908-1581ST.

Linjie Mu, Zhongzhen Huang, Yannian Gu, Shengqian Qin, Shaoting Zhang, and Xiaofan Zhang. Ehrworld: A patient-centric medical world model for long-horizon clinical trajectories. *arXiv preprint arXiv:2602.03569*, 2026.

Aniruddh Raghu, Matthieu Komorowski, Imran Ahmed, et al. Deep reinforcement learning for sepsis treatment. *arXiv preprint arXiv:1711.09602*, 2017.

Aniruddh Raghu, Matthieu Komorowski, and Sumeetpal Singh. Model-based reinforcement learning for sepsis treatment. *arXiv preprint arXiv:1811.09602*, 2018.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*, 4, 2021. doi: 10.1038/s41746-021-00455-y.

Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. *Advances in neural information processing systems*, 32, 2019.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

Satya Narayan Shukla and Benjamin M Marlin. Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations*, volume 2021, pp. 14897, 2021.

Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, 315:801–810, 2016. doi: 10.1001/jama.2016.0287.

Jiangyuan Wang, Xuyong Chen, Junwei He, Xu Xu, Shasha Xie, and Fuman Han. Chronomedical-world: A medical world model for learning patient trajectories from longitudinal care data. *arXiv preprint arXiv:2605.21963*, 2026.

Minghao Wu, Yuting Yan, Zhenyang Cai, Ke Ji, Chuangsen Fang, Ziying Sheng, Xidong Wang, Rongsheng Wang, Hejia Zhang, Shuang Li, et al. Agentifying patient dynamics within llms through interacting with clinical world model. *arXiv preprint arXiv:2605.14723*, 2026.

Qianyi Xu, Gousia Habib, Feng Wu, et al. medDreamer: Model-based reinforcement learning with latent imagination on complex EHRs for clinical decision support. *arXiv preprint arXiv:2505.19785*, 2025.

Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, et al. Lingshu: Generalist foundation model for unified multimodal medical understanding and reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Jiangtao Yan, Yanlin Qu, Yansheng Qiu, Shujian Gao, Wei Yu, Zheng Wang, Xiaodong Sun, Huixun Jia, and Diping Song. Orbit: A prognostic world model for ocular reasoning based on imagined trajectories. In *ICML*, 2026.

Yijun Yang, Zhao-Yang Wang, Qiuping Liu, Shuwen Sun, Kang Wang, Rama Chellappa, Zongwei Zhou, Alan Yuille, Lei Zhu, Yu-Dong Zhang, et al. Medical world model: Generative simulation of tumor evolution for treatment planning. *arXiv preprint arXiv:2506.02327*, 2025.

Tianhe Yu, Garrett Thomas, Lantao Yu, Stefano Ermon, James Y Zou, Sergey Levine, Chelsea Finn, and Tengyu Ma. Mopo: Model-based offline policy optimization. *Advances in neural information processing systems*, 33:14129–14142, 2020.

Tianhe Yu, Aviral Kumar, Rafael Rafailov, Aravind Rajeswaran, Sergey Levine, and Chelsea Finn. Combo: Conservative offline model-based policy optimization. *Advances in neural information processing systems*, 34:28954–28967, 2021.

APPENDIX CONTENTS

The supplementary material develops the evidence behind history-aware state estimation, multimodal forecasting, and evidence-constrained action selection. Sections A and B define the data and models; Section C presents full comparisons and component tests; Section D connects predictions to observed patient records; Section E specifies reproducibility and access conditions. Locked-test results, development studies, and synthetic experiments are identified separately throughout. Titles and page numbers below are linked.

	Contents	Page
A	**Data construction, distributions, and privacy**	15
	Public MIMIC, private MASS, and synthetic MASS-Syn	15
	MASS event construction and evaluation populations	15
	MIMIC alignment, aggregation, and patient splits	15
	Patient proportions, modality coverage, event/window counts	16
	Observed-value masks, reconstruction targets, and leakage control	16
	Data permissions and privacy	19
B	**Model specification, training, and evaluation**	19
	Model families, tensor dimensions, parameter counts	19
	Training phases and checkpoint selection	20
	Belief updates, authorization, and full policy equations	21
	Dataset-specific execution and action interfaces	26
	Prognosis comparator and outcome probe	26
	Policy fitting and decision-time search	26
C	**Complete quantitative results**	28
	Full MASS and MIMIC locked-test forecasting tables	29
	Future-state prognosis and representation forecasts	30
	Masked source reconstruction (T1)	30
	Component ablations and missing-evidence sensitivity	32
	History-enabled MIMIC prognosis follow-up	32
	Task-adapted VLM prognosis evaluation	33
	MIMIC action conditioning and candidate ranking	33
	Known-mechanism policy selection and gate diagnostics	36
	Action-conditioned branches and authorization evidence	37
D	**Case atlas: inputs, future states, and failure modes**	38
	Matched source/next-state records and complete input fields	38
	Same-patient multimodal inputs, complete outputs, and masks	40
	Diverse CT pairs and CXR/action input examples	42
	Clinical case selection and image-state comparison	43
	Longitudinal imaging and metabolomic evolution	44
E	**Reproducibility and data access**	49

A Data Construction, Splits, Ethics, and Privacy

A.1 Public MIMIC, private MASS, and synthetic selection

The paper uses two real cohorts and one simulator. They differ in access, clinical scope, and which claim they can support.

MIMIC-IV-CXR is assembled from public releases: MIMIC-IV clinical records (Johnson et al., 2023) and MIMIC-CXR chest radiographs (Johnson et al., 2019). Both are available from PhysioNet under a credentialed data-use agreement. We do not redistribute the source files. The linked records supply chest radiographs and clinical trajectories for training and evaluation. They support recursive forecasting, history retention, and logged fluid–vasopressor conditioning.

MASS is a private institutional cohort. It links pneumonia-related clinical measurements, chest CT, and BALF metabolomics. Patient-level records are not publicly released. MASS supports the pneumonia-specific representation and locked Clinical19 forecast. External repetition of those patient-level analyses requires controlled access, which is separate from the public code and aggregate-metric release (Section E).

MASS-Syn contains no patient records. It is a known-mechanism simulator and is used only to score authorized-set selection.

Throughout this appendix we keep three counting units separate: the source cohort, the model-specific data view, and the eligible evaluation population. One patient can contribute multiple events and forecast windows, so those record counts are not independent sample sizes. Figure 6 maps construction choices to evaluation populations. Figure 7 reports split proportions and observation coverage.

A.2 MASS construction and evaluation populations

MASS is the private pneumonia cohort (Section A.1). It begins with patient-linked clinical records, CT studies, BALF metabolomic profiles, and timestamped medication or organ-support intervals. Clinical values and their observation masks are aligned to the irregular event clock at recorded day-level resolution. Adjacent events form factual transitions. Interval actions describe exposure between source and target. A missing CT or BALF record does not create a synthetic observation. Clinical19 is the interface for development and for the 40-patient locked forecast. The counts below are this Clinical19 cohort.

Source linkage and record eligibility. The recorded clinical source contains 766 patients. Patient identifiers link modalities before the 83-patient locked partition is excluded, leaving 683 development patients. Clinical, CT, and BALF records need not coincide in time or be available for every patient. The 101 omics records lacking verified enrollment linkage are omitted at the record level, not counted as 101 patient exclusions. These counts describe the available source extract, not an institutional recruitment or diagnostic-screening denominator.

Clinical19 evaluation populations. The development extract contains 3,252 events and 2,569 adjacent transitions; the latter split into 1,806/428/335 train/validation/calibration transitions. Clinical19 state rows on this split number 2,608. Eligible rollout populations contain 465/104/85 patients and 1,971/461/368 windows in the same split order. The locked Clinical19 comparison uses 40 eligible patients and 40 four-step windows from the separate held-out population.

A.3 MIMIC-IV-CXR construction and temporal alignment

This section describes our alignment of the public MIMIC-IV and MIMIC-CXR releases (Section A.1). The 65,355-patient extract supplies the trajectories used for training and evaluation, including stays without a linked radiograph. Eligible chart, laboratory, and output records are linked to ICU stays. Implausible values are excluded by the frozen variable rules. Measurements are assigned to 6-hour bins relative to ICU admission, and repeated eligible values of a feature within a bin are averaged. Radiographs are linked to the same patient, stay, and time bin. The union of clinical and radiograph bins defines the event table; masks retain which features and images were observed. H4 examples

contain one source and four ordered future events from one stay. Elapsed gaps come from event times rather than fabricated intermediate states.

Final patient allocation targets 70/10/10/10% for train/calibration/validation/test. Official CXR validation and test patients retain those holdouts; remaining patients are assigned by a deterministic SHA-256 order to fill the quotas. This world-model partition is distinct from simply using the original image-file split. The locked recursive forecasting comparison uses 300 pneumonia patients drawn from the 1,178 pneumonia cases in the 6,535-patient test partition, after freezing the evaluation protocol.

Forecast windows and care intervals. A forecast window is eligible only within one ICU stay, with a source event and four ordered future events. A six-hour bin width does not require consecutive observed events to be exactly six hours apart. Clinical targets are scored only where their feature masks indicate an observed value. Linked CXR availability is retained separately from clinical completeness; patients without a radiograph remain in the clinical cohort. Fluid and vasopressor exposures describe source-to-target care intervals, rather than future clinical measurements supplied to the forecast.

A.4 Observation and split distributions

Evaluation interfaces. Present-state reconstruction recovers masked observed values; recursive forecasting predicts H1–H4 without factual intermediate observations. Future-state prognosis applies a frozen cohort-specific outcome probe. Action evaluation separates logged-care prediction from authorization and known-mechanism policy selection. MASS uses Clinical19. MIMIC uses 13 clinical variables and CXR. The two cohorts are not pooled.

Development population and eligible modeling records. The counts below itemize the private MASS development extract already defined in Section A.2. The test-excluded development view contains 683 patients, 3,252 timestamped events, 13,039 modality observations, 3,409 interval action records, and 2,569 adjacent factual transitions. Patient-level train/validation/calibration sizes are 485/106/92. The 83 locked-test patients are physically excluded from development evaluation. Clinical, CT, and BALF source observations cover 683/197/426 patients; 101 omics records without verified enrollment mapping are omitted.

The Clinical19 development model re-extracts clinical state from the same patient-disjoint cohort. The frozen state table contains 2,608 event rows and 23,724 observed feature rows: 1,853/413/342 state rows and 16,833/3,751/3,140 observations for train/validation/calibration. Nineteen clinical features are used; ten have eligible repeated-measurement targets and nine are baseline context. Grouping, units, transforms, and eligibility are frozen from train and calibration before validation.

MIMIC action construction aligns 1,150,760 adjacent transitions. Train-only cut points define four fluid and four vasopressor levels, yielding 16 supported cells (minimum train support 200). The no-recorded-action cell contains 607,910 of 802,500 train transitions (75.8%). Source extraction retains 163,213 qualifying crystalloid events and 735,644 convertible vasopressor events; excluded or unaligned records are logged rather than silently assigned.

CT embeddings are computed from a representative 2D slice by a frozen Qwen2.5-VL-7B and reduced from 3,584 to 64 dimensions by train-only PCA. BALF metabolomics is represented by a train-only 64-dimensional projection. No CT pixel is decoded. The source action ontology maps 137 medication names and organ-support records into 20 macro classes. The MASS world model evaluated in this paper keeps this ontology for auditing and disables its care-context residual. Releasing that ontology does not release the private records.

A.5 Missingness, transforms, and leakage control

Clinical transforms subtract the train median and divide by the train interquartile range, with a unit fallback only for zero or non-finite IQR. Missing clinical entries use the train median solely as a finite storage value, yielding zero after normalization while the Boolean mask remains false. The encoder receives both zero-masked values and feature-level masks, so an observed median and an unobserved value remain distinguishable. Entirely absent modalities have zero correction gain;



Figure 6: **From cohort records to model-specific evaluation populations.** (a,b) Private MASS and public MIMIC differ in clinical scope, observation clock, and modality availability. Patient splits precede model fitting; numerical transforms and projections are train-fit. (c) Forecasting and prognosis use distinct eligibility rules and must retain their own denominators. MASS prognosis uses the available endpoint through H4, not uniformly H4. Counts describe frozen data extracts, not recruitment estimates.

Table 5: **Clinical19 development data view.** The 83-patient locked test is physically excluded; the 40-patient Clinical19 recursive forecasting evaluation is reported separately. Development splits are patient-disjoint.

Quantity	Value
Development patients	683
Patient split (train/validation/calibration)	485 / 106 / 92
Events / observations / factual transitions	3,252 / 13,039 / 2,569
Transition split (train/validation/calibration)	1,806 / 428 / 335
Patients with clinical / CT / BALF	683 / 197 / 426
Clinical / CT / BALF observations	11,583 / 567 / 889
Clinical / CT latent / metabolomics dimensions	19 / 64 / 64
Actions	20 macro classes $\times$ 4 interval channels (80-D)
Retrospective / strict-online usable transitions	2,569 / 285
Locked-test patients included	0 of 83

elapsed-time dynamics still advance the prior state. Observation-value losses and evaluation use only observed targets, whereas the availability head is supervised by both observed and missing target masks. Missing CT or BALF remains masked rather than synthesized. Versioned physiologic bounds quarantine out-of-range records rather than repair them.

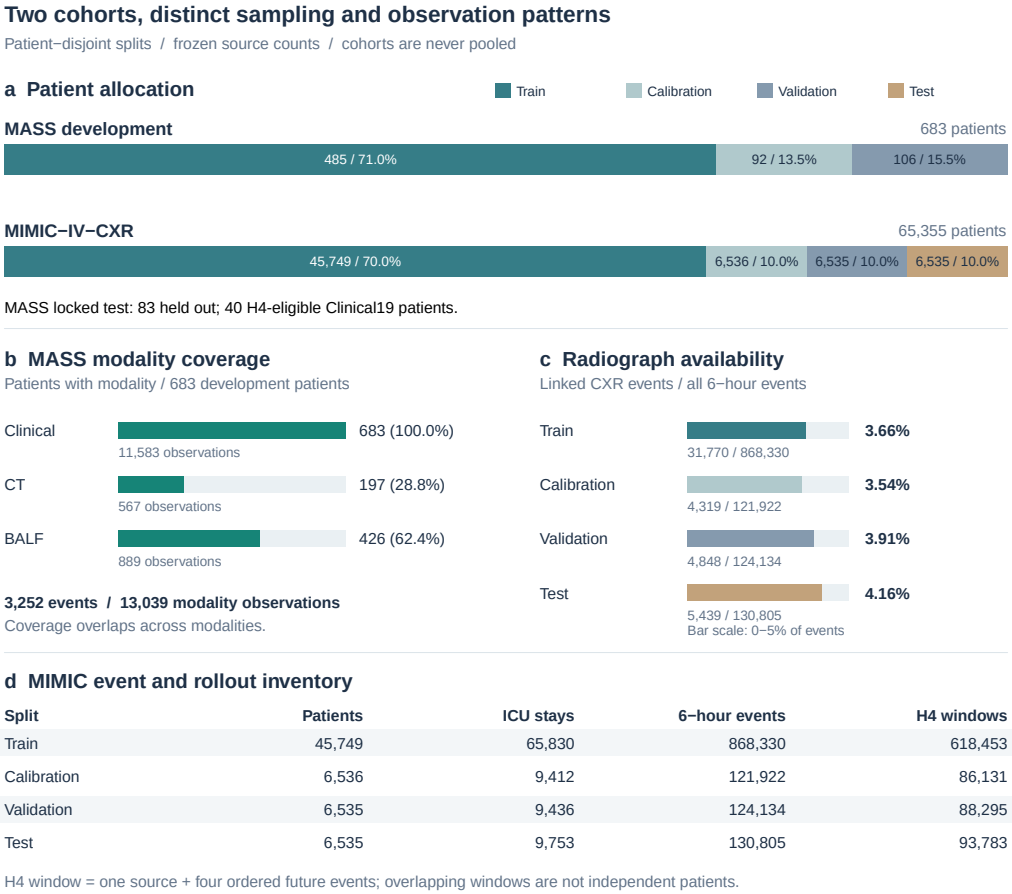


Figure 7: **Cohort composition and observation availability.** (a) Patient counts and within-cohort split proportions; MASS development excludes the 83 locked-test patients. The Clinical19 recursive forecasting panel contains 40 H4-eligible patients and 40 evaluated windows. (b) MASS patient-level modality coverage, with observation counts shown separately. Its 567 CT observations occur on 3,252 irregular events (17.4%). (c) MIMIC linked-radiograph counts divided by all 6-hour events in each split. Patient coverage in (b) and event availability in (c) have different denominators. (d) Exact MIMIC patient, stay, event, and H4-window counts. An H4 window contains one source and four ordered future events; overlapping windows are not independent patients.

