PATIENTWORLD: A PHYSIOLOGY-INFORMED WORLD MODEL FOR LEARNING PATIENT DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting how a patient's physiology evolves over time under treatment is fundamental to clinical decision making. Longitudinal electronic health records (EHRs) provide large-scale physiological and treatment trajectories from which patient dynamics can be learned. However, many clinical trajectory models either focus on direct future-state prediction without explicitly modeling treatment-conditioned transitions or encode heterogeneous physiological variables into shared representations, leaving physiological system organization and cross-system dynamics implicit. We introduce PatientWorld, a physiology-informed world model for learning reusable treatment-conditioned patient dynamics. PatientWorld organizes latent states by physiological systems, explicitly models generic treatments and cross-system effects, uses physiological priors to structure the effects of physiology-specific treatments, and models state-dependent local transition residuals. During pretraining on broad ICU trajectories, it learns cross-system interactions among physiological systems; downstream, the pretrained dynamics remain fixed while task-specific context adaptation uses task-relevant physiological relations and action-conditioned residuals to specialize transition contexts. Across MIMIC-III and MIMIC-IV, PatientWorld outperforms compared methods in nearly all multivariate trajectory forecasting settings, improves downstream adaptation over alternative reuse strategies, and achieves the highest mean on the primary decision metric among the compared world models on both tasks.

1 INTRODUCTION

Predicting how a patient's physiology evolves over time under treatment is central to sequential decision-making in intensive care. Such decisions depend not only on the patient's current state, but also on how that state may evolve over multiple steps under a given treatment sequence. Learning these patient dynamics therefore provides an important foundation for clinical trajectory modeling and sequential decision support (Raghu et al., 2018; Xu et al., 2026). As illustrated in Fig. 1(a), longitudinal electronic health records (EHRs) record multi-system physiological states together with administered treatments, yielding large-scale ICU observation–treatment trajectories (Johnson et al., 2016b; 2023b) from which such dynamics can be learned.

However, existing clinical trajectory models face two key limitations in learning these dynamics, as illustrated in Fig. 1(b). **First,** many longitudinal EHR and multivariate trajectory models primarily focus on directly predicting future clinical events or physiological states (Yang et al., 2023; He & Chiang, 2025; Chang et al., 2024). Even when treatment history is incorporated as time-varying information, the modeling objective is typically not to explicitly learn transitions conditioned on treatment actions that can be recursively rolled out along a given treatment sequence. **Second,** many longitudinal EHR and multivariate trajectory models integrate heterogeneous variables from different physiological systems into shared patient representations or generic temporal representations (Rasmy et al., 2021; Steinberg et al., 2021; Yang et al., 2023). Although such representations can capture complex statistical dependencies, states and dynamics from different systems are mixed within the same representation, without explicitly distinguishing system-specific dynamics or modeling how physiological systems interact. Consequently, both treatment-conditioned dynamics and physiological organization remain largely implicit in existing trajectory modeling.

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.

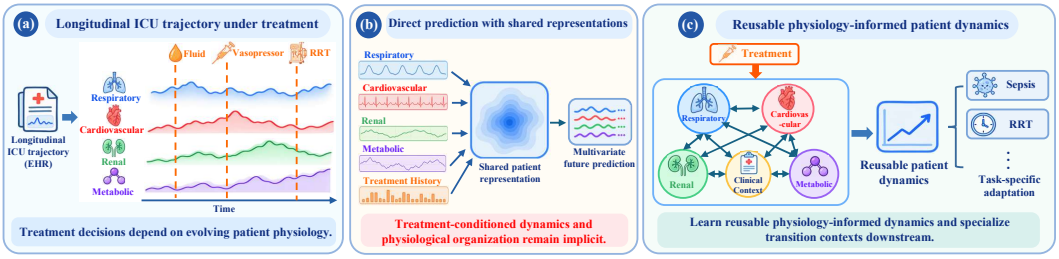


Figure 1: **Motivation for reusable physiology-informed patient dynamics.** (a) Longitudinal ICU trajectories record multi-system physiology and administered treatments. (b) Many clinical trajectory models focus on direct future-state prediction and integrate heterogeneous physiological signals into shared representations, leaving treatment-conditioned dynamics and physiological organization implicit. (c) PatientWorld learns reusable treatment-conditioned dynamics organized around physiological systems and applies task-specific context adaptation for downstream clinical tasks.

To address these limitations, we introduce PatientWorld, a physiology-informed latent world model for learning reusable treatment-conditioned patient dynamics, as shown in Fig. 1(c). Rather than relying on a shared patient representation, PatientWorld maintains separate latent states for major physiological systems alongside a global context, and learns how other physiological systems contribute to each system's transition without prespecifying a fixed interaction structure. Treatment actions condition latent transitions; the model incorporates the impacts of generic treatments, and for a subset of treatments with relatively clear physiological associations, it further uses physiological priors to structure the effects of physiology-specific treatments. In addition, we seek to reuse task-agnostic patient dynamics across clinical decision tasks while allowing task-specific information to modify the transition context rather than the pretrained dynamics backbone. We adopt a two-stage learning strategy: we first pretrain task-agnostic patient dynamics on ICU observation–treatment trajectories, and then keep the pretrained dynamics backbone fixed while applying task-specific context adaptation, which combines task-relevant physiological relations with action-conditioned residuals to specialize downstream transition contexts. Our main contributions are:

- **Physiology-informed treatment-conditioned patient dynamics.** PatientWorld factorizes patient states by physiological system and learns recursively rollable treatment-conditioned dynamics with data-driven cross-system interactions, without imposing a fixed interaction graph.
- **Task-specific context adaptation.** We freeze pretrained task-agnostic dynamics and specialize transition contexts using task-relevant physiological relations and action-conditioned residuals.
- **Evaluation from trajectory forecasting to downstream decisions.** Across MIMIC-III and MIMIC-IV, PatientWorld outperforms the compared methods in nearly all multivariate trajectory forecasting settings and achieves the highest mean primary decision metric among the compared world models on both downstream tasks.

2 Related Work

Clinical trajectory forecasting. Models increasingly target multivariate and probabilistic trajectory prediction from longitudinal EHRs (He & Chiang, 2025; Chang et al., 2024), while general time-series models advance cross-domain multivariate forecasting (Nie et al., 2023; Goswami et al., 2024; Woo et al., 2024). PatientWorld extends this setting from direct future-state prediction to learning recursively rollable, treatment-conditioned patient dynamics.

Structured patient modeling. Prior work introduces explicit structure into patient representations via clinical knowledge graphs, relational EHR modeling, and organ-wise organization (Jiang et al., 2024; Kauffman et al., 2025; Feng et al., 2024; Lal et al., 2020). In contrast, PatientWorld structures the *latent transition dynamics* themselves: physiological systems maintain distinct latent states, treatments condition their evolution, and cross-system interactions are learned dynamically.

World models for clinical decision making. Latent world models learn action-conditioned dynamics for sequential decision making (Hafner et al., 2019; Hansen et al., 2024; Zhao et al., 2026), with recent clinical applications exploring learned transitions for specific treatment decisions (Raghu

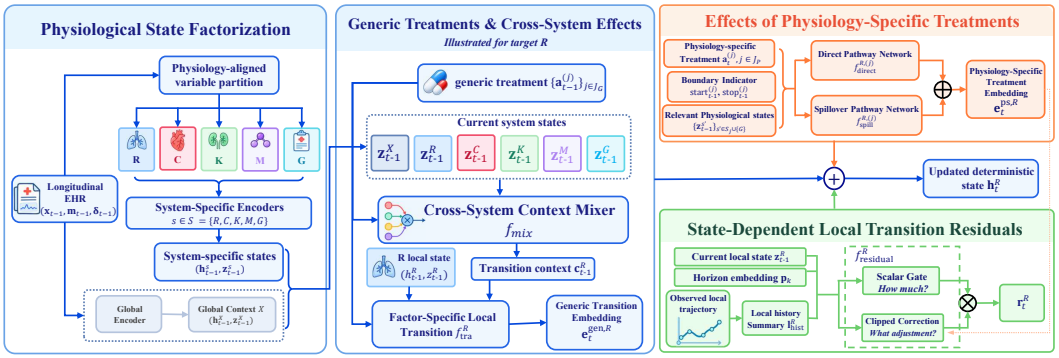


Figure 2: **Physiology-informed latent dynamics in PatientWorld.** PatientWorld factorizes longitudinal EHR observations into physiology-aligned latent states and a non-decoding global context. For target factor R, its updated deterministic state $\mathbf{h}_t^R$ combines three transition components: a generic transition embedding $\mathbf{e}_t^{\mathrm{gen},R}$ driven by cross-system context and generic treatments, a physiology-specific treatment embedding $\mathbf{e}_t^{\mathrm{ps},R}$ aggregating direct and spillover pathways, and a state-dependent local transition residual $\mathbf{r}_t^R$.

et al., 2018; Xu et al., 2026). Rather than coupling dynamics to a specific task, PatientWorld separates broad task-agnostic patient dynamics pretraining from *task-specific context adaptation*. (An extended discussion of related literature is provided in Appendix A).

3 Preliminaries

PatientWorld models each ICU stay as a discrete-time trajectory over fixed-width time bins $t = 1, \ldots, T$. At step t, the observed clinical state is $\mathbf{o}_t = (\mathbf{x}_t, \mathbf{m}_t, \boldsymbol{\delta}_t)$, where $\mathbf{x}_t \in \mathbb{R}^{D_x}$ is the measurement vector (e.g., heart rate of 90 bpm), $\mathbf{m}_t \in \{0,1\}^{D_x}$ is the observation mask, and $\boldsymbol{\delta}_t \in \mathbb{R}_{\geq 0}^{D_x}$ records elapsed time since the latest valid measurement (e.g., hours since the last blood lactate lab).