Observed, absent, and deliberately hidden values. An observed median has normalized value zero and mask one; a missing value has the same storage value but mask zero. Only the former supplies an observed-value target. During source-reconstruction pretraining, a subset of observed entries is deliberately hidden from the encoder while its original values remain reconstruction targets. This artificial training mask is distinct from natural missingness. At inference, missing-evidence perturbations remove available inputs without converting genuinely absent values into targets (Table 14). Thus the mask specifies which evidence is available, not merely a numerical imputation convention.

Preprocessing statistics, projection parameters, action-cell definitions, propensity models, and 28-day outcome-probe hyperparameters are train-fit or calibration-selected without test labels. The later Clinical19 and MIMIC training phases use calibration for checkpoint selection and validation for reporting. Locked MASS forecasts use those Clinical19 checkpoints. Bootstrap resampling uses patients as units, retaining all observations from a sampled patient.

Table 6: **Patient flow into the development population.** Record-level quarantine rows are not unique-patient exclusions. Raw screening and institutional recruitment denominators remain unavailable.

Flow step	Remaining	Removed	Evidence status
Recorded clinical source cohort	766	–	Source extract
Authoritative patient split assigned	766	0	Exact set match
Locked-test patients physically excluded	683	83	Test closed
Test-excluded development cohort	683	0	Exact non-test set
Patients with at least one event	683	0	Complete coverage

A.6 Data permissions and privacy

MASS is a private institutional cohort used with permission for retrospective research. Personal identifying information was removed from the research data. MIMIC-IV and MIMIC-CXR are accessed under the PhysioNet data-use agreement. These permissions do not authorize redistribution of patient-level records.

The planned public release contains evaluation code, preprocessing specifications, synthetic mechanisms, and aggregate metrics. It excludes patient-level MASS records and does not repackage MIMIC source files. Released tables exclude direct or hashed patient identifiers, dates, DICOM identifiers, and free-text medication names. Image exemplars printed here remain under their source agreements and are not a standalone public dataset.

B Model Specification, Training, and Evaluation Protocols

MASS uses the Clinical19 checkpoint for development and locked forecasting. MIMIC uses one history-aware setting: 13 clinical variables and CXR. The two cohorts differ in dimensions and fitted parameters. Architecture and loss definitions appear first; execution, prognosis, and planning details follow.

B.1 Cohort-specific model families and tensor interfaces

Observation modalities and internal state partitions are distinct. MASS has three observation streams: structured clinical measurements, CT representations, and BALF metabolomics. MIMIC-IV-CXR has two: structured clinical measurements and CXR representations. Both infer one patient state. The Clinical19 configuration groups its clinical variables into respiratory (5), renal (3), inflammatory (4), hemodynamic (4), and global-severity (3) blocks, with CT and BALF mapped to imaging and metabolic blocks. These seven internal state blocks parameterize the three observation modalities. MIMIC instead uses two internal blocks, clinical and radiograph.

Locked MASS forecasting uses this Clinical19 checkpoint. It partitions 19 variables into groups of sizes 5/3/4/4/3 and maps CT64 and BALF64 onto the imaging and metabolic compartments. Each of these seven compartments stores a 64-dimensional mean and diagonal log variance. The MIMIC model has 13 clinical variables and radiograph64 inputs, two 32-dimensional compartments, and hidden width 64. Its locked forecast uses the same history-aware filter that the development comparison evaluates against latest-value fusion. Both implementations use evidence-recency gating for cross-compartment messages.

Encoding, filtering, and prediction. Figure 8 is the module order in the reported checkpoints. Clinical values, BALF metabolomics, and the imaging vector each pass through a modality-specific numerical encoder: a masked value, its Boolean mask, and a learned modality embedding are concatenated and mapped by one linear layer to width H. An all-missing modality writes no evidence. MASS CT64 is a frozen Qwen2.5-VL latent prepared before training, so the VLM is not among the $5{,}338{,}172$ learned parameters. MIMIC uses CXR PCA64 the same way. A temporal imaging encoder then maintains an imaging state of width H: four learned queries and four attention heads pool the current imaging vector, and a GRU cell updates that state from the pooled token and an elapsed-time feature. On the reported vector interface the attention has a single key. Linear maps send the resulting tokens to a Z-dimensional mean and diagonal log variance for each compartment. The

multirate, care-conditioned transition advances those state blocks with a local MLP, Fourier features of elapsed time (16 on Clinical19, 8 on MIMIC), and SiLU gates on the fixed cross-compartment edges. After the last observation, rollout repeats this transition on the predicted state and decodes with the three-expert observation head. No Transformer block sits in this path. The causal Transformer and GRU comparators are trained separately.

The MASS-Syn actor is a categorical policy. It reads the flattened 3×8 latent state and a four-step one-hot code, 28 inputs in total, through two hidden layers of width 128 with $\tanh$ activations, and it emits seven action logits. That MLP has 21,127 parameters. A behavior policy of the same shape is fit by imitation and then held fixed. Imagined MASS-Syn transitions come from an ensemble of three frozen bilinear compartment models, each with latent width 8. This ensemble is the known simulator's learned transition used to score the same authorized-search interface. The simulator contains no patient records.

Observation encoder and correction. For a modality of width D_m and hidden width H, the numeric encoder concatenates masked values, the D_m-dimensional mask, and an H-dimensional learned modality embedding. A linear map $(2D_m + H) \to H$ produces its token. An all-missing modality cannot contribute observed evidence. Each compartment has separate $H \to Z$ initial-mean and initial-log-variance projections. The correction gate is an MLP $(Z + H) \to H \to Z$ with SiLU and sigmoid activations. The innovation projection maps $H \to Z$; the predicted-observation projection maps $Z \to H$. The innovation is the observed token minus this predicted token. Its gated projection is added to the prior mean; the prior variance is divided by one plus the accumulated correction gain. Token availability multiplies the gain, so an all-missing modality produces no correction. During recursive execution, decoded means and predicted availability (sigmoid threshold 0.5) supply the next decoder's source interface without posterior correction.

Dynamics and output heads. The Clinical19 graph contains 12 directed edges: reciprocal connections between global severity and each of the four clinical organ groups, between imaging and respiration, and between metabolism and inflammation. Local drift, diffusion, elapsed-time features, and gated messages update the belief. The learned time gain is bounded, with a 24-hour nominal scale and a 72-hour evidence half-life in the MASS configuration. Initial cross-edge probability is 0.05 and residual scale is 0.05. For an output of width D_m, the probabilistic head maps $Z \to 2D_m$ for dynamic mean/scale, $Z \to 3D_m$ for three-expert gate logits, and $Z \to D_m$ for availability. The experts are persistence, a learned population state, and a dynamic residual. Persistence is masked out when the source value is absent.

B.2 Training phases, losses, and stopping rules

Reported development and locked forecasts use one schedule per cohort (Table 8). MASS uses the Clinical19 base schedule. MIMIC uses the history-aware parent schedule and is selected by calibration recursive balanced RMSE across seeds 20260915–20260917; each selected checkpoint is from epoch 12. That MIMIC checkpoint uses 13 clinical variables and CXR for both the development comparison and the locked forecast. Locked MASS forecasts use the calibration-selected Clinical19 runs.

Constrained refinement and the MIMIC action adapter are later optimization phases on these checkpoints. They keep the cohort observation setting fixed. Refinement continues from the Clinical19 checkpoint and updates dynamics, observation heads, and predicted-observation projections (2,823,851 parameters). It did not satisfy its acceptance criteria, so it does not replace the locked MASS forecast. The action adapter updates 20,352 action-related parameters on the same MIMIC parent, leaves the other modules frozen, and retains epoch 7 by calibration clinical macro RMSE.

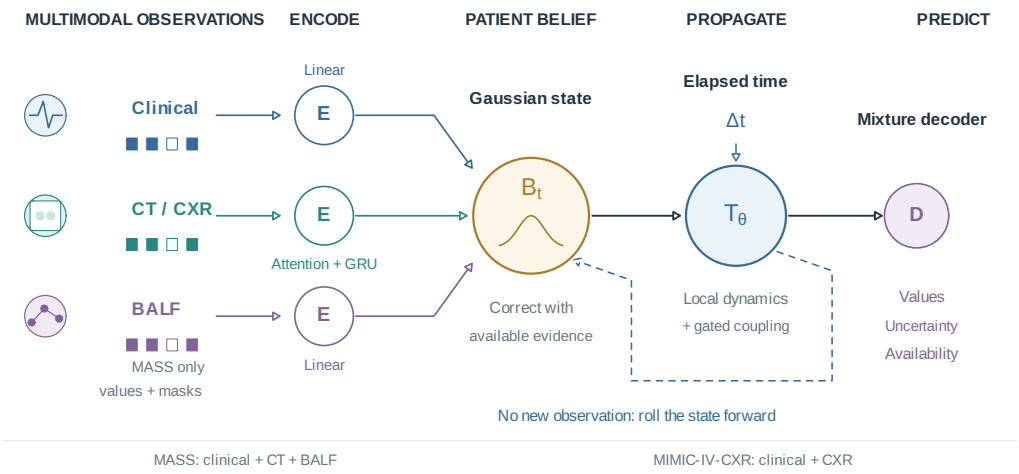


Figure 8: **Architecture of the PULMOWORLD observation filter.** Modality-specific encoders map masked observations into a Gaussian patient belief; imaging features additionally pass through attention and recurrent memory. Available evidence corrects the state, elapsed-time dynamics propagate it, and a mixture head predicts observations. The dashed loop propagates the predicted belief when no new observation arrives. MASS has three observation modalities and seven internal state blocks ($H = 256$, $Z = 64$); MIMIC has two modalities and two blocks ($H = 64$, $Z = 32$). These internal partitions are not additional modalities. The locked MIMIC forecast uses this same history-aware belief.

Table 7: **Checkpoint parameter accounting.** Counts are obtained from the stored tensor shapes; frozen image backbones, PCA, and target-network copies are excluded. The history/temporal-image row is the imaging memory above: multi-head attention plus a GRU cell. The MIMIC parent and action adapter have identical total size but different optimization scopes. Counts include allocated modules in the reported checkpoints.

Module	MASS Clinical19	MIMIC
Numeric encoders	537,600	18,304
History/temporal image encoder	939,008	62,720
State prior and initial projections	232,064	8,448
Dynamics	2,649,827	62,596
Prediction/innovation/correction	921,984	24,960
Observation and auxiliary heads	57,689	15,433
Total	5,338,172	192,461

B.3 DETAILED BELIEF UPDATES, FITTING, AND POLICY EXECUTION

The main text distinguishes observation modalities from the inferred patient state. This section gives the full parameterization and execution details. Clinical19 subdivides the clinical observation stream into five physiological groups; these are internal state blocks, not additional observation modalities. MIMIC uses two state blocks for its clinical and radiograph observations. The term compartment below denotes an internal state block.

B.3.1 PROBLEM SETUP

To connect observations collected at different times, we represent the patient through a belief inferred from the available history. We model the underlying evolution as a partially observed dynamical system. Let t index observations at calendar times τ_t, with elapsed time $\Delta t_t = \tau_{t+1} - \tau_t$. The ordered history is $\mathcal{H}_t = ((o_j, m_j, \tau_j, a_{j-1}))_{j=1}^{t}$, where o_j contains clinical measurements and modality-specific representations, and m_j identifies the observed entries. The preceding care context a_{j-1} is

included when enabled by the experimental configuration; future actions are not part of the initializing history. The model approximates the unobserved patient state Z_t with the learned belief

$$B_t = q_\theta(Z_t \mid \mathcal{H}_t). \tag{6}$$

Missing entries supply no measured value. Their availability remains explicit rather than being conflated with an observed numerical zero.

The world consists of a belief-transition kernel $\mathcal{T}_\theta(B' \mid B, \Delta t, a)$ and an observation model $\mathcal{O}_\theta(o, m \mid B)$. The latter predicts observation values and availability; the former propagates the patient belief, optionally conditioned on supplied care. Starting from $\widehat{B}_t = B_t$, an H-step open-loop forecast is defined recursively by

$$\begin{aligned} \widehat{B}_{t+h+1} &\sim \mathcal{T}_\theta\Big(\cdot \mid \widehat{B}_{t+h}, \Delta t_{t+h}, a_{t+h}\Big), \\ (\widehat{o}_{t+h+1}, \widehat{m}_{t+h+1}) &\sim \mathcal{O}_\theta\Big(\cdot \mid \widehat{B}_{t+h+1}\Big), \qquad h = 0, \dots, H-1. \end{aligned} \tag{7}$$

The kernel notation also accommodates deterministic updates of belief parameters. No factual intermediate observation is assimilated during this rollout. Evidence-recency metadata carried alongside the belief is left implicit in the operator notation.

We distinguish learning these predictive dynamics from optimizing interventions within them. Conditioning on a supplied action does not by itself establish its interventional effect. Policy learning is therefore restricted to actions supported by a separate, fixed authorization procedure, defined in Section 3.4.

B.3.2 Asynchronous multimodal belief learning

Asynchronous observations must update the patient state while retaining information from earlier measurements. We separate propagation between observations from correction by newly available evidence, allowing each observation group to refine its corresponding state block. The Clinical19 instantiation represents this belief with seven Gaussian state blocks,

$$B_t = \{B_t^c\}_{c=1}^{7}, \qquad B_t^c = \mathcal{N}(\mu_t^c, \operatorname{diag}(\exp \ell_t^c)), \tag{8}$$

corresponding to respiratory, renal, inflammatory, hemodynamic, global-severity, lung-imaging, and lung-metabolic states. The first five organize the 19 clinical variables; the last two represent CT64 and BALF64. Cohort-specific observation interfaces and the use of history-enabled filtering are specified in Sections B.1 and 4.

Before assimilating an observation at τ_t, the preceding belief is propagated over Δt_{t-1} to obtain a prior $\widetilde{B}_t$. Available evidence then corrects this prior (Figure 2(a)). Let v index the observation groups assigned to state blocks by $c(v)$. Here clinical measurements are indexed through their five predefined feature groups, rather than being assigned as a single modality to one state block. Modality encoders process masked values and masks, $(m_t^v \odot o_t^v, m_t^v)$, to supply observation tokens e_t^v. The innovation and correction are

$$\begin{aligned} \nu_t^v &= e_t^v - D_v(\widetilde{\mu}_t^{c(v)}), \\ K_t^v &= b_t^v\, \sigma\Big(G_v[\widetilde{\mu}_t^{c(v)}, \nu_t^v]\Big), \\ \mu_t^c &= \widetilde{\mu}_t^c + \sum_{v:c(v)=c} K_t^v \odot P_v(\nu_t^v), \end{aligned} \tag{9}$$

where b_t^v indicates the presence of observed evidence in group v, and D_v and P_v map between belief and token spaces. The learned gate modulates an observation-driven correction rather than replacing the propagated state. In particular, $b_t^v = 0$ removes that group's measurement correction without preventing state evolution. The same correction gains reduce the prior variance elementwise by a factor of one plus the accumulated gain, while learned nonnegative diffusion increases variance during propagation. Alongside this distribution, each state block carries evidence reliability and observation age. Observed evidence resets reliability to one and age to zero; between observations, reliability decays exponentially with a prescribed half-life and age increases with time. Compartments without initial evidence start with zero reliability. Thus, evidence age measures time since the last observation and is distinct from the elapsed time between consecutive transitions. The complete updates are specified in Section B.1.

B.3.3 SHARED-STATE DYNAMICS AND WORLD-MODEL FITTING

Recursive simulation carries each state block forward after observations stop. Each block combines a local transition with sparse cross-block messages:

$$\begin{aligned} d_t^c &= F_{\mathrm{loc}}^c(\mu_t^c, T(\Delta t_t)) + \sum_{j \to c} g_{j\to c,t} M_{j\to c}(\mu_t^j), \\ \eta_c(\Delta t_t) &= 1 - \exp[-\operatorname{softplus}(\omega_c)\Delta t_t/24], \\ \widetilde{\mu}_{t+1}^c &= \mu_t^c + \eta_c(\Delta t_t) d_t^c. \end{aligned} \tag{10}$$

Here T supplies Fourier time features, Δt_t is expressed in hours, and $g_{j\to c,t}$ is a learned message gate. In the evidence-gated configuration, it is scaled by source reliability and a target-receptivity factor that increases as target evidence becomes stale. A separate posterior-reliability option scales messages using the source and target mean log variances. These quantities describe evidence freshness and within-model belief uncertainty, respectively; the ensemble uncertainty used for planning is defined below. The gain satisfies $\eta_c(0) = 0$. The sparse graph couples physiological state blocks through global severity and links imaging to respiration and metabolism to inflammation. These connections are the message graph of the transition.

To compare supplied care conditions from a common patient state, we branch this same transition through a state-dependent action residual. The residual is initialized to zero and augments the shared transition when care is supplied:

$$\widetilde{\mu}_{t+1}^c = \mu_t^c + \eta_c(\Delta t_t)\left[d_t^c + g_c(\mu_t^c, E_a(a_t), \Delta t_t) R_c(E_a(a_t))\right]. \tag{11}$$

The encoder E_a represents care, while g_c determines its state-dependent contribution. Omitting this residual defines state-only prediction. Observation heads decode clinical variables and modality-specific representations, with separate heads for future availability. Imaging outputs are latent representations rather than synthesized CT or CXR pixels. After initialization, predicted beliefs supply subsequent transitions as in Equation (7); decoded predictions are not treated as new factual measurements.

World-model fitting. Training grounds the shared belief in observed measurements and supervises its evolution through recursive predictions. T1 supervises source-state reconstruction, while T2 supervises recursive future predictions and availability (the lower panel of Figure 2). Writing the corresponding terms as $\mathcal{L}_{\mathrm{rec}}$, $\mathcal{L}_{\mathrm{dyn}}$, and $\mathcal{L}_{\mathrm{avail}}$, world fitting can be summarized as

$$\mathcal{L}_{\mathrm{world}} = \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{dyn}}\mathcal{L}_{\mathrm{dyn}} + \lambda_{\mathrm{avail}}\mathcal{L}_{\mathrm{avail}} + \mathcal{L}_{\mathrm{aux}}. \tag{12}$$

The active terms and weights follow the stage-specific schedules in Section B.2. Reconstruction hides available source entries and minimizes their Gaussian negative log-likelihood under the observation heads. Recursive supervision applies the corresponding masked likelihood to future values, with binary cross-entropy for their availability. Value losses are evaluated only where target measurements are observed; future observations provide supervision, not intermediate rollout inputs. On MIMIC, three reconstruction epochs can precede nine rollout epochs within the same twelve-epoch budget. The history comparison crosses that history-aware input with an auxiliary reconstruction objective. $\mathcal{L}_{\mathrm{aux}}$ collects posterior-alignment, direct-path consistency, and grounding terms where applicable. The MIMIC rollout objective also supervises direct forecasts and aligns recursive beliefs with stopped-gradient direct predictions using smooth-L_1 losses. MASS-Syn fits each ensemble member by one-step latent-state mean squared error and selects its checkpoint on calibration data. After fitting and calibration, θ is frozen. Prognostic probes evaluate the information in predicted states and do not train the world dynamics.

The formulation above covers the framework's components; Section 4 specifies their use. MIMIC development comparisons and the locked forecast use one history-aware belief over 13 clinical variables and CXR. MASS-Syn instantiates action-conditioned dynamics and policy improvement on simulator latent states.

B.3.4 EVIDENCE-CONSTRAINED MODEL-BASED REINFORCEMENT LEARNING

Action-conditioned futures become useful for planning when evidence supports comparing the consequences of the candidate actions. Cap-MBRL makes this requirement explicit by determining

action admissibility before optimizing behavior. Authorization restricts the policy's action support; uncertainty-penalized policy learning and risk-budgeted selection govern improvement within that support (Figure 2(b)).

Evidence-based action authorization. The authorization audit tests both whether recorded actions inform prediction and whether their effects support interventional comparison. T3 denotes this prespecified audit, separate from world-model fitting. Let S_t denote recorded source-state features used for auditing, distinct from the learned belief. Predictive evidence compares registered state-only and action-conditioned predictors, $f(S_t, \Delta t)$ and $f_A(S_t, \Delta t, a_t)$, on patient-disjoint data. An audit cell specifies the action comparison covered by a certificate; MASS-Syn uses one cell per intervention channel. An audited cell must satisfy

$$\Delta_A = \mathcal{L}(f) - \mathcal{L}(f_A) \geq \epsilon_A, \qquad p_{A,\mathrm{adj}} \leq \alpha_A, \tag{13}$$

and pass the prespecified action-shuffling and support checks. Here $p_{A,\mathrm{adj}}$ summarizes patient-clustered evidence for the required tests. MASS-Syn evaluates ridge and extra-trees predictors, with Bonferroni adjustment for family selection and Holm correction across action cells. Together these conditions define G_A. A separate indicator G_I records whether the identification evidence required for interventional comparison is available and passes the prescribed checks. Predictive improvement alone never sets $G_I = 1$. In MASS-Syn, known transition means provide the reference action effect relative to no intervention. Identification checks bound effect-prediction error and binary action-versus-reference decision regret, and require sufficient oracle gain and patient support. Patient-clustered tests with Holm correction determine which cells pass these checks on calibration data. The resulting authorized set is

$$\mathcal{A}_{\mathrm{auth}}(B) = \{a : G_A(B, a) = 1,\ G_I(B, a) = 1\}. \tag{14}$$

The gate notation refers to fixed audit decisions. In MASS-Syn, these decisions are indexed by action channel and held constant throughout a rollout, so no mapping from imagined beliefs to source-state bins is required. The no-intervention reference a_0 is always retained, giving $\mathcal{A}^{\mathrm{Syn}}_{\mathrm{auth}} = \{a_0\} \cup \{a_j : G_A(j)G_I(j) = 1\}$. Channels without a passing certificate remain excluded at every imagined step. State-dependent extrapolation is assessed through the risk signals below, rather than through a new state-specific certificate. The audit protocol specifies support conditions and identification checks (Section B.4).

When the set is nonempty, the actor softmax is normalized only over authorized actions at each imagined step. An empty set produces abstention rather than an unauthorized control rollout. In this study, policy learning and selection are confined to known-mechanism MASS-Syn; real-cohort action conditioning remains predictive.

Policy learning from imagined trajectories. Within the authorized action set, policy learning evaluates choices through their cumulative predicted consequences. Comparing continuations from the same initial state provides the learning signal while holding the patient's starting condition fixed. On MASS-Syn, the actor π_ϕ generates K continuations from each initial belief B_i in a frozen three-member world ensemble. The members share a transition architecture but use separate random initializations and training orders. Their mean prediction supplies the next latent state. At each step, u is the variance across member predictions, averaged over latent coordinates and normalized by its calibration-set 90th percentile. It measures ensemble disagreement and does not include the Gaussian belief variance. Let $\mathbb{P}_{\theta,\phi}$ denote the resulting trajectory distribution. For continuation k, its uncertainty-penalized return and group-normalized advantage are

$$\begin{aligned} S_{ik} &= \sum_{h=1}^{H} \gamma^{h-1}(r_{ikh} - \kappa u_{ikh}), \\ \widehat{A}_{ik} &= \frac{S_{ik} - \overline{S}_i}{\max(\mathrm{Std}_k(S_{ik}), \epsilon_{\mathrm{adv}})}, \end{aligned} \tag{15}$$

where u_{ikh} is model uncertainty and $\overline{S}_i$ averages returns across continuations from the same initial belief. Rewards are derived from the MASS-Syn latent-severity objective; the reward convention and uncertainty statistic are specified in Section B.6. A behavior policy μ_b is first fitted by cross-entropy to the logged seven-way action and then frozen. These actions comprise no intervention and one

unit intervention in each of six channels; simultaneous multi-channel actions are outside this policy interface. The actor is initialized from the behavior policy. The implemented sampled objective is

$$\mathcal{L}_{\mathrm{Cap}} = -\mathbb{E}_{i,k}\left[\mathrm{sg}(\widehat{A}_{ik})\sum_{h=1}^{H}\gamma^{h-1}\log\pi_\phi(a_{ikh}\mid B_{ikh})\right] + \lambda_{\mathrm{KL}}\widehat{D}_{\mathrm{KL}}(\pi_\phi,\mu_b) + \lambda_\rho\widehat{\mathbb{E}}_{i,k}\left[(\bar{\rho}_{ik}-\bar{b}_{ik})_+\right], \tag{16}$$

where sg denotes the stop-gradient operator on the advantage. The risk statistic combines ensemble disagreement, behavior surprisal, recursive–direct discrepancy, and capability-violation diagnostics. Behavior surprisal is the negative log-probability of the chosen action under μ_b, divided by $\log 7$. Path discrepancy is the mean squared difference between recursive states and a separately fitted direct predictor conditioned on the action history, normalized on calibration data. Each nonnegative signal x is mapped to $x/(1+x)$ before weighted averaging; $\bar{\rho}$ and $\bar{b}$ are discount-normalized trajectory averages. The capability-violation signal is zero for authorized samples.

Imagined rewards enter the actor update as a score-function gradient: the stopped advantage weights the log-probability of the categorical sample. The KL term and the positive risk-budget gap are part of the same sampled objective. A nonnegative multiplier tracks mean budget violation and scales λ_ρ. The same budget is enforced again when the learned policy is used for decision-time selection. The per-step budget is

$$b_h = \mathrm{clip}(b_0 - \alpha\rho_h^{\mathrm{ref}}, b_{\min}, b_0). \tag{17}$$

During training, ρ_h^{ref} evaluates a no-intervention step from the current imagined state, retaining the preceding action history. At decision time, the reference is the candidate branch whose first action is a_0 and whose continuation follows the actor. The base decision-time budget is calibrated to a prespecified acceptance rate and then frozen. This reference is separate from the fallback trajectory used to assess improvement.

Only ϕ is optimized: world parameters, audit decisions, the behavior model, and the fallback remain fixed. The exhaustive minimizer used to evaluate selection regret is not a policy-training target.

Risk-constrained decision-time selection. At decision time, the learned policy supplies continuations for comparing immediate actions, while the risk budget and fallback comparison determine whether to adopt a proposal. MASS-Syn selection enumerates the seven possible first actions and restricts selection to authorized candidates; training samples continuations instead. After fixing the first action, each candidate branch uses the trained policy's highest-probability authorized action at every subsequent imagined state. Thus, policy learning determines the continuation used to evaluate each first action. A separate fallback branch follows its own fixed policy throughout the remaining horizon. In the reported experiment, this is the capability-mask policy, initialized from μ_b and trained within the same authorized action set before being frozen.

Let $\widehat{C}(B,a)$ and $\widehat{\rho}(B,a)$ denote a candidate's discounted cost and discount-normalized risk. The cost sums predicted latent severity and a weighted risk penalty over the remaining horizon. Let $\widehat{b}(B)$ be the shared state-dependent budget, and $\widehat{C}_{\mathrm{fb}}(B)$ the cost of the fallback's own continuation. Its first action a_{fb} must itself be authorized. If no authorized fallback exists, selection abstains. Otherwise, for a frozen improvement margin $\delta \geq 0$, define

$$\mathcal{F}(B) = \big\{a \in \mathcal{A}_{\mathrm{auth}} : \widehat{\rho}(B,a) \leq \widehat{b}(B), \ \widehat{C}(B,a) < \widehat{C}_{\mathrm{fb}}(B) - \delta\big\}. \tag{18}$$

The executed action is

$$a_{\mathrm{exec}}(B) = \begin{cases} \perp, & a_{\mathrm{fb}} \notin \mathcal{A}_{\mathrm{auth}}, \\ \arg\min_{a\in\mathcal{F}(B)} \widehat{C}(B,a), & a_{\mathrm{fb}} \in \mathcal{A}_{\mathrm{auth}},\ \mathcal{F}(B)\neq\varnothing, \\ a_{\mathrm{fb}}, & \text{otherwise}, \end{cases} \tag{19}$$

where $\perp$ denotes abstention; an unavailable fallback is treated as outside the authorized set. In MASS-Syn, retaining a_0 ensures that the masked fallback has an admissible action. The budget screens proposals that would replace the fallback; fallback execution itself is not conditional on satisfying that budget. The selected first action is executed only in MASS-Syn, and the comparison repeats at the next state. The continuation rule and cost definitions are part of the fixed decision-time protocol (Section B.6).

B.4 Dataset-specific execution and action interfaces

The shared abstraction is a masked multimodal belief updated by elapsed time; the concrete modules are sized separately for each dataset rather than sharing a fitted forecasting head. Table 8 reports the configurations used for the quantitative and case analyses.

For MIMIC, the action adapter is fit on the same history-aware parent. Its training uses 618,453 H4 windows; calibration and validation contain 86,131 and 88,295 windows. Recursive loss, direct-horizon loss, availability loss, and horizon weights are frozen in the configuration; the action adapter sets the direct-path coefficient to zero and leaves the parent dynamics fixed.

Reconstruction and forecast errors are computed per patient, feature, and horizon before aggregation. All methods on a cohort are scored on identical observed targets. Forecast carry-forward is scored only when source and target are both observed. Per-patient losses are averaged before paired bootstrap.

The action-information gate uses state-only and state-plus-action predictors on a known-signal / null synthetic control (action coefficient 0.8 or 0) and on eight real MASS cells. Acceptance requires a prespecified MAE gain, a positive patient-bootstrap lower bound, factual gains over global and propensity-matched shuffles, and at least 80% propensity overlap. Both model families must pass. The action ontology is in Table 9. The eight prespecified real MASS cells do not open an authorized set, so planning is scored in MASS-Syn.