Alongside the clinical state, the administered treatments are categorized into N_a distinct types (e.g., intravenous fluids, vasopressors). For each treatment type $j \in J = \{1, \ldots, N_a\}$, PatientWorld models three time-t features: the normalized intensity $v_t^{(j)} \in \mathbb{R}$, the availability mask $u_t^{(j)} \in \{0,1\}$ (indicating whether the treatment channel is available in the data), and the activity indicator $e_t^{(j)} \in \{0,1\}$ (indicating whether the treatment is actively occurring). The type-specific treatment vector and full treatment vector are given by $\mathbf{a}_t^{(j)} = [v_t^{(j)}, u_t^{(j)}, e_t^{(j)}] \in \mathbb{R}^3$ and $\mathbf{a}_t = \mathrm{concat}_{j \in J}\, \mathbf{a}_t^{(j)} \in \mathbb{R}^{3N_a}$. We represent treatment boundaries with two binary indicators from adjacent activity events: $\mathrm{start}_t^{(j)} = \mathbf{1}(e_{t-1}^{(j)} = 0 \wedge e_t^{(j)} = 1 \wedge u_t^{(j)} = 1)$ and $\mathrm{stop}_t^{(j)} = \mathbf{1}(e_{t-1}^{(j)} = 1 \wedge e_t^{(j)} = 0 \wedge u_t^{(j)} = 1)$. Here, $\mathbf{1}(\cdot)$ denotes the indicator function, and we set $e_0^{(j)} = 0$ for all $j \in J$.

Building upon the recurrent state-space model (RSSM) formulation (Hafner et al., 2019) , we transition from step $t-1$ to t under the previous treatment vector $\mathbf{a}_{t-1}$. The recurrent transition network f_{transit} updates the deterministic state $\mathbf{h}_t$ from the previous deterministic state $\mathbf{h}_{t-1}$, latent state $\mathbf{z}_{t-1}$, and treatment input: $\mathbf{h}_t = f_{\mathrm{transit}}(\mathbf{h}_{t-1}, \mathbf{z}_{t-1}, \mathbf{a}_{t-1})$. The prior network f_{prior} maps $\mathbf{h}_t$ to a diagonal-Gaussian transition prior $q_{\mathrm{prior},t}(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_t^{\mathrm{prior}}, \mathrm{diag}((\boldsymbol{\sigma}_t^{\mathrm{prior}})^2))$, where $(\boldsymbol{\mu}_t^{\mathrm{prior}}, \boldsymbol{\sigma}_t^{\mathrm{prior}}) = f_{\mathrm{prior}}(\mathbf{h}_t)$. When the current observation tuple $\mathbf{o}_t$ is available, the observation encoder f_{obs} maps it to an observation representation, and the posterior network f_{post} combines this representation with $\mathbf{h}_t$ to define a diagonal-Gaussian posterior $q_{\mathrm{post},t}(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_t^{\mathrm{post}}, \mathrm{diag}((\boldsymbol{\sigma}_t^{\mathrm{post}})^2))$, where $(\boldsymbol{\mu}_t^{\mathrm{post}}, \boldsymbol{\sigma}_t^{\mathrm{post}}) = f_{\mathrm{post}}(\mathbf{h}_t, f_{\mathrm{obs}}(\mathbf{o}_t))$. Without sampling, the forward dynamics propagate distribution means, while the scale parameters $\boldsymbol{\sigma}_t^{\mathrm{prior}}$ and $\boldsymbol{\sigma}_t^{\mathrm{post}}$ are used exclusively to enforce prior–posterior consistency during training. We set $\mathbf{z}_t := \boldsymbol{\mu}_t^{\mathrm{post}}$ during training and filtering, and $\mathbf{z}_t := \boldsymbol{\mu}_t^{\mathrm{prior}}$ during open-loop rollout.

4 METHODOLOGY

PatientWorld learns treatment-conditioned patient dynamics from longitudinal EHR trajectories through three core components. First, we introduce a physiology-informed dynamics backbone that explicitly separates the internal evolution of physiological systems from cross-system interactions and clinical treatments (Sec. 4.1). Second, we pretrain this task-agnostic backbone on broad clinical trajectories to learn general patient dynamics (Sec. 4.2). Finally, for downstream clinical tasks, we keep the pretrained dynamics frozen and adapt only the task-specific transition contexts (Sec. 4.3).

4.1 PHYSIOLOGY-INFORMED LATENT DYNAMICS

Physiological state factorization. Traditional trajectory models typically compress all diverse clinical measurements into a single shared latent representation. However, mixing heterogeneous variables into one monolithic state obscures the distinct internal dynamics of individual physiological systems. To explicitly respect the natural organization of human physiology, PatientWorld factorizes the latent patient state into physiology-aligned, system-specific states. Extending the temporal mechanics introduced in Sec. 3, we partition the observed clinical variables into distinct physiological systems $S = \{R, C, K, M, G\}$ (detailed in Appendix C.1), alongside a non-decoding global context X, defining $S' = S \cup \{X\}$.

Consequently, instead of maintaining a single monolithic state, we decompose the memory and belief into system-specific deterministic recurrent states $\mathbf{h}_t^s \in \mathbb{R}^{d_h}$ and mean latent states $\mathbf{z}_t^s \in \mathbb{R}^{d_z}$ for each $s \in S'$, where d_h and d_z denote their scalar dimensionalities. Correspondingly, each physiological system $s \in S$ receives its own specific observations $\mathbf{o}_t^s$. To advance the trajectory from $t-1$ to t, we replace the monolithic transition f_{transit} with system-specific transition networks f_{transit}^s. As elaborated in the subsequent paragraphs, we compute the updated deterministic state $\mathbf{h}_t^s$ by synthesizing three distinct components: the generic treatments and cross-system effects, the effects of physiology-specific treatments, and state-dependent local transition residuals. Once aggregated, this updated $\mathbf{h}_t^s$ parameterizes the system-specific prior and posterior distributions described in Sec. 3.

Modeling generic treatments and cross-system effects. Some treatments (e.g., intravenous fluids) exert broad, systemic effects that influence the entire patient state, which we define as generic treatments (J_G). In contrast, other treatments (e.g., diuretics) possess highly targeted physiological effects on specific systems, which we define as physiology-specific treatments (J_P). We explicitly partition the full treatment type set J into these two subsets, such that $J_G \cup J_P = J$ and $J_G \cap J_P = \emptyset$.

A target system $s \in S'$ does not evolve in isolation. Its trajectory is jointly shaped by its own internal dynamics, continuous interactions with other systems, and the broad impacts of generic treatments. We need to feed these external influences into the target system, but doing so directly would entangle external conditioning with the system's local transition rules. To keep them neatly separated, we bundle all external information—the previous mean latent states of all systems $\{\mathbf{z}_{t-1}^s\}_{s \in S'}$ and the generic treatments $\{\mathbf{a}_{t-1}^{(j)}\}_{j \in J_G}$—into a single cross-system transition context embedding $\mathbf{c}_{t-1}^s$. We compute this context embedding as:

$$(\mathbf{c}_{t-1}^s)_{s \in S'} = f_{\text{mix}}(\{\mathbf{z}_{t-1}^s\}_{s \in S'}, \{\mathbf{a}_{t-1}^{(j)}\}_{j \in J_G})$$

where f_{mix} is a shared context embedding mixer network. The target system s then applies its own system-specific recurrent network f_{transit}^s, using its previous deterministic state $\mathbf{h}_{t-1}^s$, its own previous mean latent state $\mathbf{z}_{t-1}^s$, the context embedding $\mathbf{c}_{t-1}^s$, and the generic treatments $\{\mathbf{a}_{t-1}^{(j)}\}_{j \in J_G}$:

$$\mathbf{e}_t^{\text{gen},s} = f_{\text{transit}}^s(\mathbf{h}_{t-1}^s, \mathbf{z}_{t-1}^s, \mathbf{c}_{t-1}^s, \{\mathbf{a}_{t-1}^{(j)}\}_{j \in J_G})$$

Thus, this step outputs the generic transition embedding $\mathbf{e}_t^{\text{gen},s}$, which serves as a base transition component encapsulating the system's intrinsic evolution, implicitly learned cross-system interactions, and the broad impacts of generic treatments. Details of f_{mix} and f_{transit}^s are in Appendix D.1

Modeling the effects of physiology-specific treatments. Unlike generic treatments that broadly influence the entire patient state, physiology-specific treatments are primarily administered to focus on specific physiological systems. However, because human physiology is highly interconnected, a treatment aimed at a primary target inevitably triggers ripple effects across other systems. To accurately capture both the strong primary clinical intent and these inevitable physiological spills,

PatientWorld models physiology-specific treatments through two complementary routes: a direct pathway for the predefined target systems, and a spillover pathway for weaker secondary effects on other physiological systems. The goal is to compute a physiology-specific treatment embedding $\mathbf{e}_t^{\mathrm{ps},s}$ for each system $s \in S'$ by combining these two pathways.

For each specific treatment type $j \in J_P$, let $S_j \subset S$ denote the predefined target systems for the direct pathway. As detailed in Appendix C.2, S_j is defined based on established clinical literature (e.g., assigning diuretics primarily to the renal and cardiovascular systems). Unlike existing methods that either omit explicit physiological targeting or rely on shared representations without structurally isolating these pathways (He & Chiang, 2025; Chang et al., 2024), this direct pathway introduces a novel physiology-aligned inductive bias. To condition the effect of a specific treatment j on the current patient state, we collect the previous mean latent states $\{\mathbf{z}_{t-1}^{s'}\}_{s' \in S_j \cup \{G\}}$, where G provides general clinical-context information rather than acting as an additional effect target. Together with the treatment vector $\mathbf{a}_{t-1}^{(j)}$ and scalar boundary indicators $\mathrm{start}_{t-1}^{(j)}, \mathrm{stop}_{t-1}^{(j)}$, these inputs define the treatment-specific direct and spillover embeddings:

$$\mathbf{e}_t^{\mathrm{direct},s,(j)} = f_{\mathrm{direct}}^{s,(j)}\left(\{\mathbf{z}_{t-1}^{s'}\}_{s' \in S_j \cup \{G\}}, \mathbf{a}_{t-1}^{(j)}, \mathrm{start}_{t-1}^{(j)}, \mathrm{stop}_{t-1}^{(j)}\right), \quad s \in S_j,$$

$$\mathbf{e}_t^{\mathrm{spill},s,(j)} = f_{\mathrm{spill}}^{s,(j)}\left(\{\mathbf{z}_{t-1}^{s'}\}_{s' \in S_j \cup \{G\}}, \mathbf{a}_{t-1}^{(j)}, \mathrm{start}_{t-1}^{(j)}, \mathrm{stop}_{t-1}^{(j)}\right), \quad s \in S' \setminus (S_j \cup \{G\}).$$

Here, $f_{\mathrm{direct}}^{s,(j)}$ is a dedicated, treatment-specific direct pathway network that applies the targeted effect to the predefined target set S_j, while $f_{\mathrm{spill}}^{s,(j)}$ is a treatment-specific spillover pathway network that captures secondary effects on physiological systems outside S_j and G. Details are in Appendix C.2.