B.5 Prognosis comparator and outcome probe

MASS CLARITY comparator. This arm is fitted independently to Clinical19 clinical variables and CT64, and does not receive BALF. It uses four Transformer layers, hidden width 256, eight attention heads, dropout 0.1, batch size 64, AdamW at 3×10^{-4} with weight decay 10^{-4}, gradient clipping 1, and three training seeds. The 1,971 training pairs fit parameters; 368 calibration pairs select among 40 epochs (selected epoch 6). Clinical and CT squared errors have weight 1; availability cross-entropy has weight 0.1. The unchanged shared probe evaluates the 99 validation patients, including 28 deaths. Clinical values and availability are updated recursively; nonclinical probe coordinates remain at source values, matching the Direct-VLM observable-interface scoring convention.

The 28-day outcome probe is L2 logistic regression on train-standardized features. It receives source observables together with endpoint-minus-source changes and elapsed time, not the future state alone. The source summary includes clinical values/masks and CT/BALF representation summaries. The 99-patient MASS panel uses the available endpoint: H2, H3, and H4 for 2, 46, and 51 patients, respectively. C, temperature, and threshold are selected on calibration. The matched shuffle uses the same fitted probe.

B.6 Policy fitting and decision-time search

When $G_A \cap G_I$ exposes $\mathcal{A}_{\mathrm{auth}}(B_t)$, we optimize a two-layer state–action MLP π_ϕ on MASS-Syn only. Each hidden layer has width 128 and a $\tanh$ nonlinearity. The input is the 24-dimensional latent state plus a four-step indicator, and the output is seven action logits ($21{,}127$ parameters; Figure 8). The state cost and return are

$$c(B_h) = \sum_{k=1}^{3} w_k[(B^k_{h,1})^2 + 0.5(B^k_{h,2})^2], \quad R(\tau) = -\sum_{h=1}^{H} \gamma^{h-1} c(B_h), \tag{20}$$

with $w = (1, 0.65, 0.45)$. Model uncertainty, behavior surprisal, recursive–direct discrepancy, and capability violation form a bounded risk ρ_h. The no-action risk tightens $b(B_t) = \mathrm{clip}(b_0 - \alpha\rho_0(B_t), b_{\min}, b_0)$. Fitting uses the root-mean-square standardized group-relative advantage in Equation (16), sampled KL to behavior, and a nonnegative dual multiplier on mean budget violation. At inference a feasible action is accepted only when it beats the authorized-set fallback by a frozen margin.

What training compares. The behavior policy is trained by cross-entropy on the logged category. Category 0 is the zero action. Categories 1–6 are the six one-hot channels. That policy is then frozen. Actor training samples a category from the authorized softmax and rolls it through the ensemble. Several samples that share an initial state form a group. Each trajectory receives its

Table 8: **Model architecture and training hyperparameters.** Each cohort has one observation setting. Parameter counts include learned world-model modules but exclude frozen image encoders and train-fitted PCA. Constrained refinement continues from the Clinical19 checkpoint. The MIMIC action adapter initializes from the history-aware parent and updates only action residual and gate parameters.

Phase	Inputs and state	Optimization	Selection and size
MASS multimodal	Clinical19; CT64; BALF64; seven 64-D compartments; hidden 256; 16 Fourier time features; bounded continuous time; action branch disabled.	Base prediction: AdamW, batch 16, learning rate 5×10^{-5}, at most 20 epochs. Constrained refinement: 10^{-5}, at most 12 epochs, recursive/direct horizon weights 0.05/0.25/1.5/2.2.	Calibration selection; seeds 20260915–20260917; 5,338,172 learned parameters per checkpoint.
MIMIC	13 clinical variables; CXR PCA64; two compartments (clinical/radiograph); hidden 64; compartment 32; 8 Fourier time features; four recursive horizons.	AdamW; batch 256; learning rate 10^{-3}; weight decay 10^{-4}; gradient clipping 5.0; 12 epochs; horizon weights 0.25/1/1.5/2.	Calibration recursive balanced RMSE; seeds 20260915–20260917; 192,461 learned parameters; action disabled.
MIMIC action-conditioned	Same parent plus two transformed factual action channels in a 6-D reserved interface; action residual and gate enabled.	AdamW on action adapter only; batch 256; learning rate 3×10^{-4}; weight decay 10^{-4}; gradient clipping 1.0; at most 8 epochs; patience 2 after epoch 3.	Calibration clinical macro RMSE; selected epoch 7; 20,352 action-related parameters within the 192,461-parameter model.

Table 9: Frozen MASS action ontology. Twenty privacy-safe macro classes are collapsed by a within-group maximum into ten mechanism blocks.

Mechanism block	Member macro classes
Broad-spectrum beta-lactam	Antipseudomonal beta-lactam/inhibitor, carbapenem, cephalosporin
Resistant Gram-positive	Gram-positive antibiotic
Other antibacterial	Fluoroquinolone, tetracycline, aminoglycoside, polymyxin, other antimicrobial
Antifungal	Antifungal
Antiviral	Antiviral
Immunomodulation	Systemic corticosteroid, other immunomodulator
Nutrition/microbiome support	Enteral nutrition, parenteral nutrition, probiotic
Acid suppression	Acid suppression
Ventilation support	Invasive mechanical ventilation
Extracorporeal support	Continuous renal replacement therapy, ECMO

discounted negative severity cost minus an ensemble-uncertainty penalty. Its advantage is the centered, standardized score within the group (Equation (15)); thus reinforcement rewards better predicted trajectories without requiring oracle action labels. The KL term penalizes departure from the behavior distribution. The positive part of the group's mean budget violation enters Equation (16) on the same categorical samples (Section B.3). Decision-time selection applies the same budget to the action that is executed.

Inference and selection-regret evaluation. At each decision, the ranker enumerates candidate first actions and continues each trajectory with the fitted policy through the frozen world-model ensemble. A separate trajectory follows the authorized fallback. Candidate scores combine discounted predicted state cost and bounded risk. The optional policy-preference penalty has coefficient zero in the reported experiment. The authorization mask and state-dependent risk budget restrict eligibility; the fallback is retained unless a feasible candidate clears the score margin. The selected first action is executed, and ranking repeats at the next state. For evaluation only, the known simulator supplies each policy's realized discounted cost and the lowest cost among authorized action sequences. Their difference is selection regret; the oracle is not supplied to the planner. Figure 5d compares this quantity across methods with zero unauthorized action rate. It evaluates selection within the permitted set, not the separate regret imposed by restricting that set.

C Complete Quantitative Results

This section supplements the main experiments with full forecasting comparisons, source-reconstruction and history controls, and action-evidence diagnostics. Each comparison identifies its evaluation population: locked clinical tests, development component studies, or known-mechanism simulation. These populations address complementary questions and are not pooled.

C.1 Evaluation metrics and uncertainty

Macro NMAE averages feature-wise errors normalized by training IQR. Consecutive-change error is

$$\mathrm{NMAE}_{\Delta} = \frac{|(\hat{y}_h - \hat{y}_{h-1}) - (y_h - y_{h-1})|}{\mathrm{IQR}_{\mathrm{train}}}, \qquad \hat{y}_0 = y_0. \tag{21}$$

H4 direction agreement measures sign agreement, not change magnitude. EHR-S@15 averages feature-level success at relative error at most 15%, with denominator $\max(|y|, 0.1\,\mathrm{IQR}_{\mathrm{train}})$. It is a retrospective compatibility score, not a clinical acceptability threshold. Whole-patient bootstrap uses 2,000 resamples for forecasting and action evaluation, and 5,000 for seed-averaged MASS trends and paired future-state prognosis comparisons. Component-specific resampling protocols are stated with their tables.

The unauthorized-action rate is the fraction of executed actions outside the permitted set. A zero rate means the hard mask was respected. It does not measure whether an authorized comparison was identified. Selection regret compares decisions against the best action in the permitted set. Full definitions are in Section B.6.

C.2 COMPLETE LOCKED-TEST FORECASTING COMPARISON

Table 10: **Locked-test longitudinal simulation.** H1–H4 report macro NMAE; Δ-NMAE scores consecutive changes (patient-cluster 95% CIs); H4 dir. scores direction agreement. Bold and underline mark the best and second-best task-trained result in each column. Ties share marks.

MIMIC-IV-CXR pneumonia locked test (300 patients)

Method	H1↓	H2↓	H3↓	H4↓	H4 EHR-S@15↑	Δ-NMAE↓	H4 dir.↑
Controls							
Persistence	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
Train median	0.7188	0.7009	0.6839	0.7090	48.8%	0.471 [0.454,0.489]	59.6%
Temporal models							
GRU	0.5068	**0.5483**	0.5818	0.5934	52.1%	0.412 [0.396,0.428]	60.4%
Causal Transformer	**0.5005**	0.5527	0.5877	0.5954	52.5%	0.411 [0.395,0.428]	60.1%
GRU-D	0.5238	0.6815	0.8418	0.9600	38.3%	0.444 [0.429,0.461]	58.1%
General-purpose Direct VLMs							
Qwen2.5-VL-3B†	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
Qwen2.5-VL-7B	2.0176	2.8403	2.1843	1.8233	44.7%	0.682 [0.572,0.795]	56.7%
Qwen2.5-VL-72B	0.9069	2.2291	0.8848	3.0042	44.0%	0.774 [0.446,1.410]	57.0%
Qwen3-VL-8B-Instruct	4.5230	5.5157	6.1501	6.2845	46.0%	0.467 [0.449,0.486]	57.0%
Qwen3-VL-30B-A3B	33.2483	35.0294	40.3677	25.2325	45.2%	0.581 [0.469,0.754]	56.1%
InternVL3-8B†	0.6740	0.7491	0.7692	0.7752	49.4%	0.443 [0.425,0.461]	56.8%
LLaVA-OV-1.5-8B	2.4461	1.0740	3.4511	1.2832	38.3%	0.644 [0.564,0.742]	55.5%
Medical Direct VLMs							
Lingshu-7B	1.0245	1.1025	1.1310	1.1103	46.9%	0.458 [0.430,0.500]	56.8%
Hulu-Med-7B	24.1641	11.6647	14.1120	9.8788	47.0%	0.897 [0.564,1.500]	57.7%
Hulu-Med-32B	54.5972	34.4492	45.6867	22.0711	47.8%	0.588 [0.447,0.845]	57.6%
Task-adapted VLM and medical world model							
Qwen2.5-VL-7B SFT	0.6484	1.6620	0.7182	0.7221	49.5%	0.875 [0.589,1.415]	56.4%
CLARITY (adapted)¶	0.5110	0.5746	0.6129	0.6328	50.8%	0.531 [0.513,0.550]	61.4%
Ours							
PULMOWORLD	0.5034	0.5489	**0.5722**	**0.5780**	**54.9%**	**0.407 [0.392,0.423]**	**61.7%**

MASS Clinical19 locked test (40 patients; three training seeds)

Method	H1↓	H2↓	H3↓	H4↓	H4 EHR-S@15↑	Δ-NMAE↓	H4 dir.↑
Controls							
Persistence	0.5726	0.6727	0.5951	0.5563	38.5%	0.523 [0.449,0.610]	50.0%
Train median	0.5444	0.6134	0.5672	0.5563	38.5%	0.509 [0.439,0.578]	50.0%
Temporal models							
GRU	0.5438	0.6648	0.5984	0.6908	39.7%	0.507 [0.445,0.577]	52.2%
Causal Transformer	0.5543	0.7039	0.6080	0.6995	35.0%	0.503 [0.440,0.572]	52.2%
GRU-D	0.5237	0.6397	0.5718	0.5442	42.2%	0.503 [0.436,0.575]	**56.7%**
General-purpose Direct VLMs							
Qwen2.5-VL-3B	2.3613	2.8903	6.8647	16.4983	3.8%	0.527 [0.454,0.616]	63.3%
Qwen2.5-VL-7B	0.7887	1.2337	0.8895	1.0127	30.4%	0.561 [0.479,0.651]	60.0%
Qwen2.5-VL-72B	4.9553	0.9430	0.8230	0.9236	36.4%	0.532 [0.453,0.624]	50.0%
Qwen3-VL-8B-Instruct	0.8699	1.1391	0.8965	1.0233	45.5%	0.525 [0.450,0.610]	60.0%
Qwen3-VL-30B-A3B	2.9213	3.1071	6.2477	1.3304	29.1%	0.526 [0.453,0.611]	60.0%
InternVL3-8B†	0.5726	0.6727	0.5951	0.5563	38.5%	0.523 [0.449,0.610]	50.0%
LLaVA-OV-1.5-8B	6.3862	6.7173	17.9637	18.4743	0.0%	0.558 [0.475,0.661]	66.7%
Medical Direct VLMs							
Lingshu-7B	4.8341	2.8278	9.5505	10.0853	11.2%	2.390 [0.915,4.076]	66.7%
Hulu-Med-7B	6.3582	6.6926	17.8894	18.4743	0.0%	0.522 [0.449,0.610]	66.7%
Hulu-Med-32B	0.9901	1.2992	0.8678	1.0012	44.3%	0.534 [0.463,0.618]	56.7%
Task-adapted VLM and medical world model							
Qwen2.5-VL-7B SFT	4.8415	1.3606	2.2658	1.5103	26.2%	2.519 [1.625,3.660]	53.3%
CLARITY (adapted)¶	**0.5217**	0.8358	0.6554	0.6302	**49.9%**	0.601 [0.542,0.664]	53.3%
Ours							
PULMOWORLD	0.5218	**0.5997**	**0.5658**	**0.5418**	46.8%	**0.498 [0.431,0.569]**	51.1%

¶CLARITY is independently adapted and fitted per cohort. †This Direct VLM row repeats the persistence control of its panel on every cell. MASS and MIMIC task-trained models and PULMOWORLD are three-seed means. EHR-S is a retrospective compatibility score, not a clinical acceptability threshold.

Benchmark conventions. CLARITY is adapted separately for each cohort. On MASS, that comparator is fitted to clinical variables and CT64, without BALF, and uses calibration selection. SFT is fitted separately on the clinical forecast interface: Clinical19 for MASS and the 13-variable

plus CXR interface for MIMIC. MASS persistence carries the last observed interface forward with elapsed time retained; train median uses the train-split median vector. Both adapted models use the cohort's shared outcome probe. PULMOWORLD forecast and prognosis remain on the reported checkpoints: Clinical19 with CT and BALF for MASS, and 13 clinical variables with CXR for MIMIC (Section C.6).

C.3 FUTURE-STATE PROGNOSIS AND EXTENDED ROLLOUT DISPLAYS

One frozen outcome probe per cohort reads future-state summaries. The probe is shared across methods within a cohort and is not shared between cohorts. MIMIC locked test has 300 pneumonia patients and 32 in-hospital deaths, scored at H4. MASS uses a separate 99-patient frozen-validation panel with 28 deaths by day 28, scored at the available horizon: H2, H3, and H4 for 2, 46, and 51 patients. Fallback handling is in Section C.2. Relative to the current-state control, predicted states raised AUROC from 0.7106 to 0.7135 on MIMIC and from 0.6856 to 0.7186 on MASS. AUPRC and Brier score also improved. The complete prognosis comparison is retained in the main text (Table 3), alongside the MASS clinical and representation forecast display (Figure 3).

Figure 4a shows observed CXR references; panel (b) shows lower PULMOWORLD representation error than CLARITY and persistence on that window. Panels (c–e) show selected clinical trajectories after observations stop. The CT64 and BALF64 development displays in Figure 3 are separate from locked-test clinical scores. H1 MSE was $0.836/0.836$, compared with $0.868/0.889$ for GRU. These checks concern representation forecasts.