The final physiology-specific treatment embedding for a target system $s \in S'$ aggregates the direct effects from treatments where s is a primary target and the scaled spillover effects from treatments where s is a secondary target, formulated as $\mathbf{e}_t^{\mathrm{ps},s} = \sum_{j \in J_P:\, s \in S_j} \mathbf{e}_t^{\mathrm{direct},s,(j)} + \alpha \sum_{j \in J_P:\, s \in S' \setminus (S_j \cup \{G\})} \mathbf{e}_t^{\mathrm{spill},s,(j)}$. Here, α is a small scalar spillover scale (detailed in Appendix D.1) that keeps secondary spillover effects weaker than the direct pathway.

State-dependent local transition residuals. Recursive open-loop rollout accumulates prediction error over increasing rollout depth. To mitigate this cumulative drift and stabilize multi-step rollouts, we introduce a bounded correction residual $\mathbf{r}_t^s$ for the observed physiological systems $s \in S$.

Because this drift worsens the further we predict into the unobserved future, the correction must be aware of its relative position. Suppose an observed history window of length W ends at time t_0, containing time steps $[t_0 - W + 1, t_0]$. When rolling out future states at times $t > t_0$, we define the relative rollout depth as $k = t - t_0$, up to a maximum prediction horizon H (i.e., $k \in \{1, 2, \ldots, H\}$). To compute the correction, we rely on a fixed local history summary $\mathbf{l}_{\mathrm{hist}}^s$ extracted from the observation window, the previous mean latent state $\mathbf{z}_{t-1}^s$, the physiology-specific treatment effect $\mathbf{e}_t^{\mathrm{ps},s}$, and a dedicated temporal embedding $\mathbf{p}_k$. Unlike absolute clinical clock time, $\mathbf{p}_k$ functions as a relative prediction-horizon embedding (analogous to a positional embedding). It is retrieved from a learnable lookup table parameterized by the relative step k, informing the network whether it is predicting the immediate next step or further into the future. This formulation allows the model to learn distinct, horizon-specific correction patterns. The residual vector is computed as:

$$\mathbf{r}_t^s = f_{\mathrm{residual}}^s(\mathbf{z}_{t-1}^s, \mathbf{l}_{\mathrm{hist}}^s, \mathbf{e}_t^{\mathrm{ps},s}, \mathbf{p}_k), \quad s \in S$$

where f_{residual}^s is a dedicated correction network (see Appendix D.1 for its details). The residual vector $\mathbf{r}_t^s$ provides a local-history-dependent correction during open-loop rollout.

Final updated system state. The final updated deterministic state $\mathbf{h}_t^s$ combines the three transition components: $\mathbf{e}_t^{\mathrm{gen},s}$ provides the base transition under cross-system context and generic treatments, $\mathbf{e}_t^{\mathrm{ps},s}$ injects physiology-specific treatment effects, and $\mathbf{r}_t^s$ supplies the bounded local-history correction for observed physiological systems. Thus, it is computed as:

$$\mathbf{h}_t^s = \begin{cases} \mathbf{e}_t^{\mathrm{gen},s} + \mathbf{e}_t^{\mathrm{ps},s} + \mathbf{r}_t^s, & s \in S, \\ \mathbf{e}_t^{\mathrm{gen},X} + \mathbf{e}_t^{\mathrm{ps},X}, & s = X. \end{cases}$$

The resulting $\mathbf{h}_t^s$ parameterizes the system-specific latent prior $q_{\mathrm{prior},t}^s(\mathbf{z})$ described in Sec. 3. Let $\boldsymbol{\mu}_{\mathrm{prior},t}^s$ denote its mean; for each system $s \in S$, PatientWorld predicts future physiological values and observation-mask logits at time t via a system-specific decoder network f_{dec}^s, such that $(\hat{\mathbf{x}}_t^s, \hat{\mathbf{m}}_t^s) = f_{\mathrm{dec}}^s(\mathbf{h}_t^s, \boldsymbol{\mu}_{\mathrm{prior},t}^s)$.

4.2 TASK-AGNOSTIC PATIENT WORLD-MODEL PRETRAINING

PatientWorld is pretrained on ICU observation–treatment trajectories with two complementary prediction modes. The posterior-conditioned branch uses the next observed physiology during training to update latent states and supervise one-step filtering, whereas the open-loop branch starts from the same history and recursively propagates transition priors under the future treatment sequence without future physiological observations. Both branches predict physiological values and observation masks, and pretraining uses no reward, policy, or task-specific decision objective.

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{posterior}} + \lambda_{\text{rollout}}\mathcal{L}_{\text{rollout}} + \lambda_{\text{overshoot}}\mathcal{L}_{\text{overshoot}} + \mathcal{L}_{\text{aux}}. \tag{1}$$

Here, the first three terms follow standard latent world-model objectives for posterior-conditioned prediction, open-loop rollout prediction, and multi-step prior–posterior consistency (Hafner et al., 2019), while the auxiliary term provides subsystem-level and local-residual regularization; details are in Appendix D.2.

4.3 TASK-SPECIFIC CONTEXT ADAPTATION

When applying a world model to different downstream tasks, the critical treatments and cross-system dependencies often change, but the patient's underlying physiological dynamics remain the same. Existing methods typically handle new tasks by either fully fine-tuning the entire model—which overwrites the reusable dynamics—or by relying on a single, fixed interaction graph that cannot adapt to different clinical settings (Xu et al., 2026; Jiang et al., 2024). To address this issue, we design a task-specific context adaptation method, which modifies the pretrained context $\mathbf{c}^s_{t-1}$ to compute a task-adapted context $\tilde{\mathbf{c}}^{\tau,s}_{t-1}$ for each physiological system, while keeping the pretrained local transition network f^s_{transit} strictly frozen.

To explicitly capture these task-specific requirements, we must identify the clinically relevant inter-system influences and critical treatments for the given task. For a downstream task τ, we define a sparse directed relation graph $\mathcal{G}_\tau = (S, E_\tau)$, where the directed edges E_τ indicate task-specific inter-system influences. Specifically, these graphs are defined based on established clinical literature, where a directed edge $(s', s) \in E_\tau$ exists if the source system s' exerts a clinically relevant influence on the target system s under the specific context of task τ (Appendix E.1). Based on this graph, we collect the previous mean latent states of the incoming source systems for a target system s, denoted as $\{\mathbf{z}^{s'}_{t-1}\}_{s':(s',s)\in E_\tau}$. Alongside this graph, we identify a task-specific treatment subset $J_\tau \subset J$ and strictly collect the treatment vectors for these task-relevant treatments as $\{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau}$ (e.g., J_{sepsis} contains fluid and vasopressor treatments). To integrate the graph-based relations from the incoming source systems, the current system state, and the selected task treatments, we define a task- and system-specific adapter network $f^{\tau,s}_{\text{adapt}}$(see Appendix E.2 for the implementation details). The adapted context is then computed by adding this adapter's residual output to the pretrained context vector:

$$\tilde{\mathbf{c}}^{\tau,s}_{t-1} = \mathbf{c}^s_{t-1} + f^{\tau,s}_{\text{adapt}}\left(\{\mathbf{z}^{s'}_{t-1}\}_{s':(s',s)\in E_\tau}, \mathbf{z}^s_{t-1}, \{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau}\right), \qquad s \in S.$$

Finally, $\tilde{\mathbf{c}}^{\tau,s}_{t-1}$ replaces $\mathbf{c}^s_{t-1}$ as the input context to the frozen f^s_{transit}.

During training, we optimize only the task-specific adapter networks, keeping all the pretrained networks described in 4.1 frozen. The adaptation process remains a predictive world-model objective over the downstream trajectories. The overall loss $\mathcal{L}_{\text{adapt}}$ combines normalized one-step physiological prediction error $\mathcal{L}_{\text{1step}}$, multi-step open-loop prediction error $\mathcal{L}_{\text{multi}}$, and future observation-mask prediction loss $\mathcal{L}_{\text{mask}}$:

$$\mathcal{L}_{\text{adapt}} = \mathcal{L}_{\text{1step}} + \lambda_{\text{multi}}\mathcal{L}_{\text{multi}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}}$$

Further loss details, optimization configurations and hyperparameters are detailed in Appendix E.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Datasets and tasks. We evaluate PatientWorld on two EHR datasets, MIMIC-III (Johnson et al., 2016b) and MIMIC-IV (Johnson et al., 2023b). Following Sec. 3, each ICU stay is discretized into

Table 1: **Multivariate patient trajectory forecasting.** Each model occupies two rows, with pooled normalized RMSE on top and MAE below. Global H3/H6 pool observed targets over rollout steps 1:3/1:6, respectively. Values are mean ± standard deviation over three seeds; lower is better. Light rose and blue denote the best and second-best means.