The case figure is retained in the main text (Figure 4). Its CXR panel compares PULMOWORLD, CLARITY, persistence, and the train mean in the same standardized representation space, not pixel space. The clinical panels include observed values, PULMOWORLD, the causal Transformer, GRU, and persistence. The six-hour labels denote event windows rather than exact image acquisition times (Figure 12). Case selection and interpretation are detailed in Section D.6.

C.4 T1 SOURCE COMPLETION

T1 here is masked source completion on the numerical interface used by the forecasting baselines. The frozen PULMOWORLD checkpoint is evaluated on exactly the same 271 MIMIC requests, 718 hidden cells, train-only transforms, and scorer as the Direct VLMs. Requested current clinical values are zeroed and masked before encoding; the action-disabled model then decodes the masked belief after a 10^{-3}-h transition. This is a source-completion comparison. T1 scores masked source completion on this checkpoint, not the three-seed mean used for locked recursive forecasts. The history-versus-fusion comparison is the development experiment in Section 4.2.

Table 11: **Masked source reconstruction under two evaluation protocols.** The MIMIC column is the frozen locked test: 271 requests, 718 masked cells, the same train-only IQR and scorer as the Direct VLMs in Table 12. The MASS column is the Clinical19 development validation: 105 examples and 210 masked values. It is not the 40-patient locked forecast, so the columns are not a paired comparison. Lower NMAE, RMSE, NLL, and Brier score are better. Coverage closer to 95% is better.

	MIMIC locked test	MASS Task 1 validation
Requests / masked values	271 / 718	105 / 210
Masked NMAE	0.783 cell; 0.834 macro	0.667
Std. RMSE / Gaussian NLL	1.143	3.228
EHR-S@10/15/25	34.6 / 42.8 / 58.8	–
95% coverage / availability Brier	–	90.0% / 0.278

Table 12: **T1 source completion on the numerical interface.** MIMIC-IV-CXR uses the frozen locked-test cohort (271 requests; 718 scored cells), a 25% stateless observed-cell mask, and train-only median/IQR normalization. The PULMOWORLD row uses that same mask, cohort, and scorer. MASS uses the Clinical19 locked-test protocol (20 requests; 37 scored cells) and has no matched PULMOWORLD row. EHR-S is feature-macro in both panels.

MIMIC-IV-CXR locked test

Method	Cell NMAE↓	Macro NMAE↓	Std. RMSE↓	EHR-S@10/15/25 (%↑)
PULMOWORLD (frozen)	**0.783**	**0.834**	**1.143**	34.6 / 42.8 / 58.8
Qwen2.5-VL-3B	7.834	7.029	69.054	18.4 / 22.4 / 30.5
Qwen2.5-VL-7B	7.764	6.419	17.717	12.4 / 16.1 / 21.2
Qwen2.5-VL-72B	15.103	14.914	123.536	20.1 / 24.5 / 32.5
Qwen3-VL-8B-Instruct	21.294	22.602	163.906	29.2 / 35.8 / 48.3
Qwen3-VL-30B-A3B	26.089	25.263	146.563	7.2 / 8.5 / 11.8
InternVL3-8B	1.338	1.304	5.524	35.5 / 43.5 / 55.9
LLaVA-OV-1.5-8B	13.884	13.429	101.546	14.9 / 17.6 / 23.0
Lingshu-7B	9.233	8.559	80.423	19.7 / 24.1 / 33.2
Hulu-Med-7B	6.452	6.333	70.518	31.4 / 38.8 / 51.1
Hulu-Med-32B	114.072	118.842	362.066	29.6 / 36.5 / 49.8

MASS Clinical19 locked test

Method	Cell NMAE↓	Macro NMAE↓	Std. RMSE↓	EHR-S@10/15/25 (%↑)
Qwen2.5-VL-3B	4.327	4.895	13.644	16.1 / 16.1 / 18.0
Qwen2.5-VL-7B	17.607	21.532	100.550	24.5 / 40.5 / 51.8
Qwen2.5-VL-72B	1.556	1.494	1.874	20.0 / 23.3 / 33.3
Qwen3-VL-8B-Instruct	4.041	4.573	13.587	16.1 / 16.1 / 16.1
Qwen3-VL-30B-A3B	6.309	7.350	19.122	10.0 / 15.6 / 15.6
InternVL3-8B	0.572	0.592	0.757	43.0 / 56.2 / 69.4
LLaVA-OV-1.5-8B	1.867	1.899	2.207	16.7 / 19.4 / 22.2
Lingshu-7B	1.491	1.528	1.859	26.7 / 33.7 / 35.5
Hulu-Med-7B	0.573	0.593	0.758	43.0 / 56.2 / 69.4
Hulu-Med-32B	1.451	1.466	1.798	22.2 / 22.2 / 31.1

Standardized RMSE uses train-IQR normalization. The MIMIC PULMOWORLD row uses the same frozen history-aware checkpoint: requested current clinical values are hidden before encoding and decoded after an effectively zero-time (10^{-3} h) transition. Bold marks the lowest error in the MIMIC panel. These are source-completion diagnostics, not recursive forecasting or control comparisons.

MASS source-reconstruction validation. A separate MASS Task 1 validation evaluates 105 held-out examples (210 masked source values) under the Clinical19 reconstruction curriculum (Table 11). It obtains masked NMAE 0.667, masked Gaussian NLL 3.228, 95% interval coverage 90.0%, and availability Brier score 0.278. Transferring this value into the MASS panel above would mix distinct masks, examples, and scoring protocols.

C.5 COMPONENT ABLATIONS AND MISSING-EVIDENCE SENSITIVITY

The component studies test three distinct capabilities: retaining observation history, predicting clinical values and modality representations from a shared state, and selecting actions under fixed authorization. The following map links each capability to its controls and states the interpretation supported by those controls.

Design	Corresponding evidence	Boundary
Asynchronous observation updates	MIMIC development history comparison (Table 1)	On MIMIC, history-aware filtering improves prognosis versus latest-value fusion, with and without auxiliary reconstruction.
Multimodal evolving state	CXR/CT/BALF representation prediction; clinical rollouts; frozen-probe prognosis	The shared state supports clinical forecasts and CT/CXR/BALF representation forecasts under the same evaluation.
Treatment-conditioned transitions	Logged-versus-zero/shuffled conditioning and same-source candidate branches	Logged care changes the predicted trajectory.
Frozen-world planning	Authorized-set regret and risk-budget comparison on MASS-Syn	Authorized search lowers selection regret in MASS-Syn.

The main text summarizes history retention in Table 1. Reconstruction-curriculum and missing-evidence controls are reported below.

C.5.1 HISTORY-ENABLED MIMIC PROGNOSIS FOLLOW-UP

The development comparison uses one frozen H4-eligible window per patient on the MIMIC development set (3,758 patients; 524 in-hospital deaths), and three training seeds. History-aware filtering updates the state from ordered observations. Latest-value fusion merges each channel to its most recent value. The comparison is reported in Table 1: without reconstruction, filtering improves AUROC by 0.0213 (95% CI $[0.0132, 0.0295]$). With reconstruction, the AUROC, AUPRC, and Brier changes relative to latest-value fusion are $+0.0381$, $+0.0778$, and -0.0110. Both comparisons use this MIMIC history-aware belief.

C.5.2 MIMIC CURRICULUM AND MISSING-EVIDENCE CONTROLS

The development controls use the same MIMIC history-aware model as the locked forecast. Reconstruction replaces the first three of twelve training epochs rather than adding compute. Inference-only dropout uses probabilities 25%, 50%, and 75% for each observed clinical value and, independently, for each whole CXR channel. Patient windows are paired across arms, with three training seeds. No locked-test data are used.

Table 13: **Matched component ablations on MIMIC development.** Top: clinical RMSE, mean $\pm$ sample SD across three training seeds. Bottom: paired patient-level RMSE reduction; positive favors the named component. Each patient's window/horizon MSE is averaged before taking the square root, then averaged across seeds. These contrasts therefore have a different aggregation from the top-panel scores. Intervals use 5,000 patient-bootstrap resamples over 4,036 patients, conditional on the three seeds.

Time input	Training / input arm	Clinical RMSE ↓
Elapsed	Rollout only	0.7688 ± 0.0022
Elapsed	Reconstruction + rollout	0.7646 ± 0.0021
Paired component contrast		Gain [95% CI]
Reconstruction curriculum		0.00371 $[0.00196, 0.00535]$

Table 14: **Inference with missing source evidence.** Clinical RMSE, mean ± sample SD over the same three seeds. Source dropout masks each observed clinical value and drops each whole CXR channel at the stated probability, without retraining. Perturbations are paired across arms. These rows mask observed clinical values and drop whole CXR channels together. They are development robustness diagnostics.

Source perturbation	Elapsed-time rollout	With reconstruction
None	0.7688 ± 0.0022	0.7646 ± 0.0021
25% source dropout	0.8264 ± 0.0010	0.8239 ± 0.0015
50% source dropout	0.8891 ± 0.0007	0.8884 ± 0.0036
75% source dropout	0.9504 ± 0.0015	0.9500 ± 0.0055

C.6 TASK-ADAPTED VLM PROGNOSIS EVALUATION

Qwen2.5-VL-7B SFT is trained and scored on the clinical forecast interface: Clinical19 targets for MASS, and the 13-variable plus CXR interface for MIMIC.

C.7 MIMIC PREDICTIVE ACTION CONDITIONING

The MIMIC action-conditioned PULMOWORLD is a supervised world model whose input includes logged fluid and vasopressor doses. Table 16 reports factual-versus-zero and shuffled-action checks together with the authorization gate. Relative RMSE reductions versus the frozen state-only parent are 0.20% on all validation windows, 0.61% on action-exposed windows, and 0.79% at H4.

Table 15: **From action-conditioned prediction to authorized policy improvement.** The capability ladder separates an action representation, predictive action information (G_A), causal identification (G_I), and policy optimization. MASS-Syn tests Cap-MBRL after known-mechanism identification. MIMIC candidates are means of train-transition action vectors within train-fitted dose cells; no non-train row contributes to a prototype.

Capability and evidence ladder

World	Action representation	Action evidence	Identification G_I	Authorized treatment set	Planner / policy result and claim
Real MASS	10 mechanism blocks; 8 prespecified cells	$0/8$ cells pass the full gate	Closed: observational	$\varnothing$	Eight prespecified cells; planning is scored in MASS-Syn
MIMIC-IV-CXR	4×4 fluid–vasopressor cells; 16 train-only prototypes	$+0.61\%$ RMSE gain on exposed windows; $+0.79\%$ at H4; logged beats zero/shuffles	Closed: no causal design	$\varnothing$; no-action fallback a_0	Logged care changes the forecast; no nonzero action is authorized
MASS-Syn	Known-mechanism action space	$3/3$ signals open; $3/3$ matched nulls close	Ground-truth audit	Nonempty where supported	Cap-MBRL regret 0.0180 vs. fallback 0.0910; unauthorized action rate 0

The MIMIC candidate grid, dose thresholds, and replay protocol are in Table 16. Since $G_I = 0$, no non-zero candidate is authorized. The MASS-Syn value 0 is the rate of executed actions outside the permitted set. Bootstrap intervals and the MIMIC planner audit are in Table 16.

Table 16: **MIMIC action interface: prediction is enabled, control fails closed.** Fluid and vasopressor doses are discretized on the train split into 16 joint templates (4×4; thresholds 5.83/12.36 ml/kg and 0.065/0.148 mcg/kg/min NEE). Template 0 is no-action (75.8% of train transitions). Each candidate prototype is the within-cell mean of only those train-transition action vectors; calibration, validation, and test rows are excluded from prototype construction. Identification contrasts reuse the frozen action-conditioned PULMOWORLD on the same 2048-window / 1380-patient validation subsample as Cap-MBRL (without accessing the test set). Relative RMSE reduction compares logged doses with the named contrast on action-exposed windows ($n = 750$); the state-only parent is the frozen natural-history model, and the all-window parent reduction is 0.20%. Positive values mean logged doses are more predictive. The ranking searches the 16 templates by imagined death probability.

A. Frozen candidate-action pool

Fluid \ vasopressor	None	Low	Medium	High
None	a_0	a_1	a_2	a_3
Low	a_4	a_5	a_6	a_7
Medium	a_8	a_9	a_{10}	a_{11}
High	a_{12}	a_{13}	a_{14}	a_{15}

B. Predictive action-information contrasts (action-exposed)

Contrast vs logged doses	All-H RMSE red. ↑	H4 RMSE red. ↑
State-only parent	0.61% $[0.45, 0.80]$	0.79% $[0.59, 1.03]$
Zero action	0.82% $[0.24, 1.68]$	1.26% $[0.39, 2.37]$
Global shuffle	0.47% $[0.18, 0.79]$	0.73% $[0.24, 1.28]$
Support-matched shuffle	0.48% $[0.16, 0.84]$	0.66% $[0.21, 1.18]$

C. Fail-closed policy interface on the same windows

Method	Treat. rate	Unauth.	Imagined death P	Δ vs current
Cap-MBRL (hard gate)	**0.000**	**0.000**	0.136	-0.013
Logged factual sequence (audit)	0.366	0.366	0.139	-0.010
Unrestricted diagnostic planner	0.740	0.740	0.125	-0.024

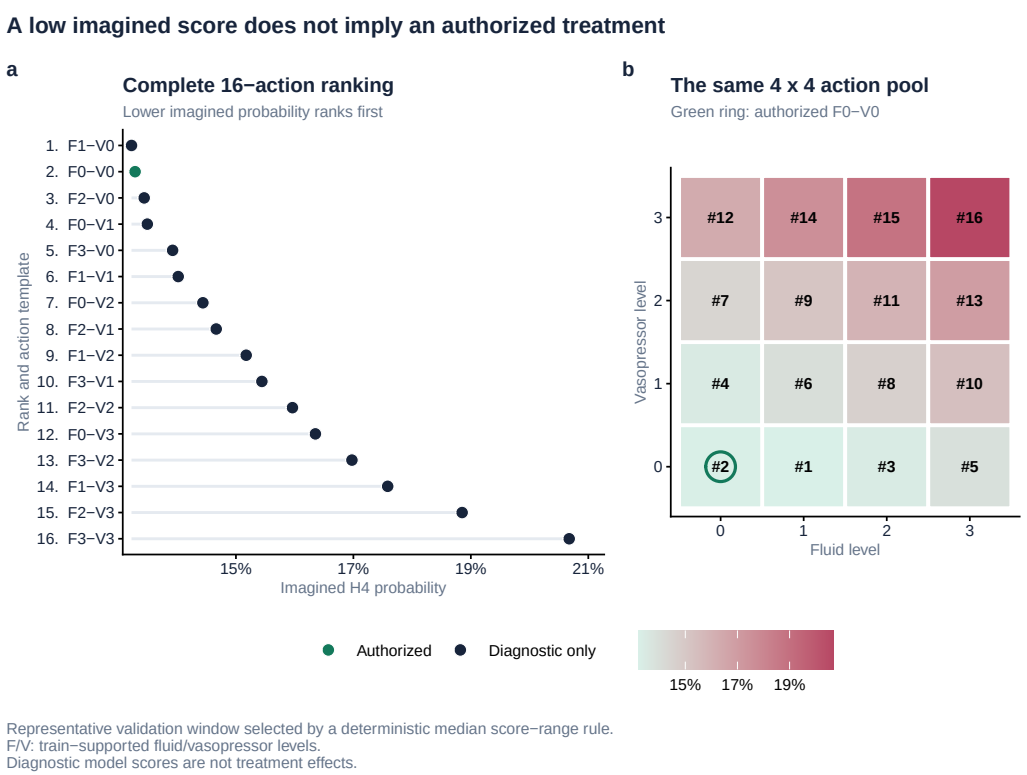


Figure 9: **A complete single-window ranking of all 16 MIMIC action templates.** (a) The frozen action-conditioned model ranks every train-derived fluid–vasopressor template by imagined H4 mortality probability for one representative validation window. The case is selected deterministically as the median score-range window among those whose imagined optimum is not the no-op fallback template. (b) The same ranks in the 4×4 action grid. F1–V0 has the lowest model score, and the displayed branch retains F0–V0.