	MIMIC-III				MIMIC-IV			
	Global H3		Global H6		Global H3		Global H6	
Model	**Cond.**	**Blind**	**Cond.**	**Blind**	**Cond.**	**Blind**	**Cond.**	**Blind**
GRU	0.6310±.0018	0.6955±.0191	0.6659±.0018	0.7140±.0152	0.5801±.0003	0.6470±.0049	0.6188±.0009	0.6928±.0080
	0.4314±.0022	0.4831±.0148	0.4654±.0008	0.5034±.0124	0.3817±.0006	0.4536±.0050	0.4159±.0007	0.5022±.0109
Transformer	0.6241±.0026	0.6356±.0041	0.6570±.0027	0.6859±.0086	0.5691±.0006	0.6219±.0077	**0.6006**±.0006	0.6765±.0100
	0.4263±.0035	0.4376±.0042	0.4572±.0022	0.4817±.0091	0.3704±.0002	0.4388±.0082	0.3990±.0004	0.4826±.0089
DLinear	0.6462±.0035	0.6810±.0305	0.6740±.0017	0.6956±.0172	0.6066±.0001	0.7162±.0187	0.6416±.0001	0.7526±.0343
	0.4420±.0024	0.4733±.0260	0.4736±.0003	0.4936±.0144	0.4029±.0008	0.5272±.0143	0.4352±.0006	0.5583±.0315
PatchTST	0.6340±.0055	0.6989±.0067	0.6656±.0039	0.7159±.0211	0.5810±.0006	0.6389±.0140	0.6203±.0006	0.6846±.0126
	0.4336±.0059	0.4891±.0097	0.4667±.0038	0.5072±.0194	0.3820±.0011	0.4520±.0109	0.4155±.0006	0.4955±.0141
iTransformer	0.6298±.0012	0.6824±.0403	0.6673±.0013	0.7330±.0278	0.5798±.0001	0.5989±.0074	0.6247±.0006	0.6899±.0263
	0.4260±.0018	0.4923±.0483	0.4594±.0004	0.5277±.0321	0.3724±.0012	0.4052±.0120	0.4117±.0007	0.4947±.0313
MOMENT	0.6958±.0011	0.7113±.0033	0.7196±.0009	0.7233±.0032	0.6435±.0013	0.6483±.0035	0.6789±.0014	0.6875±.0044
	0.4764±.0014	0.4970±.0053	0.5042±.0002	0.5061±.0028	0.4306±.0025	0.4366±.0048	0.4613±.0005	0.4758±.0041
Moirai 2.0	0.6322±.0004	0.6419±.0046	0.6624±.0004	0.6773±.0039	0.5864±.0068	0.6094±.0042	0.6190±.0003	0.6679±.0104
	0.4263±.0010	0.4360±.0058	0.4588±.0004	0.4730±.0027	0.3792±.0042	0.4165±.0063	0.4097±.0004	0.4746±.0115
TFT-multi	0.6287±.0001	0.6428±.0063	0.6646±.0017	0.6864±.0050	0.5822±.0017	0.6227±.0067	0.6166±.0001	0.6707±.0026
	0.4300±.0014	0.4439±.0068	0.4639±.0010	0.4852±.0015	0.3815±.0031	0.4399±.0139	0.4120±.0003	0.4831±.0123
TDSTF	0.9021±.0089	0.9021±.0089	0.9026±.0022	0.9026±.0022	0.8374±.0071	0.8374±.0071	0.8463±.0090	0.8463±.0090
	0.6471±.0078	0.6471±.0078	0.6536±.0005	0.6536±.0005	0.5592±.0050	0.5592±.0050	0.5710±.0049	0.5710±.0049
Phys-JEPA	0.6307±.0028	1.1106±.0573	0.6622±.0022	0.9597±.0653	0.5800±.0018	1.1875±.1459	0.6160±.0021	1.1023±.1683
	0.4342±.0019	0.8608±.0472	0.4627±.0022	0.7452±.0804	0.3836±.0020	0.9607±.1161	0.4154±.0021	0.8657±.1696
MedDreamer	0.6439±.0021	0.6764±.0072	0.6696±.0014	0.7190±.0086	0.5940±.0005	0.6463±.0119	0.6279±.0008	0.7032±.0036
	0.4431±.0014	0.4704±.0035	0.4706±.0013	0.5129±.0031	0.3951±.0014	0.4580±.0132	0.4233±.0017	0.5088±.0047
PatientWorld	**0.6234**±.0009	**0.6251**±.0005	**0.6546**±.0009	**0.6591**±.0020	**0.5687**±.0010	**0.5780**±.0021	0.6052±.0005	**0.6233**±.0012
	0.4222±.0016	**0.4240**±.0013	**0.4524**±.0019	**0.4576**±.0035	**0.3654**±.0014	**0.3734**±.0012	**0.3976**±.0009	**0.4130**±.0007

4-hour bins and split at the patient level into training, validation, and test sets. During pretraining, we train separate models on MIMIC-III and MIMIC-IV. Given a 12-step (48 h) history, we train separate H3 and H6 configurations to predict the next 3 and 6 steps (12 and 24 h) via recursive open-loop latent rollout. In adaptation, both downstream tasks are constructed from MIMIC-IV, with task construction adapted from prior work on sepsis treatment (Komorowski et al., 2018) and RRT initiation timing for acute kidney injury (Grolleau et al., 2024). For downstream experiments, all variants that reuse pretrained weights initialize from the same MIMIC-IV H6-pretrained PatientWorld checkpoint; PatientWorld Adapt keeps the pretrained backbone fixed. Cohort construction and preprocessing are in Appendix B.

Baselines & Evaluation Protocol. For trajectory forecasting, we compare against diverse task-trained forecasters, time-series foundation models, clinical trajectory models, and world-model baselines under a shared binned trajectory representation, data split, and normalized RMSE/MAE evaluation protocol (conditional and blind). For downstream adaptation, PatientWorld Adapt is compared against alternative downstream training and reuse strategies (e.g., Full Fine-tuning). For downstream policy evaluation, we include CQL (Kumar et al., 2020) and representative forecasting baselines. All dynamics-based methods share the same task-specific action space, planner, and strict offline policy evaluation (OPE) protocol with Reward-FQE (sepsis) and HFD60 FQE (RRT) as primary metrics. Trainable models are reported over three seeds as mean ± standard deviation. Implementation, baselines, and evaluation metrics are detailed in Appendices B, F, and G.

5.2 Performance on Multivariate Patient Trajectory Forecasting

PatientWorld achieves the strongest overall forecasting performance, with larger gains under blind forecasting and at longer horizons. As shown in Table 1, PatientWorld obtains the lowest mean forecasting error in 15 of the 16 combinations of dataset, horizon, conditioning setting, and metric across MIMIC-III and MIMIC-IV. Under the common forecasting protocol, it outperforms a broad range of task-specific forecasters, time-series foundation models, clinical trajectory models, and world-model baselines. Under conditional forecasting, the strongest meth-

Table 2: **Task-specific adaptation of pretrained patient dynamics.** Global RMSE@H pools masked prediction errors over future steps $1{:}H$, while Macro RMSE@H6 reports the task-variable macro average. Results are mean ± standard deviation over three seeds. Lower is better. Best results are in **bold** and second-best results are underlined.

	Sepsis Treatment				RRT Initiation Timing			
	Global RMSE ↓			Macro RMSE ↓	Global RMSE ↓			Macro RMSE ↓
Method	H1	H3	H6	H6	H1	H3	H6	H6
From Scratch	0.5380±0.0008	0.5921±0.0010	0.6317±0.0010	0.5897±0.0006	0.5881±0.0022	0.6375±0.0018	0.6829±0.0020	0.6440±0.0024
Zero-shot	0.5469±0.0000	0.6060±0.0000	0.6524±0.0000	0.6219±0.0000	0.5749±0.0000	0.6351±0.0000	0.6921±0.0000	0.6507±0.0000
Full Fine-tuning	0.5394±0.0009	0.5927±0.0006	0.6320±0.0003	0.5888±0.0013	0.5674±0.0010	0.6228±0.0008	**0.6684±0.0010**	0.6352±0.0026
Random Graph Adapt	0.5318±0.0002	0.5895±0.0001	0.6305±0.0001	0.5881±0.0004	0.5692±0.0005	0.6253±0.0005	0.6743±0.0003	0.6377±0.0006
PatientWorld Adapt	**0.5310±0.0003**	**0.5882±0.0001**	**0.6284±0.0002**	**0.5871±0.0003**	**0.5665±0.0001**	**0.6218±0.0002**	0.6698±0.0001	**0.6338±0.0002**

ods perform relatively similarly, with Transformer slightly outperforming PatientWorld only on conditional RMSE at H6 on MIMIC-IV. The differences become more pronounced under blind forecasting and at longer horizons. At H6 on MIMIC-IV, PatientWorld reduces blind RMSE and MAE by 6.7% and 13.0%, respectively, relative to the strongest competing baseline. A similar pattern is observed on MIMIC-III, where the advantage also increases under blind forecasting.

The pretraining ablation in Fig. 3 examines physiological factorization, physiology-specific treatment pathways, physiological alignment, and state-dependent local transitions. Replacing the dense RSSM with physiology-factorized dynamics reduces conditional and blind forecasting errors on both datasets. Adding the physiology-aligned pathway for physiology-specific treatments to the factorized model further improves blind forecasting on both datasets. Under the complete architecture, shuffling the primary treatment-to-system assignments consistently degrades performance relative to PatientWorld, supporting the value of physiological alignment under the same architecture and parameterization. Removing the state-dependent local residual increases forecasting error, with the complete PatientWorld model achieving the lowest Global RMSE@H3 across all four settings.

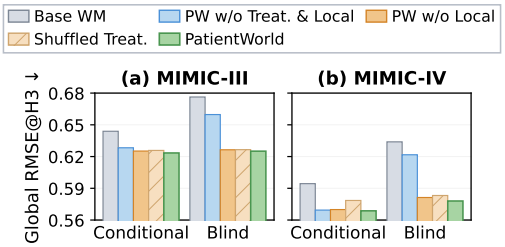


Figure 3: **Pretraining ablation study.** Global RMSE@H3 pools observed targets over rollout steps 1:3 on MIMIC-III and MIMIC-IV.

5.3 Comparison of Downstream Adaptation Methods

PatientWorld Adapt shows that pretrained patient dynamics can be reused across distinct downstream tasks through task-specific transition-context adaptation. Table 2 separates the roles of pretraining, task specialization, and physiological structure in downstream reuse. PatientWorld Adapt outperforms From Scratch and Zero-shot across both tasks: the former supports the value of pretraining, while the latter shows that pretrained dynamics still benefit from task-specific specialization. With the pretrained backbone frozen, PatientWorld Adapt outperforms Full Fine-tuning in seven of eight comparisons, except RRT Global RMSE@H6, indicating that downstream specialization can be achieved while retaining the pretrained dynamics. It also outperforms Random Graph Adapt on all eight metrics under the same adapter and training protocol, supporting the value of task-relevant physiological relations. Together with pretraining shuffled physiology-specific treatment-routing control, these results support physiology-aligned priors across both stages.

5.4 Analysis of Treatment Conditioning and Action Responsiveness

PatientWorld's forecasting advantage increases under blind forecasting for both Action and No-action segments. Across both datasets, differences from Transformer are generally small under conditional forecasting, whereas PatientWorld's advantage increases under blind forecasting on both subsets and generally grows at longer horizons (Fig. 4(a)). Full Action/No-action results for trajectory forecasting methods are in Appendix G.2.

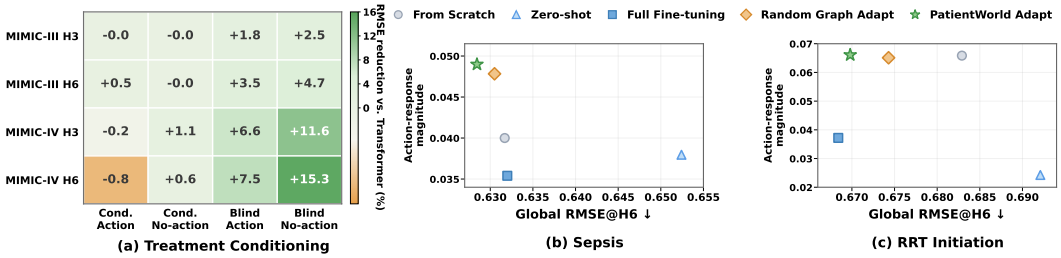


Figure 4: **Treatment conditioning and action responsiveness.** (a) Relative RMSE reduction of PatientWorld over Transformer across Action/No-action subsets under conditional and blind forecasting. (b–c) Global RMSE@H6 versus action-response magnitude after task-specific adaptation for sepsis and RRT initiation timing. Action-response magnitude measures variation in predicted trajectories under alternative admissible actions and is not a causal treatment-effect estimate.

Table 3: **Downstream decision utility under offline evaluation.** Results are mean ± standard deviation over three seeds. For sepsis, Reward-FQE is the primary utility metric, while Mean Supp. and Low Supp. summarize behavior-policy support; Low Supp. denotes the fraction with support below 0.01. For RRT initiation timing, HFD60 is evaluated using both FQE and WIS, and WIS ESS measures the effective support of the importance-sampling estimate. Best results are in **bold** and second-best results are underlined.

	Sepsis Treatment				RRT Initiation Timing			
Method	Reward-FQE ↑	Est. Mort. ↓	Mean Supp. ↑	Low Supp. ↓	HFD60 FQE ↑	HFD60 WIS ↑	Hosp. Mort. FQE ↓	WIS ESS ↑
CQL	6.382 ± 0.084	19.13 ± 0.77%	0.099 ± 0.011	26.07 ± 1.04%	36.640 ± 0.110	40.673 ± 0.080	19.14 ± 0.10%	57.59 ± 4.75%
MedDreamer	5.831 ± 0.344	20.11 ± 1.18%	0.247 ± 0.046	7.76 ± 3.53%	38.657 ± 0.064	**42.050 ± 0.695**	15.98 ± 0.37%	61.06 ± 8.69%
Transformer	6.133 ± 0.217	20.39 ± 0.62%	0.246 ± 0.063	10.49 ± 7.48%	38.420 ± 0.219	35.940 ± 1.276	**14.76 ± 0.51%**	20.64 ± 1.67%
Moirai 2.0	5.889 ± 0.472	24.75 ± 1.19%	0.207 ± 0.138	22.45 ± 26.63%	38.150 ± 0.556	37.200 ± 4.259	16.29 ± 0.31%	30.64 ± 20.39%
TFT-multi	6.072 ± 0.141	21.25 ± 1.47%	**0.283 ± 0.007**	4.77 ± 1.18%	37.683 ± 0.058	31.970 ± 1.388	16.08 ± 0.38%	15.64 ± 2.50%
Phys-JEPA	5.810 ± 0.772	21.82 ± 1.27%	0.212 ± 0.079	11.24 ± 4.66%	37.447 ± 0.029	25.123 ± 1.617	16.27 ± 0.29%	5.06 ± 1.01%
PatientWorld	**6.613 ± 0.129**	**17.28 ± 1.07%**	0.268 ± 0.002	**4.23 ± 2.05%**	**39.317 ± 0.025**	40.700 ± 0.070	15.69 ± 0.12%	**75.42 ± 3.21%**

PatientWorld Adapt retains predictive accuracy while producing action-dependent trajectories. Holding patient history fixed and varying the admissible action, PatientWorld Adapt exhibits the largest action-response magnitude on both tasks while achieving the best Global RMSE@H6 on sepsis and near-best accuracy on RRT (Fig. 4(b–c)). Action-response magnitude is a sensitivity diagnostic, not a performance metric or causal treatment-effect estimate.

5.5 Offline Evaluation of Downstream Decision Policies

As shown in Table 3, PatientWorld achieves the highest mean estimate on the primary FQE metric for both downstream tasks (cumulative reward for sepsis, HFD60 for RRT) and the lowest estimated mortality on sepsis. Complementary estimators provide additional context: MedDreamer attains the highest WIS-estimated HFD60, while Transformer has the lowest FQE-estimated hospital mortality. PatientWorld consistently demonstrates strong support diagnostics across both tasks (e.g., highest WIS effective sample size on RRT). Complementary WDR results are reported in Appendix G.6.

6 Conclusion & Future Work

We introduced PatientWorld, a physiology-informed world model that combines physiology-organized latent states, task-agnostic pretraining, and task-specific context adaptation. Across MIMIC-III and MIMIC-IV, it achieves the strongest overall multivariate forecasting performance and the highest mean primary decision metric among the compared world models on both downstream tasks. These results support reusing a common pretrained dynamics backbone while specializing transition contexts across the evaluated clinical decision tasks. Downstream evaluation remains retrospective and offline; future work will examine additional clinical settings.

AI USE STATEMENT

Generative AI tools, including ChatGPT and Codex, were used to aid or polish writing and literature retrieval and discovery. All AI-assisted outputs were reviewed by the authors. Reported results were verified against the underlying experimental artifacts, cited sources were checked by the authors, and AI-assisted code was reviewed and tested where applicable. The authors take responsibility for the final methodology, experiments, analyses, claims, text, figures, and code presented in this work.

ETHICS STATEMENT

This study uses de-identified MIMIC-III and MIMIC-IV data accessed under the corresponding PhysioNet credentialing and data-use requirements, and no new patient data were collected. All downstream decision analyses are retrospective and offline. The learned dynamics and evaluated policies are intended for research evaluation rather than direct clinical deployment, and their outputs should not be interpreted as causal treatment-effect estimates.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce the study throughout the main paper and appendices. Appendix A provides an extended discussion of related work; Appendix B describes dataset versions, cohort construction, preprocessing, and data splits; Appendix C specifies the physiological variable partition and treatment definitions; Appendix D details the pretraining architecture, objectives, loss weights, and optimization; Appendix E describes task-specific adaptation; Appendix F documents baseline implementations, ablations, and additional pretraining analyses; and Appendix G defines evaluation metrics, downstream policy construction, and offline policy-evaluation procedures. Trainable results are reported over three seeds unless otherwise specified. The accompanying code and reproducibility materials are available at `https://anonymous.4open.science/r/PatientWorld-0AB4/`.

REFERENCES

André Borges and Luís Bento. Organ crosstalk and dysfunction in sepsis. *Annals of Intensive Care*, 14(1):147, 2024.

Ping Chang, Huayu Li, Stuart F. Quan, Shuyang Lu, Shu-Fen Wung, Janet Roveda, and Ao Li. A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in intensive care unit. *Computer Methods and Programs in Biomedicine*, 246:108060, 2024. ISSN 0169-2607. doi: 10.1016/j.cmpb.2024.108060. URL `https://www.sciencedirect.com/science/article/pii/S0169260724000567`.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL `https://aclanthology.org/D14-1179/`.

Claudia S Cohn and Beth H Shaz. Blood and its components. *JAMA*, 330(19):1903–1904, 2023.

Luca Di Lullo, Patrick Bronson Reeves, Antonio Bellasi, and Claudio Ronco. Cardiorenal syndrome in acute kidney injury. *Seminars in nephrology*, 39:31–40, 2019.

G Michael Felker, David H Ellison, Wilfried Mullens, Zachary L Cox, and Jeffrey M Testani. Diuretic therapy for patients with heart failure: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 75(10):1178–1195, 2020.

Xue Feng, Siyi Zhu, Yanfei Shen, Huaiping Zhu, Molei Yan, Guolong Cai, and Gangmin Ning. Multi-organ spatiotemporal information aware model for sepsis mortality prediction. *Artificial Intelligence in Medicine*, 147:102746, 2024.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 16115–16152. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/goswami24a.html.

François Grolleau, François Petit, Stéphane Gaudry, Élise Diard, Jean-Pierre Quenot, Didier Dreyfuss, Viet-Thi Tran, and Raphaël Porcher. Personalizing renal replacement therapy initiation in the intensive care unit: a reinforcement learning-based strategy with external validation on the AKIKI randomized controlled trials. *Journal of the American Medical Informatics Association*, 31(5):1074–1083, 2024. doi: 10.1093/jamia/ocae004.