C.8 AUTHORIZATION-GATED SELECTION ON KNOWN-MECHANISM WORLDS

The policy comparison scores selection after the identification gate opens on MASS-Syn. The confirmatory experiment compares unrestricted and penalty-based world-model RL, an authorized fallback, an authorized soft-penalty policy, and risk-budgeted policy improvement (Table 17 and Figure 5). The aggregate across five independent environments met the prespecified confirmatory criteria; individual environments are replicates, not separate confirmatory studies.

Table 17: **Authorization-gated world-model RL on known-mechanism MASS-Syn.** Means over five independent simulator-parameter and split seeds in the confirmatory experiment; the aggregate met the prespecified criteria. Lower is better. Selection regret is policy error within the authorized set. Rows with nonzero unauthorized action rates can choose actions outside the permitted set, so their selection regret is not comparable. The authorized soft-penalty policy combines the mask with uncertainty, support, and path-consistency terms, without Cap-MBRL's budgeted candidate selection. Among rows with zero unauthorized action rate, bold and underline mark the best and second-best selection regret. Rates are fractions, not percentages.

Method	Selection regret	Unauth. rate
Diagnostic controls without hard authorization		
Unrestricted WM-RL	–	0.3219
Uncertainty-aware	–	0.3226
Support-aware	–	0.3371
Path-consistency weighting	–	0.3158
Policy comparison within the same authorized action set		
Authorized + soft penalties	0.3636	**0.0000**
Authorized-set fallback	<u>0.0910</u>	**0.0000**
Cap-MBRL	**0.0180**	**0.0000**
No action	–	0.0000

Interpreting the policy controls. The ungated controls test whether uncertainty, support, or path-consistency penalties can replace explicit authorization. Their lower unrestricted regret is not evidence of better selection within the authorized set. The gated comparison holds that set fixed. Cap-MBRL reduces mean selection regret from 0.0910 to 0.0180. Across the five environments, selection regret improves in four and is unchanged in one.

C.9 ACTION-CONDITIONED BRANCHES AND AUTHORIZATION EVIDENCE

The main-text case uses the first identity-hash-ordered validation window with all three displayed source variables observed (212 eligible windows among 2,048). Selection does not use future measurements or model outputs. Fluid templates pool recorded crystalloid boluses (normal saline and lactated Ringer's); vasopressors use norepinephrine-equivalent rates, not a drug-specific intervention. For Actions 1–4, the displayed fluid/rate pairs are $(0, 0)$, $(0, 0.252)$, $(18.710, 0)$, and $(20.904, 0.292)$, respectively, in mL/kg and μg/kg/min. These are inverse transforms of mean training-set templates in model space, not mean administered doses. The combined template therefore differs slightly in dose from the two single-channel templates. The same template is applied in every six-hour rollout interval.

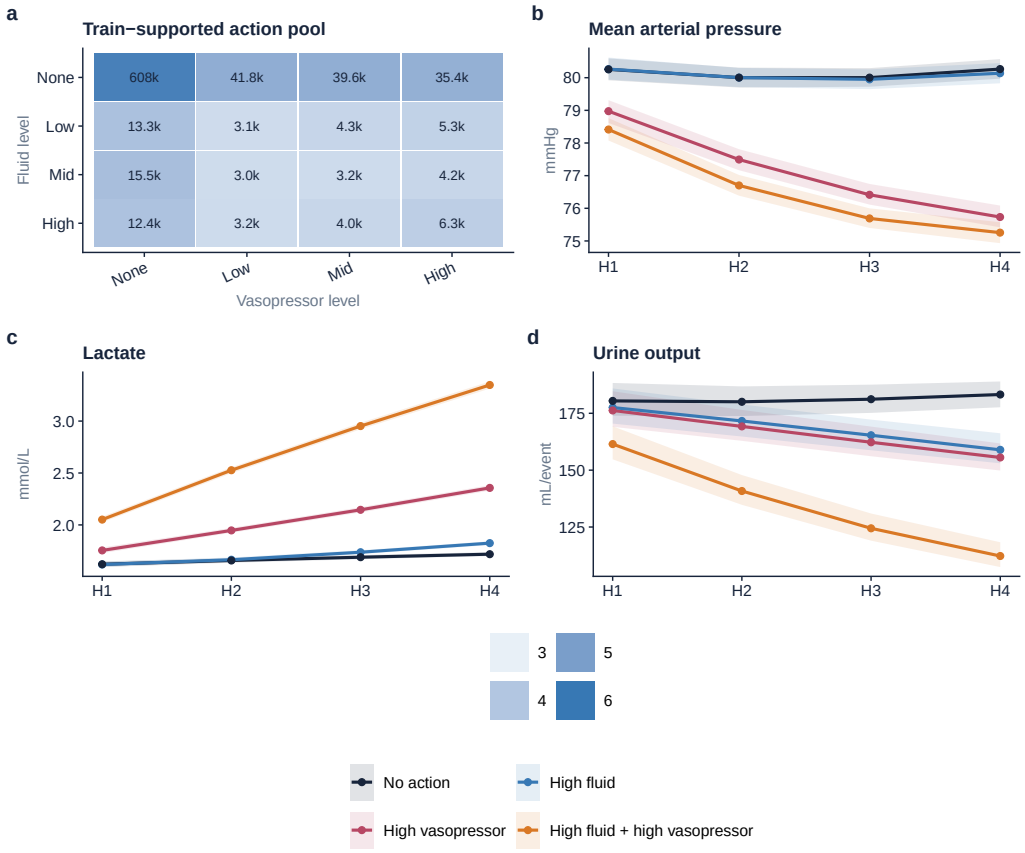


Figure 10: **What an action is and how it changes a frozen simulation.** (a) The MIMIC action pool is a 4×4 train-supported grid over fluid and vasopressor levels; annotations give train-transition counts. The four corners define the diagnostic candidates replayed in panels (b–d). Repeating these four train-derived templates over H1–H4 separates predicted mean arterial pressure, lactate, and urine-output trajectories. Lines are patient-level means over 2,048 validation windows and bands are patient-cluster 95% CIs. This action sensitivity is predictive input to the gate.

D CASE ATLAS: INPUTS, FUTURE STATES, AND FAILURE MODES

D.1 MATCHED SOURCE AND NEXT-STATE RECORDS

The case atlas connects observed clinical and imaging evidence to model inputs, recursive predictions, and their limitations. It first presents matched records, then a complete same-patient forecast, followed by additional clinical, imaging, and metabolomic examples. The images are observed scans, not reconstructions or model generations. The numbers are canonical physical-unit values before train-fitted normalization; "missing" remains missing through an explicit mask. Exact drug names and identifiers are mapped to privacy-safe action families before modeling.

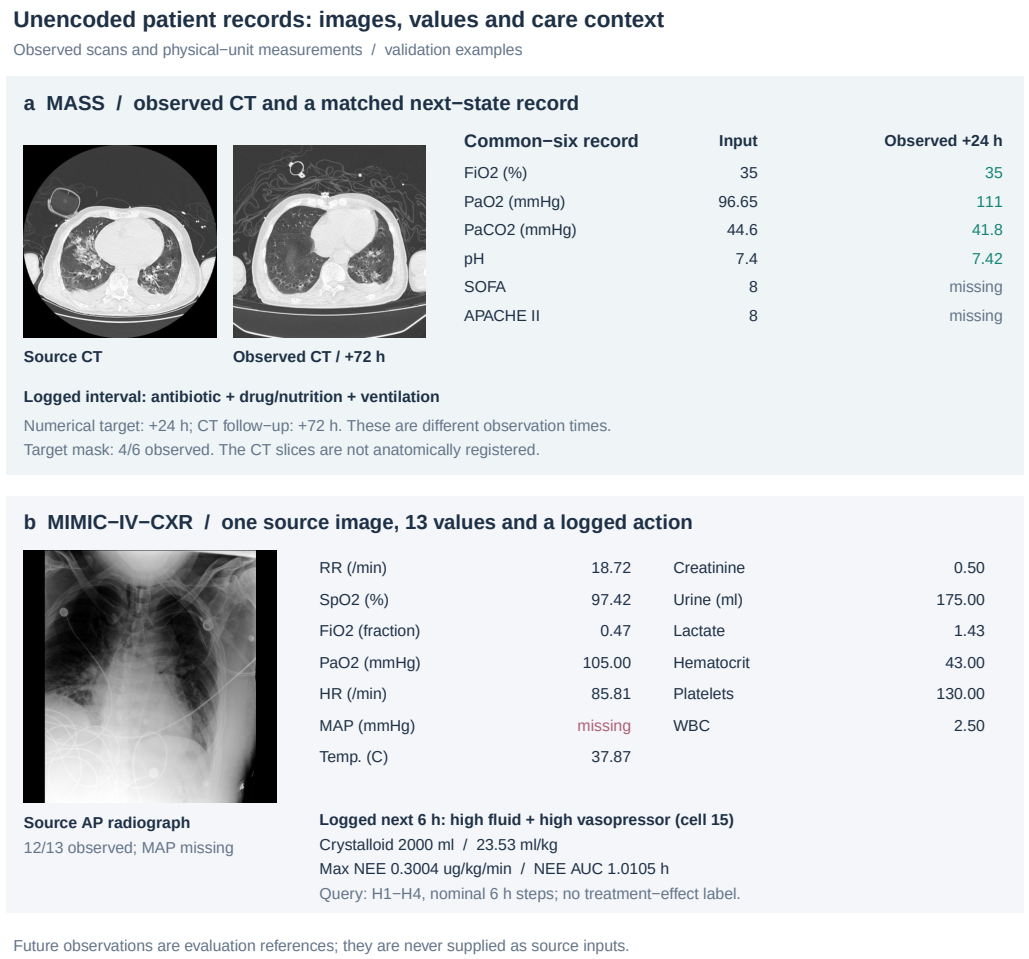


Figure 11: **Integrated pre-encoder case records.** (a) A MASS validation CT pair is shown beside its matched Clinical19 source and observed next-state record. Numerical references are at +24 h, whereas the observed CT follow-up is at +72 h. Missing target cells remain missing; neither images nor target values are generated. The fixed lung-window slices need not depict identical anatomy and do not support pixelwise comparison. The full Clinical19 field list is in Table 18. (b) A MIMIC validation radiograph, all 13 source clinical values, the missing MAP mask, and its logged fluid/vasopressor context are grouped together. Images retain their grayscale intensities; fitting and padding preserve aspect ratio. Future observations are evaluation references only.

D.2 CLINICAL AND ACTION RECORDS BEFORE ENCODING

The examples below report the measurements supplied to the model. They preserve physical units, missing-value masks, event timing, and action semantics; they do not contain PCA coordinates, normalized values, or latent states. MASS exposes all 19 source clinical variables, whereas MIMIC-IV-CXR exposes 13 variables and retains a false mask for the missing MAP cell.

Table 18: **Full pre-encoder clinical records.** MASS shows the Clinical19 extraction (19/19 observed), the same interface as the action evaluation record in Figure 11; MIMIC shows the 13-field source record in that figure (12/13 observed). Missing MAP remains masked. Units are shown where documented for these records.

MASS: Clinical19		MIMIC-IV-CXR: 13 clinical	
Variable	Source value	Variable	Source value
FiO_2 (fraction)	0.35	FiO_2 (fraction)	0.467
PaO_2 (mmHg)	91.3	PaO_2 (mmHg)	105
$PaCO_2$ (mmHg)	44.6	SpO_2 (%)	97.42
pH	7.40	Respiratory rate (/min)	18.72
Respiratory rate (/min)	12	Heart rate (/min)	85.81
Lactate (mmol/L)	1.50	Lactate	1.43
MAP (mmHg)	70	MAP (mmHg)	missing
Heart rate (/min)	165	Temperature (°C)	37.87
Temperature (°C)	35.3	Creatinine	0.50
Platelets	28	Platelets	130
WBC	9.37	WBC	2.50
Creatinine	88	Urine output (ml)	175
Bilirubin	16.4	Hematocrit	43
SOFA	8		
Renal SOFA	0		
APACHE II	8		
CRP	28.7		
PCT	0.45		
BUN	13.02		
Observed mask	19/19	Observed mask	12/13

Time and action interface. The MASS action evaluation example queries the next event after 24 h under logged antibiotic, drug/nutrition, and ventilation context; the supported alternative retains drug/nutrition and ventilation only and is a simulation query without an observed counterfactual target. MIMIC queries H1–H4 at nominal 6 h steps. Its logged cell 15 contains high fluid and high vasopressor exposure: 2000 ml crystalloid (23.53 ml/kg), maximum NEE 0.3004 μg/kg/min, and NEE AUC 1.0105 h. Alternative candidates come from the train-supported 4×4 pool; ranking is diagnostic unless the authorization gate opens. Radiology reports are not model inputs.

D.3 FROM RAW RECORDS TO MODEL INTERFACES

MASS supplies Clinical19 values and masks, an optional CT, an optional BALF metabolomic profile, elapsed time, and action context at each asynchronous event. The full action representation has 20 macro classes with presence, interval-overlap fraction, source-time activity, and within-interval initiation channels; the action labels shown here form the coarser case-display interface. MIMIC supplies 13 clinical values and masks, an optional CXR, elapsed time, and four continuous fluid/vasopressor quantities mapped to a train-fitted 16-cell candidate pool. The displayed reports aid human audit only. Action fields in both cohorts record logged care.

D.4 A COMPLETE SAME-PATIENT MULTIMODAL INPUT–OUTPUT EXAMPLE

This example links the existing complete-CXR validation window to all clinical inputs and forecasts from the same patient, admission, and source state. The window was selected for radiograph coverage, not prediction accuracy. The source contains seven observed clinical variables, six masked variables, and an observed CXR representation.