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, 2019.

Nick Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 47376–47405, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/cf73d57b6dcda32b293df7c2d5341f49-Paper-Conference.pdf.

Rosemary He and Jeffrey N Chiang. Simultaneous forecasting of vital sign trajectories in the ICU. *Scientific Reports*, 15(1):14996, 2025.

Pengcheng Jiang, Cao (Danica) Xiao, Adam Cross, and Jimeng Sun. GraphCare: Enhancing healthcare predictions with personalized knowledge graphs. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 16839–16866, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/484d254ff80e99d543159440a06db0de-Paper-Conference.pdf.

Michael Joannidis, Lui G Forni, Sebastian J Klein, Patrick M Honore, Kianoush Kashani, Marlies Ostermann, John Prowle, Sean M Bagshaw, Vincenzo Cantaluppi, Michael Darmon, et al. Lung–kidney interactions in critically ill patients: consensus report of the acute disease quality initiative (ADQI) 21 workgroup. *Intensive care medicine*, 46(4):654–672, 2020.

Alistair Johnson, Tom Pollard, and Roger Mark. MIMIC-III Clinical Database. *PhysioNet*, September 2016a. doi: 10.13026/C2XW26. URL https://doi.org/10.13026/C2XW26. Version 1.4.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV. *PhysioNet*, January 2023a. doi: 10.13026/6mm1-ek67. URL https://doi.org/10.13026/6mm1-ek67. Version 2.2.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1):160035, 2016b.

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, 2023b.

Justin Kauffman, Emma Holmes, Akhil Vaid, Alexander W Charney, Patricia Kovatch, Joshua Lampert, Ankit Sakhuja, Marinka Zitnik, Benjamin S Glicksberg, Ira Hofer, et al. InfEHR: Clinical phenotype resolution through deep geometric learning on electronic health records. *Nature Communications*, 16(1):8475, 2025.

Arif Khwaja. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clinical Practice*, 120(4):c179–c184, 2012.

Antoine Kimmoun, Emmanuel Novy, Thomas Auchet, Nicolas Ducrocq, and Bruno Levy. Hemodynamic consequences of severe lactic acidosis in shock states: from bench to bedside. *Critical Care*, 19(1):175, 2015.

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*, 24:1716–1720, 2018. doi: 10.1038/s41591-018-0213-5. URL https://www.nature.com/articles/s41591-018-0213-5.

John Kreit. Respiratory-cardiovascular interactions during mechanical ventilation: Physiology and clinical implications. *Comprehensive Physiology*, 12(3):3425–3448, 2022.

Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. Conservative Q-learning for offline reinforcement learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 1179–1191. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/0d2b2061826a5df3221116a5085a6052-Paper.pdf.

Amos Lal, Guangxi Li, Edin Cubro, Sarah Chalmers, Heyi Li, Vitaly Herasevich, Yue Dong, Brian W Pickering, Oguz Kilickaya, and Ognjen Gajic. Development and verification of a digital twin patient model to predict specific treatment response during the first 24 hours of sepsis. *Critical Care Explorations*, 2(11):e0249, 2020.

Lies Langouche and Greet Van den Berghe. Glucose metabolism and insulin therapy. *Critical Care Clinics*, 22(1):119–129, 2006.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When less is more for time series forecasting. *arXiv preprint arXiv:2511.11698*, 2025.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *International Conference on Learning Representations*, 2024. URL https://mlanthology.org/iclr/2024/liu2024iclr-itransformer/.

Weizhi Nie, Weichao Liu, Honglin Guo, and Yuting Su. Phys-JEPA: Physics-informed latent world models for multivariate time-series forecasting, 2026. URL https://arxiv.org/abs/2606.16076.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. In *International Conference on Learning Representations*, 2023. URL https://mlanthology.org/iclr/2023/nie2023iclr-time/.

Aniruddh Raghu, Matthieu Komorowski, and Sumeetpal Singh. Model-based reinforcement learning for sepsis treatment, 2018. URL https://arxiv.org/abs/1811.09602.

Janani Rangaswami, Vivek Bhalla, John EA Blair, Tara I Chang, Salvatore Costa, Krista L Lentine, Edgar V Lerma, Kenechukwu Mezue, Mark Molitch, Wilfried Mullens, et al. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the american heart association. *Circulation*, 139(16):e840–e878, 2019.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*, 4(1):86, 2021. doi: 10.1038/s41746-021-00455-y. URL https://doi.org/10.1038/s41746-021-00455-y.

Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean-Daniel Chiche, Craig M. Coopersmith, Richard S. Hotchkiss, Mitchell M. Levy, John C. Marshall, Greg S. Martin, Steven M. Opal, Gordon D. Rubenfeld, Tom van der Poll, Jean-Louis Vincent, and Derek C. Angus. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, 315(8):801–810, 2016. doi: 10.1001/jama.2016.0287. URL https://jamanetwork.com/journals/jama/fullarticle/2492881.

Ethan Steinberg, Ken Jung, Jason A Fries, Conor K Corbin, Stephen R Pfohl, and Nigam H Shah. Language models are an effective representation learning technique for electronic health record data. *Journal of Biomedical Informatics*, 113:103637, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure: On behalf of the working group on sepsis-related problems of the European Society of Intensive Care Medicine (see contributors to the project in the appendix). *Intensive Care Medicine*, 22(7):707–710, 1996. doi: 10.1007/BF01709751. URL https://doi.org/10.1007/BF01709751.

Alexander P Vlaar, Simon Oczkowski, Sanne De Bruin, Marije Wijnberge, Massimo Antonelli, Cecile Aubron, Philippe Aries, Jacques Duranteau, Nicole P Juffermans, Jens Meier, et al. Transfusion strategies in non-bleeding critically ill adults: a clinical practice guideline from the european society of intensive care medicine. *Intensive Care Medicine*, 46(4):673–696, 2020.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified Training of Universal Time Series Forecasting Transformers. In *International Conference on Machine Learning*, volume 235, pp. 53140–53164, 2024. URL https://mlanthology.org/icml/2024/woo2024icml-unified/.

Qianyi Xu, Gousia Habib, Feng Wu, Dilruk Perera, and Mengling Feng. medDreamer: Model-based reinforcement learning with latent imagination on complex EHRs for clinical decision support. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '26, pp. 1693–1704, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400722585. doi: 10.1145/3770854.3780292. URL https://doi.org/10.1145/3770854.3780292.

Zhichao Yang, Avijit Mitra, Weisong Liu, Dan Berlowitz, and Hong Yu. TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nature Communications*, 14(1):7857, 2023. doi: 10.1038/s41467-023-43715-z.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128, June 2023. ISSN 2159-5399. doi: 10.1609/aaai.v37i9.26317. URL http://dx.doi.org/10.1609/aaai.v37i9.26317.

Yi Zhao, Aidan Scannell, Wenshuai Zhao, Yuxin Hou, Tianyu Cui, Le Chen, Dieter Büchler, Arno Solin, Juho Kannala, and Joni Pajarinen. Efficient reinforcement learning by guiding world models with non-curated data. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=oBXfPyi47m.

A Extended Related Work

Building upon the core trajectory forecasting and clinical representation literature, recent clinical models increasingly target complex multivariate settings, including joint ICU trajectory forecasting and probabilistic forecasting of sparse physiological time series (He & Chiang, 2025; Chang et al., 2024). Concurrently, general time-series models have advanced multivariate forecasting and cross-domain transfer through specialized architectures and large-scale pretraining (Nie et al., 2023; Liu et al., 2024; Goswami et al., 2024; Woo et al., 2024).

Within structured patient modeling, GraphCare and InfEHR organize patient information as structured graphs for clinical prediction and representation learning (Jiang et al., 2024; Kauffman et al., 2025), while rGAT-mGRU more directly models organ-specific temporal dynamics and multi-organ interactions (Feng et al., 2024). Related digital-twin work further represents organ states, disease progression, and treatment relations for patient-response simulation (Lal et al., 2020).

For sequential decision making, clinical model-based methods have used learned patient transitions or simulators to guide treatment decisions (Raghu et al., 2018). Most closely related to our world-model baseline setting, MedDreamer combines latent world modeling with irregular EHR modeling and latent imagination for sepsis treatment and mechanical ventilation tasks (Xu et al., 2026).

B Data and Cohort Construction

B.1 Pretraining Cohorts

Pretraining is conducted separately on MIMIC-III v1.4.0 (Johnson et al., 2016a) and MIMIC-IV v2.2 (Johnson et al., 2023a), with an independent PatientWorld model trained for each database. Both datasets are de-identified and accessed under the corresponding PhysioNet credentialing and data-use requirements. We use an ICU stay as the trajectory unit and retain adult ICU stays with age ≥ 18 years and ICU length of stay ≥ 4 h. Longitudinal clinical observations and treatments are discretized on a common 4-h grid.

Each database is partitioned at the patient level into training, validation, and test sets using an 80%/10%/10% split (seed 9899), with all ICU stays from the same patient assigned to the same split. Detailed cohort statistics are reported in Table 4.

Table 4: **Pretraining cohort statistics.** MIMIC-III and MIMIC-IV are processed and used for pretraining separately.

Dataset	Split	Patients	ICU stays	Transitions
MIMIC-III	Train	30,718	42,361	1,035,056
	Valid	3,840	5,330	132,044
	Test	3,840	5,353	132,773
	Total	**38,398**	**53,044**	**1,299,873**
MIMIC-IV	Train	40,662	58,103	1,191,799
	Valid	5,083	7,175	140,184
	Test	5,082	7,304	147,146
	Total	**50,827**	**72,582**	**1,479,129**

B.2 Temporal Discretization and Preprocessing

Pretraining inputs comprise 29 longitudinal clinical variables and a canonical treatment-action schema represented on the same 4-h grid. For each clinical variable, we retain its value, an observation mask, and the elapsed time since its most recent observation; elapsed time is capped at 168 h. Within each bin, variables configured for latest-value aggregation retain the most recent available observation, while the remaining observation variables use the preprocessing pipeline's value/count aggregation. Missing observations are represented explicitly by the observation mask. Normalization statistics are estimated exclusively from observed values in the training split and are

then fixed for validation and test data. MIMIC-III and MIMIC-IV use the same pretraining temporal representation and preprocessing specification.

B.3 Downstream Task Cohorts

Both downstream tasks are constructed from MIMIC-IV and inherit the patient-level training, validation, and test partition used for the corresponding pretraining trajectories.

Sepsis treatment. The base sepsis cohort follows an adaptation of Microsoft's publicly available `mimic_sepsis` implementation,[1] which implements Sepsis-3-style cohort construction derived from the corrected AI Clinician preprocessing workflow (Singer et al., 2016; Komorowski et al., 2018). Because the public implementation was developed for MIMIC-III, we adapt its cohort-construction logic to MIMIC-IV and then construct PatientWorld-specific sequential decision trajectories.

Suspected infection is identified by pairing antibiotic administration with a culture event occurring within $[-72, +24]$ h relative to the antibiotic. The sepsis reference time is defined as the first timestamp within the suspected-infection window at which SOFA ≥ 2. To avoid using information unavailable at decision time, the first eligible decision is aligned to the first 4-h boundary no earlier than both the suspected-infection known time and the SOFA availability time.

Starting from this point, we construct a decision window of up to 48 h. At each transition, longitudinal information available through the current decision time defines the patient history, treatments recorded during the following 4 h define the current action, and physiology at the next 4-h step defines the prediction target. The sepsis action space combines five intravenous-fluid bins and five vasopressor bins, yielding 25 discrete treatment actions.

RRT initiation timing. The RRT initiation task follows the sequential decision setting of Grolleau et al. (Grolleau et al., 2024), adapted to MIMIC-IV and the 4-h PatientWorld trajectory representation. The cohort is anchored at the first pre-RRT KDIGO stage-3 state (Khwaja, 2012) and includes only patients who have not yet received active RRT at that time. Candidate decision points are constructed at 0, 24, and 48 h after the anchor and retained only when the patient is alive, the required historical information is available, and active RRT has not previously started.

To focus on non-emergent RRT initiation timing, we exclude decision points with severe hyperkalemia ($K \geq 6.0\ \mathrm{mmol/L}$) or severe metabolic acidosis ($\mathrm{pH} \leq 7.20$ or $\mathrm{HCO_3^-} \leq 12\ \mathrm{mmol/L}$). ICU stays with ongoing RRT before the beginning of the pretraining trajectory are also excluded, so all formal decision rows precede the first active RRT event. The task uses a binary action space: *defer RRT* or *initiate RRT* within the following 24 h.

B.4 Downstream Cohort Statistics

Table 5 summarizes the formal downstream decision datasets used for task-specific adaptation and downstream evaluation. These counts correspond to the complete downstream datasets before OPE-specific eligibility filtering.

Sepsis uses the formal test replay directly for offline policy evaluation. For RRT, the formal test split contains 3,777 episodes and 8,007 decision rows; OPE is anchored at stage 0, yielding 3,711 eligible episodes with a valid stage-0 decision row. HFD60 is fully observed in the formal test cohort and therefore causes no additional episode-level exclusion. The complete OPE protocol is described in Appendix G.

C Physiological Variable Partition and Treatment Definitions

C.1 Physiological Variable Partition

Critical care assessment commonly organizes patient status around organ and functional systems (Vincent et al., 1996). Following this clinical organization and the physiological factoriza-

[1] `https://github.com/microsoft/mimic_sepsis`

Table 5: **Downstream cohort statistics.** Counts correspond to the formal downstream decision datasets before OPE-specific eligibility filtering.

Task	Split	Patients	ICU stays / episodes	Decision rows
Sepsis	Train	22,883	28,028	253,574
	Valid	2,756	3,318	29,557
	Test	2,842	3,505	31,802
	Total	**28,481**	**34,851**	**314,933**
RRT	Train	23,680	29,918	63,310
	Valid	2,920	3,637	7,615
	Test	2,995	3,777	8,007
	Total	**29,595**	**37,332**	**78,932**

Table 6: **Physiological variable partition used by PatientWorld.** Variables are assigned to physiological factors before encoding.

Factor	EHR variables
Respiratory (R)	RespRate, SpO2, FiO2, PaO2, PaCO2
Cardiovascular (C)	HeartRate, SysBP, DiaBP, MeanBP, ShockIndex
Renal/electrolyte (K)	Creatinine, BUN, UrineOutput, Sodium, Potassium
Metabolic/acid-base (M)	Chloride, Bicarbonate, Glucose, Lactate, pH
Clinical context (G)	Temperature, GCS_Eye, GCS_Verbal, GCS_Motor, GCS_Total, Hemoglobin, WBC, Platelets, Bilirubin

tion introduced in Sec. 4.1, we partition the observed variables into $S = \{R, C, K, M, G\}$, corresponding to respiratory, cardiovascular, renal/electrolyte, metabolic/acid-base, and clinical context, respectively. The partition is defined before encoding and is shared across pretraining and downstream adaptation. Table 6 lists the complete variable assignment used in our experiments.

C.2 Generic and Physiology-Specific Treatment Dynamics

As described in Sec. 4.1, PatientWorld explicitly partitions the treatment type set into generic treatments (J_G) and physiology-specific treatments (J_P). Generic treatments are integrated directly into the cross-system transition context without imposing predefined physiological targets. For each physiology-specific treatment $j \in J_P$, PatientWorld uses a direct pathway for predefined primary target systems $S_j \subset S$ and a spillover pathway for secondary effects.

The pretraining action schema uses audited active treatment families extracted on the 4-hour grid: intravenous fluids, vasopressors/inotropes, respiratory support, diuretics, insulin, renal replacement therapy, and blood-product transfusion. Among these, respiratory support, diuretics, insulin, renal replacement therapy, and blood-product transfusion use the physiology-specific treatment pathways described below, while intravenous fluids and vasopressors/inotropes remain under the generic treatment-conditioning pathway.

Table 7 lists the physiology-specific treatments, their action definitions, and the primary target-system mappings S_j used in our implementation. The selected primary target sets are motivated by established physiological roles and treatment–system interactions in critical care. Specifically, insulin is assigned to the metabolic system based on its role in glucose regulation (Langouche & Van den Berghe, 2006); diuretics are assigned to the renal and cardiovascular systems through their effects on renal sodium and fluid handling and cardiovascular volume status (Felker et al., 2020); respiratory support is assigned to the respiratory and cardiovascular systems to reflect respiratory–cardiovascular interactions during ventilatory support (Kreit, 2022); and renal replacement therapy is assigned to the renal and metabolic systems through its role in solute, electrolyte, and acid–base regulation (Khwaja, 2012). Blood-product transfusion is represented as an aggregate channel spanning packed red blood cells (PRBC), fresh frozen plasma (FFP), platelets, and cryoprecipitate, motivating its cardiovascular and hematologic/systemic targets (Cohn & Shaz, 2023; Vlaar et al., 2020).

These mappings specify primary physiological targets used as structural priors and are not intended to represent exhaustive treatment effects. Treatments without a designated primary target set, including intravenous fluids and vasopressors/inotropes, remain under the generic treatment-conditioning pathway J_G.

Table 7: **Physiology-specific treatments and primary target-system mappings.** Active treatment families without a primary target set belong to J_G and are represented through the generic treatment-conditioning pathway.

Physiology-specific treatment	Action definition	S_j
Insulin	Total administered insulin units within each time bin	M
Diuretics	Total furosemide-equivalent dose within each time bin	K, C
Respiratory support	Maximum ordinal respiratory-support level within each time bin	R, C
Renal replacement therapy	Fraction of the time bin with active renal replacement therapy	K, M
Blood-product transfusion	Total transfusion volume within each time bin	C, G

Treatment representation and direct/spillover pathways. For each physiology-specific treatment $j \in J_P$, the model first constructs a treatment message. The state gate conditions on the mean latent states of the treatment-associated systems and the clinical-context group:

$$\mathbf{g}_t^{(j)} = \sigma\Big(F_{\text{gate}}\left((\mathbf{z}_{t-1}^{s'})_{s' \in S_j \cup \{G\}}\right)\Big),$$

where G provides clinical-context information for gating but is not treated as an additional direct target system. The treatment-intensity encoder receives the five scalar treatment features used in the implementation:

$$\mathbf{u}_t^{(j)} = F_{\text{treat}}^{(j)}\left(v_{t-1}^{(j)}, u_{t-1}^{(j)}, e_{t-1}^{(j)}, \text{start}_{t-1}^{(j)}, \text{stop}_{t-1}^{(j)}\right).$$

The treatment message is

$$\mathbf{m}_t^{(j)} = \mathbf{g}_t^{(j)} \odot \mathbf{u}_t^{(j)}.$$

Both F_{gate} and $F_{\text{treat}}^{(j)}$ are two-layer MLPs; F_{gate} uses a sigmoid output and $F_{\text{treat}}^{(j)}$ uses a $\tanh$ output.

The direct pathway applies treatment j only to its predefined target systems:

$$\mathbf{e}_t^{\text{direct},s,(j)} = D^{s,(j)}(\mathbf{m}_t^{(j)}), \qquad s \in S_j,$$

where $D^{s,(j)}$ is a treatment- and system-specific linear projection. The spillover pathway uses a separate treatment-specific projection to produce weaker residual effects over the factorized state space:

$$(\mathbf{e}_t^{\text{spill},s,(j)})_{s \in S'} = B^{(j)}(\mathbf{m}_t^{(j)}).$$

The final physiology-specific treatment embedding is aggregated across all physiology-specific treatments:

$$\mathbf{e}_t^{\text{ps},s} = \sum_{j \in J_P} \left[\mathbf{1}(s \in S_j)\mathbf{e}_t^{\text{direct},s,(j)} + \alpha_j \mathbf{e}_t^{\text{spill},s,(j)}\right].$$

In the main text, α_j is denoted compactly as the small spillover scale α. In our implementation, the default spillover scale is 0.1, while respiratory support uses a zero spillover scale. The spillover projection also produces an output for Global Context X, whereas X does not receive the local-history residual.