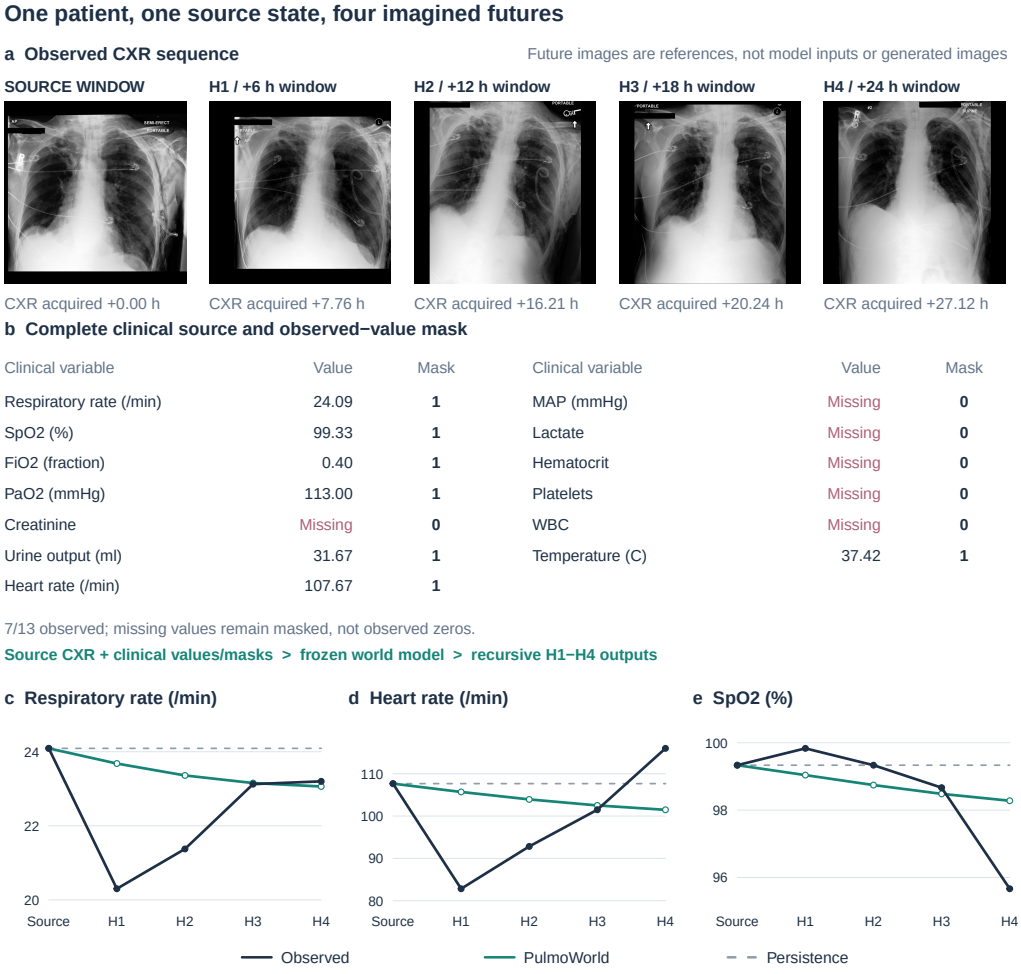


Figure 12: **From one multimodal source record to recursive future outputs.** (a) Observed AP radiographs from the source and four future windows. Acquisition offsets are relative to the source radiograph, whereas horizon labels denote nominal six-hour event-window offsets. One image is displayed per window; the underlying image counts are 1, 1, 2, 2, and 3. The model uses the pooled window representation, not only the displayed view. (b) All 13 clinical input fields and their observed-value masks. (c–e) Three clinical coordinates decoded from the same four-step rollout, with observed references and source-value persistence. All other coordinates are retained in Table 19. Only the source observations initialize the rollout; future images and measurements serve as evaluation references. Images retain their grayscale intensities with aspect-preserving fitting and padding. This single-patient illustration has no calibrated individual prediction intervals.

Complete numerical outputs and missing-reference handling. The frozen natural-history model executes four six-hour transitions without future observation updates. Its zero model-action vectors are an evaluation setting, not evidence that the patient received no treatment. Table 19 reports every clinical decoder mean, including coordinates whose future reference is unavailable. Missing references are never replaced with zero for evaluation.

Table 19: **Complete clinical outputs for the same multimodal case.** Each future entry is prediction / observed reference (P / O); "–" means unobserved, with reference mask zero. Source values use physical units before normalization. Decoder means are shown even when predicted availability is below 0.5; these means are not treated as available inputs at the next step. Unrounded values, normalized inputs, masks, availability probabilities, and observed-cell errors are retained in the accompanying source data.

Variable	Source	H1: P / O	H2: P / O	H3: P / O	H4: P / O
Respiratory rate (/min)	24.09	23.68 / 20.30	23.36 / 21.38	23.15 / 23.12	23.06 / 23.20
SpO_2 (%)	99.33	99.04 / 99.83	98.74 / 99.33	98.48 / 98.67	98.28 / 95.67
FiO_2 (fraction)	0.40	0.40 / 0.40	0.40 / 0.40	0.40 / 0.40	0.40 / 0.40
PaO_2 (mmHg)	113.00	100.99 / –	96.96 / –	97.35 / 161.00	97.74 / –
Creatinine	–	1.29 / –	1.32 / –	1.35 / 0.60	1.38 / –
Urine output (ml)	31.67	45.09 / 30.83	57.67 / 48.75	68.89 / 143.33	78.46 / 125.00
Heart rate (/min)	107.67	105.70 / 82.83	103.94 / 92.83	102.52 / 101.50	101.48 / 116.00
MAP (mmHg)	–	79.72 / –	80.03 / –	80.16 / –	80.15 / –
Lactate	–	1.57 / –	1.60 / –	1.63 / –	1.65 / –
Hematocrit	–	28.10 / –	27.52 / –	27.04 / 28.60	26.65 / –
Platelets	–	217.71 / –	220.57 / –	224.20 / 212.00	228.82 / –
WBC	–	12.49 / –	12.54 / –	12.58 / 9.90	12.60 / –
Temperature (°C)	37.42	37.37 / 37.28	37.32 / 37.81	37.28 / 37.44	37.26 / 37.19

What the mask and output probabilities mean. Observed FiO_2 equals 0.40 and maps to a normalized value of zero with mask one. Missing MAP also has a zero placeholder, but mask zero distinguishes it from an observed value. For example, the H1 MAP decoder mean is 79.72 mmHg with predicted observation probability 0.275; no MAP reference is observed at any displayed horizon. This probability describes availability, not the probability that 79.72 mmHg is accurate. The decoded mean is an availability-conditioned output.

Table 20: **CXR representation outputs from the same rollout.** RMSE is over 64 train-standardized PCA coordinates, not image pixels. Persistence repeats the source representation. These are the same image-state predictions already used in the complete-CXR case; no new image selection is performed.

CXR representation error	H1	H2	H3	H4
PulmoWorld	0.902	0.780	0.939	0.756
Persistence	1.228	1.258	1.318	1.129

A separate gated-decision output. The MIMIC validation action-ranking case in Figure 9 is a different patient example. Among 16 fluid–vasopressor templates, F1–V0 has the lowest imagined mortality score (13.222%), versus 13.283% for F0–V0. Identification evidence is absent, so the reported branch retains F0–V0. Authorized-set selection is evaluated in MASS-Syn (Table 17).

D.5 DIVERSE SOURCE, FOLLOW-UP, AND ACTION-CONTEXT EXAMPLES

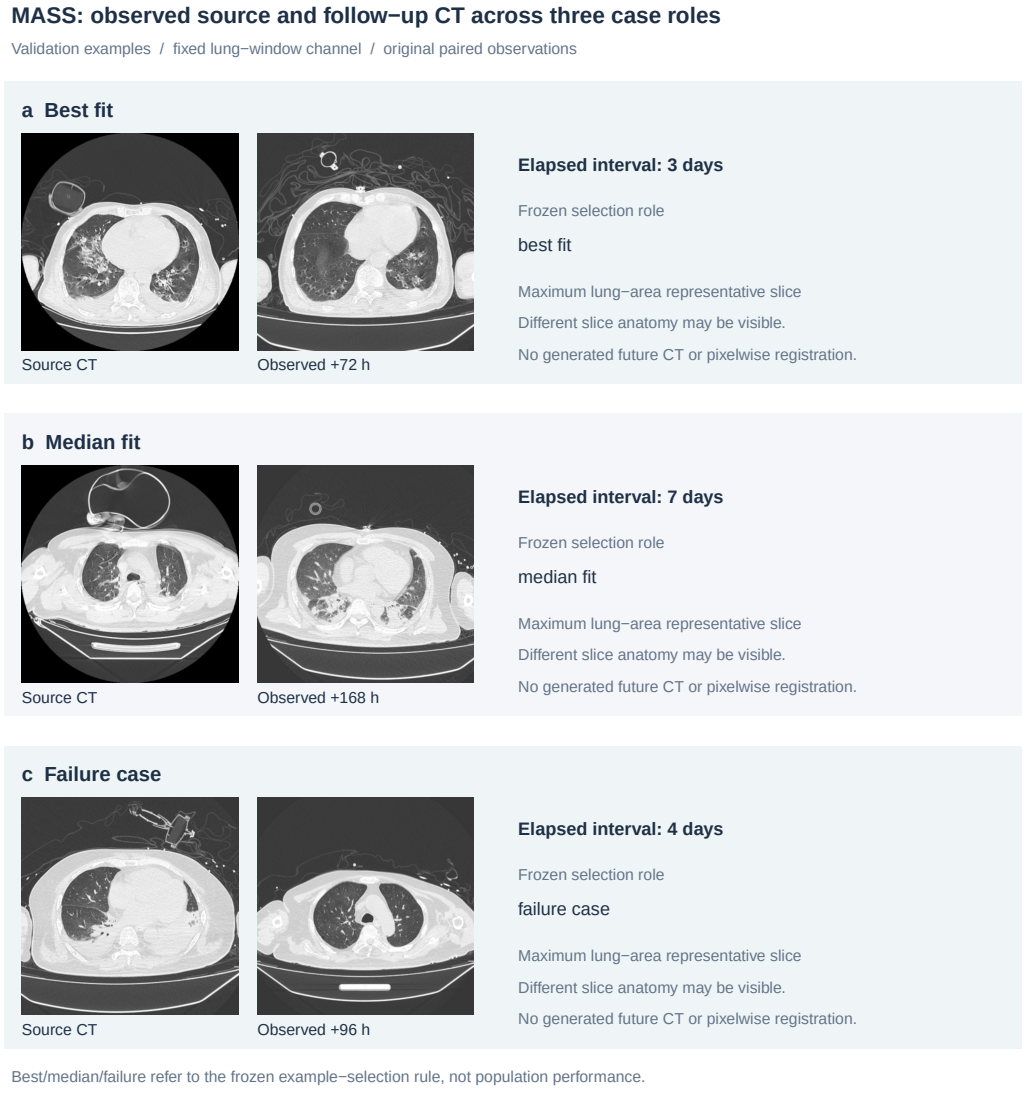
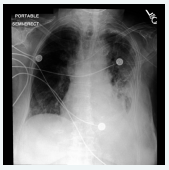


Figure 13: **Three MASS temporal CT examples.** The frozen best-fit, median-fit, and failure-case roles are all retained. Source and future scans are observed validation images, separated by 3, 7, and 4 days, respectively. Only the original lung-window channel is displayed. Representative slices are not anatomically registered.

MIMIC: diverse observed inputs and logged care contexts

Four validation examples / all 13 clinical fields / missing cells retain their masks

a No recorded hemodynamic action



RR (/min)	18.38	MAP (mmHg)	missing	**Cell 0 / next 6 h**
SpO2 (%)	95.12	Lactate	0.80	Fluid 0 ml
FiO2	missing	Hematocrit	29.30	0.00 ml/kg
PaO2 (mmHg)	missing	Platelets	100.00	Max NEE 0.0000
Creatinine	0.80	WBC	5.40	AUC 0.0000 h
Urine (ml)	80.00	Temp. (C)	35.04	
HR (/min)	63.62			

b High vasopressor



RR (/min)	20.09	MAP (mmHg)	97.11	**Cell 3 / next 6 h**
SpO2 (%)	99.70	Lactate	3.10	Fluid 0 ml
FiO2	0.40	Hematocrit	36.80	0.00 ml/kg
PaO2 (mmHg)	215.00	Platelets	44.00	Max NEE 0.1601
Creatinine	1.20	WBC	10.00	AUC 0.3308 h
Urine (ml)	143.75	Temp. (C)	37.58	
HR (/min)	120.56			

c High fluid



RR (/min)	17.64	MAP (mmHg)	missing	**Cell 12 / next 6 h**
SpO2 (%)	99.17	Lactate	1.90	Fluid 1000 ml
FiO2	0.45	Hematocrit	missing	15.22 ml/kg
PaO2 (mmHg)	missing	Platelets	missing	Max NEE 0.0000
Creatinine	0.70	WBC	missing	AUC 0.0000 h
Urine (ml)	583.33	Temp. (C)	37.18	
HR (/min)	116.00			

d High fluid + high vasopressor



RR (/min)	18.72	MAP (mmHg)	missing	**Cell 15 / next 6 h**
SpO2 (%)	97.42	Lactate	1.43	Fluid 2000 ml
FiO2	0.47	Hematocrit	43.00	23.53 ml/kg
PaO2 (mmHg)	105.00	Platelets	130.00	Max NEE 0.3004
Creatinine	0.50	WBC	2.50	AUC 1.0105 h
Urine (ml)	175.00	Temp. (C)	37.87	
HR (/min)	85.81			

Max NEE: ug/kg/min. Logged care is conditioning context, not an identified treatment effect.

Figure 14: **Four MIMIC input cases with different logged care contexts.** Each row combines the original source radiograph, all 13 clinical values, missingness, and the following six-hour action record from the same validation case. The examples cover no recorded hemodynamic action and different fluid/vasopressor exposures. Dose summaries are factual context; these rows contain no observed counterfactual outcomes or treatment recommendations. They are distinct from the main-text forecasting cases.

D.6 Clinical case selection and image-state comparison

Selection of the main-text clinical examples. The respiratory-rate, SpO_2, and MAP panels in Figure 4 show separate MIMIC patients. Eligibility requires observed source and H1–H4 values, nonconstant ground truth, lower four-step MAE for PulmoWorld than each of GRU, causal Transformer, and persistence, source-inclusive correlation at least 0.8, and MAE divided by the observed range at most 0.25. Among eligible cases, selection minimizes the worst displayed method's normalized MAE, rather than maximizing the winning margin. Nine respiratory-rate, nine SpO_2, and one MAP trajectory meet these outcome-dependent criteria. Their PulmoWorld MAEs are 1.794

breaths/min, 0.630 percentage points, and 1.441 mmHg, respectively. Table 2 reports the locked-test comparison.

Additional trajectory examples. Among 13 complete MASS PaO_2 trajectories, one has lower four-step MAE for PULMOWORLD than both the causal Transformer and CLARITY. Its observed H1–H4 values are 68.50, 79.00, 106.00, and 75.00 mmHg, whereas the predictions are 88.74, 89.55, 86.49, and 84.80 mmHg (MAE 15.03 mmHg). The unobserved source value prevents a persistence comparison. For MIMIC, median-error examples among PULMOWORLD wins over GRU, causal Transformer, and persistence yield MAEs of 0.98 percentage points for SpO_2 and 4.91 mmHg for MAP. There are 62 such wins among 296 eligible SpO_2 trajectories and 18 among 93 eligible MAP trajectories.

A complete CXR sequence. The full observed sequence and its same-patient clinical forecasts appear in Figure 12, with complete numerical outputs in Tables 19 and 20. This validation window was selected for complete image coverage rather than forecast accuracy. Event-window labels are not exact acquisition times; multiple views contribute to some pooled window representations. All four methods in the main-text image-error panel are scored in the same train-standardized, 64-dimensional coordinate system. CLARITY predictions are transformed from raw PCA coordinates before scoring; the train mean is zero in this space. PULMOWORLD improves on CLARITY and source persistence in this example. The comparison concerns representation prediction, not image synthesis.

D.7 IMAGING AND METABOLOMIC EVOLUTION

Locked-test recursive forecasting scores clinical variables only. Imaging evidence is therefore reported here through distinct diagnostics: MASS CT/BALF latent errors and clinical forecast errors using frozen MIMIC CXR encoders (Table 21). The subsequent CXR figure separately scores image latents. Direct VLMs are omitted from the MASS latent panel because they never emit CT64 or BALF64. The reported scores are representation forecasts for those modalities.

Together, the following analyses distinguish forecasting a compact modality representation from recovering an image or identifying an individual metabolite. Aggregate errors are followed by spatial CT diagnostics, CXR coverage and error summaries, and longitudinal BALF examples. The latter also show how sparse repeat sampling limits later-horizon evaluation.

Spatial CT diagnostics. The CT examples below complement vector-level forecast errors by comparing observed and predicted token changes. They include a favorable, an intermediate, and an unfavorable case; representative slices are not anatomically registered. These displays assess representation dynamics rather than image synthesis.

Table 21: **Development and validation of imaging and metabolomic representations.** MASS CT64/BALF64 are Clinical19 development rollouts (three-seed mean standardized MSE; lower is better). Direct VLMs never emit those latents, so they are omitted rather than shown as dashes. MIMIC reports clinical forecast RMSE using frozen CXR PCA64 encoders and independently fitted world models on validation without accessing test patients. These are fitted-encoder comparisons, not prompted numerical forecasts or the 300-patient locked-test comparison. These latent forecasts are separate from locked-test recursive forecasting, which uses the Clinical19 and 13-variable clinical targets. Bold and underline mark the best and second-best displayed values in each column.

MASS CT64 / BALF64 (development)

Method	CT H1↓	CT H2↓	CT H3↓	CT H4↓	BALF H1↓	BALF H2↓	BALF H3↓	BALF H4↓
GRU	<u>0.868</u>	0.848	<u>0.884</u>	<u>0.894</u>	<u>0.889</u>	<u>0.926</u>	<u>0.792</u>	<u>0.913</u>
Causal Transformer	0.884	0.862	0.898	0.900	0.893	0.934	0.810	0.928
Deterministic WM	0.895	0.879	0.941	0.986	0.897	0.964	0.847	1.000
RSSM	0.881	<u>0.846</u>	0.898	<u>0.894</u>	0.896	0.932	0.803	0.922
PULMOWORLD	**0.836**	**0.834**	**0.875**	**0.881**	**0.836**	**0.921**	**0.785**	**0.912**

MIMIC clinical forecasting with frozen CXR encoders (validation RMSE)

Model	Source input	Full validation↓	CXR-paired↓
Persistence	Clinical	0.8625	0.8755
GRU	Clinical	0.7711	0.7456
Causal Transformer	Clinical	0.7686	0.7480
GRU-D	Clinical	0.9289	0.9359
Qwen2.5-VL-7B + fitted WM	Clinical+CXR64	0.7653	0.7448
Qwen2.5-VL-72B + fitted WM	Clinical+CXR64	**0.7651**	**0.7438**
Qwen3-VL-30B-A3B + fitted WM	Clinical+CXR64	<u>0.7652</u>	<u>0.7442</u>

MASS numbers are the CT64/BALF64 block of the Clinical19 development three-seed comparison. MIMIC encoder rows use frozen CXR encoders and a fitted clinical head on validation. They are not the locked 13-variable forecast and they are not Direct-VLM forecasts. Remaining encoder variants sit in 0.7649–0.7654 on the full validation split.

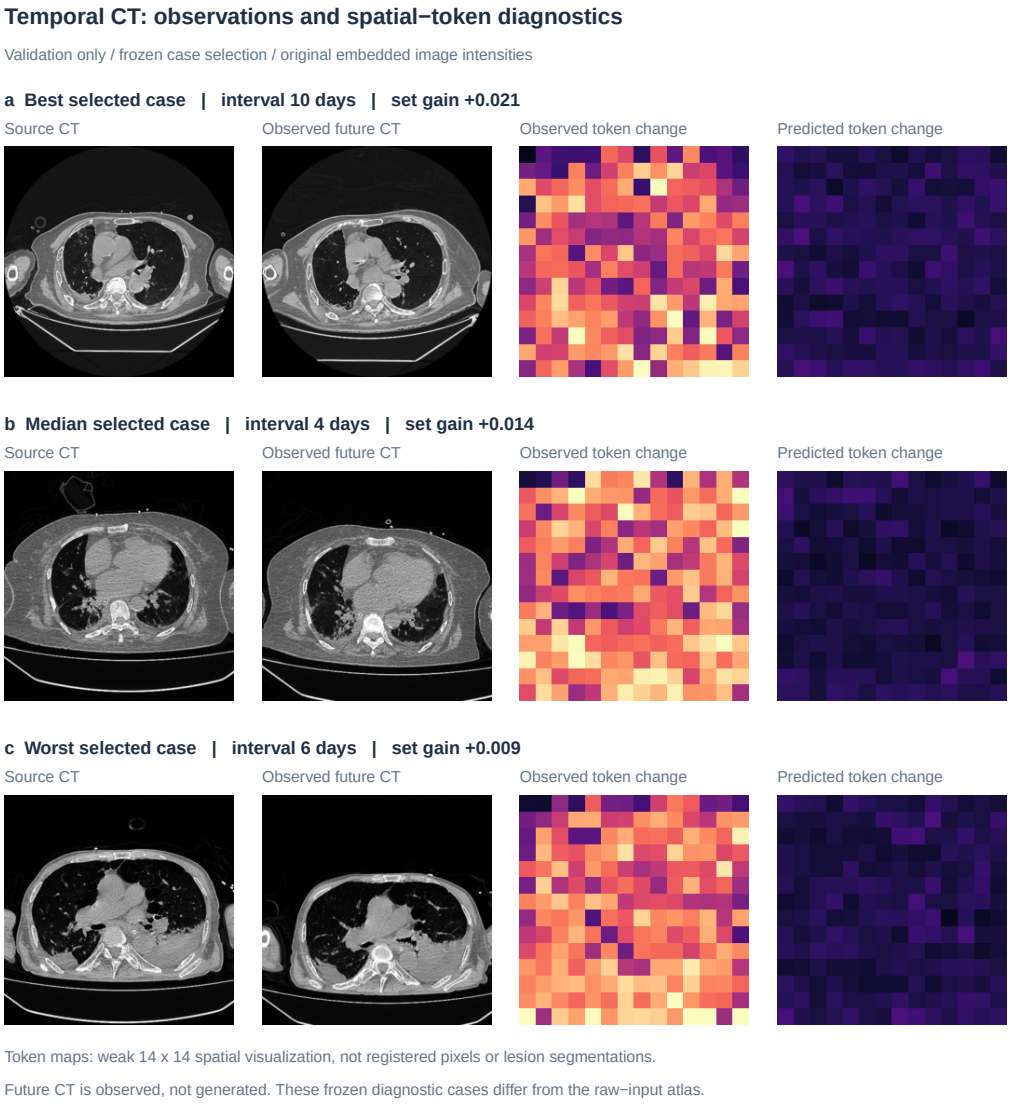


Figure 15: **MASS CT observations and token-space evolution.** Validation CT pairs (best / median / worst set-distance gain after an SSIM screen). Columns are current 2D CT, observed future 2D CT, observed token change, and predicted token change. Future CT is observed, not generated. No locked-test patient is shown. These are the frozen spatial-token diagnostic cases, separate from the raw-input atlas.

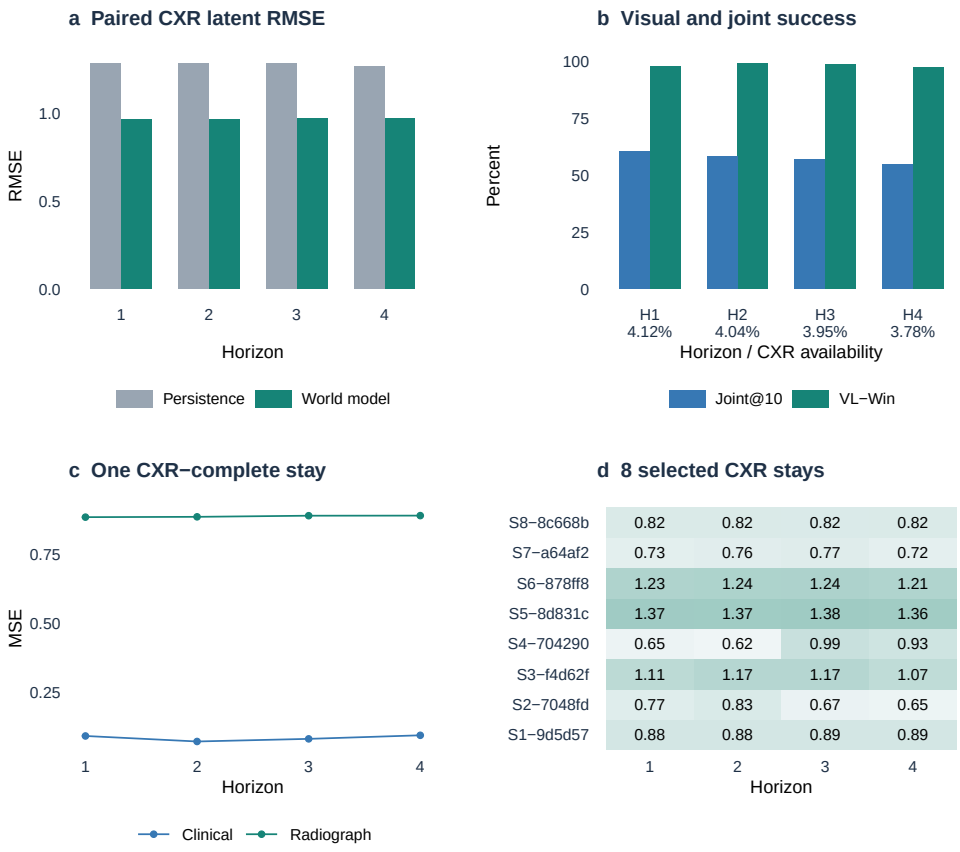


Figure 16: **MIMIC paired-CXR latent evolution.** Validation data, separate from the 300-patient locked test; observed rather than generated films. (a) Frozen Qwen2.5-VL-7B encoder + world model versus source-film persistence. (b) VL-Win is the fraction of source/target-CXR pairs whose predicted representation is closer to the target than persistence. Joint@10 is the fraction of eligible clinical cells with at most 10% source-normalized error and a simultaneous visual win. The error denominator is the larger of the absolute source value and 0.1 times the training IQR (with a numerical floor of 10^{-8}); these rates have different denominators. Tick labels give the share of windows with a target CXR (≈4%). (c) One hashed stay with a radiograph at every horizon. (d) Eight selected low-clinical-MSE stays with CXR at H1–H4. Stay labels are pseudonyms, not MIMIC identifiers.

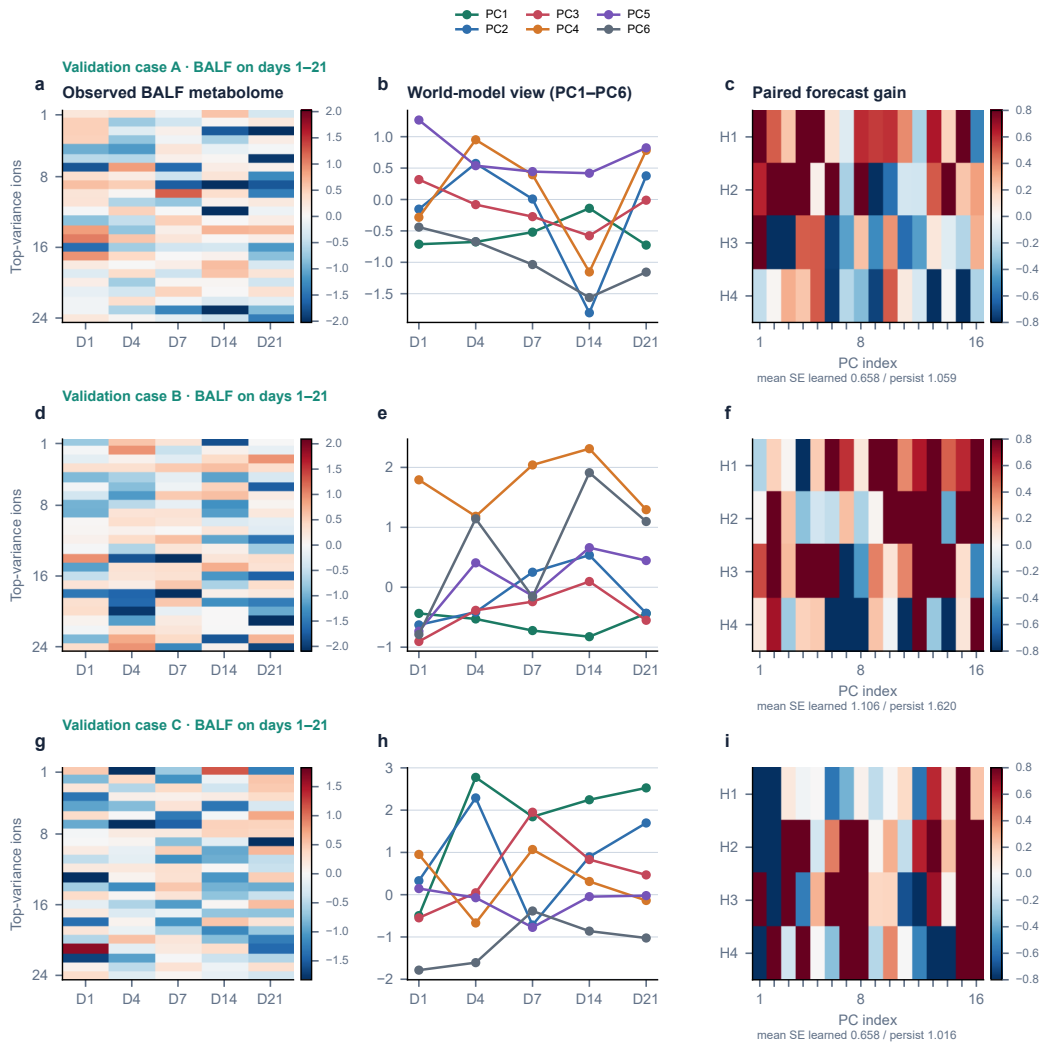


Figure 17: **Observed BALF metabolomic sequences and predicted state changes.** Three development-validation patients with five BALF visits on protocol days 1/4/7/14/21 (3 of 63 validation patients). (a,d,g) 24 highest-variance ions after the train-only log1p+IQR map; rows index selected ions and colorbars report transformed intensities. (b,e,h) PCA64 scores for the same visits (PC1–PC6), with a shared legend. (c,f,i) recursive forecasting paired MSE gain versus carry-forward on PC1–PC16; positive means PULMOWORLD is closer than persistence. The annotations report mean squared errors for the learned forecast and persistence. Direct VLMs do not emit BALF64. Validation examples only, not the locked-test recursive forecasting ranking.

E Reproducibility and Data Access

Materials and access. We will release evaluation code, preprocessing specifications, synthetic mechanisms, and aggregate results. MIMIC-IV and MIMIC-CXR must be obtained under the PhysioNet data-use agreement (Johnson et al., 2023; 2019); we do not redistribute these records. MASS was used with permission after personal identifying information was removed. Its patient-level clinical, CT, and BALF records are not part of the public release. Any qualified request for controlled access remains subject to institutional permission and the applicable data-use conditions.

Evaluation separation. The locked MASS and MIMIC tests were evaluated after the respective protocols were fixed. Test labels were excluded from training, model selection, calibration, and threshold fitting. Development ablations and history-enabled prognosis use separate validation populations, as specified in Section C.5. Uncertainty estimates resample patients for clinical comparisons; MASS-Syn policy comparisons instead summarize independent simulated environments. These sources of uncertainty should not be interpreted interchangeably.

Scope of reproducibility. The appendix specifies cohort construction, model dimensions, loss functions, training schedules, comparator interfaces, and case-selection criteria. Clinical forecasting, representation prediction, and synthetic policy selection remain distinct evaluations: Cap-MBRL is tested on known-mechanism MASS-Syn, while real-cohort action analyses assess care-conditioned forecasts and authorization.