In our implementation, α corresponds to the fixed residual scale in the family-conditioned treatment module. The default value is 0.1; respiratory support uses a zero spillover scale in the reported configuration. The spillover projection additionally contains an output for Global Context X, whereas X does not receive the state-dependent local-history residual defined in Sec. 4.1.

D Architecture and Pretraining Details

D.1 Model Architecture

PatientWorld maintains deterministic and stochastic latent states over $S' = \{R, C, K, M, G, X\}$. For each observed physiological factor $s \in S = \{R, C, K, M, G\}$, the encoder receives the normalized variables assigned to that factor together with their observation masks and $\log(1 + \boldsymbol{\delta}_t^s)$ features. Global Context X is encoded from the full physiological observation and serves as a non-decoding shared context for cross-system transition modeling; physiological predictions remain factor-specific. Each factor uses a 128-dimensional deterministic state $\mathbf{h}_t^s$ and a 32-dimensional stochastic state $\mathbf{z}_t^s$.

Cross-system context mixer. For the physiology-factorized pretraining model, the cross-system context mixer receives the concatenation of the previous stochastic states from all six factors (R, C, K, M, G, X) and the generic treatments $\{\mathbf{a}_{t-1}^{(j)}\}_{j \in J_G}$. The generic treatment representations reserve three features for each canonical treatment channel: the normalized treatment value, availability indicator, and event indicator. For treatment families handled by the physiology-specific treatment pathways (J_P), the corresponding generic-pathway feature triplets are zeroed, so that these treatments enter the transition strictly through the direct and spillover pathways rather than the generic conditioning pathway. The mixer is a shared MLP consisting of $\mathrm{Linear} \to \mathrm{LayerNorm} \to \mathrm{SiLU} \to \mathrm{Linear}$. Its output is reshaped into six target-specific cross-system transition contexts $(\mathbf{c}_{t-1}^s)_{s \in S'}$. With six 32-dimensional stochastic states and seven three-feature treatment-action slots, the mixer input has dimension 213, and its output comprises six 128-dimensional contexts. The mixer parameters are shared across target factors, while the subsequent recurrent transitions remain factor specific.

Factor-specific recurrent transition. For each observed physiological factor $s \in S$, the generic transition component is implemented by a factor-specific GRU cell:

$$\mathbf{e}_t^{\mathrm{gen},s} = \mathrm{GRU}^s\left(\mathbf{h}_{t-1}^s, \left[\mathbf{z}_{t-1}^s; \mathbf{c}_{t-1}^s; \mathbf{a}_{t-1}^{\mathrm{gen}}\right]\right),$$

where $\mathbf{a}_{t-1}^{\mathrm{gen}}$ denotes the pretraining treatment vector after removing the physiology-specific treatment channels handled by the direct/spillover pathways. Global Context X uses an analogous recurrent transition with its own parameters, but it is not decoded into physiological observations.

Each observed factor then applies its own recurrent local transition conditioned on its current stochastic state, cross-system transition context, and generic treatment input. The physiology-specific treatment effect $\mathbf{e}_t^{\mathrm{ps},s}$ and the state-dependent local transition residual $\mathbf{r}_t^s$ described in Sec. 4.1 are added to this local recurrent update before the next stochastic prior is formed. Global Context X has its own recurrent transition and stochastic prior but no direct physiological observation decoder.

Spillover Scale (α). As introduced in Sec. 4.1, the spillover pathway allows physiology-specific treatments to weakly influence non-target physiological systems. In our implementation, the spillover scale α is set to a fixed coefficient of 0.1 by default rather than a learned scalar. The residual spillover pathway for respiratory support is explicitly disabled by setting its coefficient to 0. Global Context X also receives a weak spillover contribution but does not receive the state-dependent local-history residual defined in Sec. 4.1.

For posterior inference, factor-specific observations update the stochastic latent states. During open-loop rollout, future observations are unavailable and the transition prior is propagated recursively; its mean is used as the stochastic state at the subsequent rollout step. Each observed factor predicts physiological values and observation masks, with an additional anchor-to-future delta head used only for auxiliary supervision.

The local trajectory summary $\mathbf{l}_{\mathrm{hist}}^s$ contains five per-variable statistics: the latest observed value, local slope, volatility, missingness, and time since the most recent observation. In addition to the compact formulation in Sec. 4.1, the implementation supplies a learned relative prediction-horizon embedding $\mathbf{p}_k$ to both the residual gate and correction:

$$\rho_t^s = \sigma\big(F_{\text{hist-gate}}^s(\mathbf{z}_{t-1}^s, \mathbf{l}_{\mathrm{hist}}^s, \mathbf{p}_k)\big), \qquad \boldsymbol{\epsilon}_t^s = \mathrm{clip}\big(F_{\text{hist-res}}^s(\mathbf{z}_{t-1}^s, \mathbf{l}_{\mathrm{hist}}^s, \mathbf{e}_t^{\mathrm{ps},s}, \mathbf{p}_k), -1, 1\big). \quad (2)$$

Table 8: **PatientWorld pretraining architecture.** Implementation-specific settings not fully specified in the main text.

Component	Setting
Physiological factors	R, C, K, M, G, with non-decoding global context X
Deterministic state	128 dimensions per factor
Stochastic state	32 dimensions per factor
Observed-factor input	Normalized values, observation masks, and $\log(1+\boldsymbol{\delta}_t^s)$
Open-loop stochastic state	Transition-prior mean
Prediction heads	Factor-specific value and observation-mask heads; auxiliary delta head
Local trajectory summary	Latest value, slope, volatility, missingness, and time since last observation
Local residual conditioning	Current factor state, trajectory summary, physiology-specific treatment effect, and rollout-step embedding

Here, $\mathbf{p}_k$ is indexed by the current rollout step k, rather than by the configured forecasting horizon. It is used only by the local residual gate and correction networks and does not enter the shared cross-system mixer, the factor-specific recurrent transition f^s_{transit}, or the latent prior/posterior parameterizations. The resulting residual is $\mathbf{r}_t^s = \rho_t^s \boldsymbol{\epsilon}_t^s$. The gate is scalar-valued for each physiological factor, whereas the correction has the same dimensionality as the deterministic state. Global Context X does not receive this local-history residual.

D.2 Pretraining Objective

The compact objective in Eq. 1 combines posterior-conditioned prediction, open-loop prior rollout, latent overshooting, and auxiliary physiological supervision.

Posterior-conditioned prediction. The posterior branch is supervised by future physiological-value prediction, future observation-mask prediction, and one-step prior–posterior consistency:

$$\mathcal{L}_{\text{posterior}} = w_{\text{post}} \left(\mathcal{L}_{\text{post,val}} + \alpha_{\text{mask}} \mathcal{L}_{\text{post,mask}} + \beta_{\text{post}} \mathcal{L}_{\text{post,KL}} \right). \tag{3}$$

The value term is masked mean squared error on normalized observed targets:

$$\mathcal{L}_{\text{post,val}} = \frac{\sum_{t,i} m_{t,i} \left(\hat{x}^{\text{post}}_{t,i} - x_{t,i} \right)^2}{\sum_{t,i} m_{t,i}}, \tag{4}$$

where $m_{t,i}$ denotes the future observation mask. $\mathcal{L}_{\text{post,mask}}$ is binary cross entropy on future observation masks, and $\mathcal{L}_{\text{post,KL}}$ penalizes disagreement between the one-step transition prior and the corresponding filtering posterior.

Open-loop rollout and latent overshooting. Starting from the same observed history, the rollout branch recursively propagates transition priors under the subsequent treatment sequence and optimizes

$$\mathcal{L}_{\text{rollout}} = \mathcal{L}_{\text{roll,val}} + \alpha_{\text{mask}} \mathcal{L}_{\text{roll,mask}}, \tag{5}$$

where the value and mask terms use the same targets and masking rules as the posterior branch. Latent overshooting, $\mathcal{L}_{\text{overshoot}}$, penalizes KL disagreement between recursively propagated rollout priors and future observation-conditioned posteriors. The first forecast step is excluded because one-step prior–posterior consistency is already included in $\mathcal{L}_{\text{posterior}}$.

Auxiliary physiological supervision. The auxiliary objective is

$$\mathcal{L}_{\text{aux}} = \lambda_{\text{level}} \mathcal{L}_{\text{level}} + \lambda_{\Delta} \mathcal{L}_{\Delta} + \lambda_{\text{gate}} \mathcal{L}_{\text{gate}} + \lambda_{\text{non}} \mathcal{L}_{\text{non}}. \tag{6}$$

$\mathcal{L}_{\text{level}}$ supervises factor-specific physiological levels, while $\mathcal{L}_{\Delta}$ supervises changes from the history anchor to future physiological values; both are evaluated only at observed targets.

The local-transition residual is regularized by $\mathcal{L}_{\text{gate}}$ and $\mathcal{L}_{\text{non}}$. Event targets are constructed from training windows only. A physiological factor is marked as an event at a future step when at least one

Table 9: **Pretraining loss weights.**

Learning signal	Coefficient	Value
Posterior prediction	w_{post}	1.5
Mask prediction	α_{mask}	0.35
Posterior prior–posterior consistency	β_{post}	10^{-4}
Open-loop rollout	λ_{rollout}	0.5
Latent overshooting	$\lambda_{\text{overshoot}}$	10^{-4}
Subsystem level prediction	λ_{level}	0.02
Subsystem delta prediction	λ_{Δ}	0.02
Local residual gate supervision	λ_{gate}	0.05
Non-event residual penalty	λ_{non}	0.05

Table 10: **Pretraining optimization and training configuration.** Settings are shared across MIMIC-III and MIMIC-IV unless otherwise indicated.

Setting	Value
Optimizer	AdamW
Learning rate	2×10^{-4}
Weight decay	10^{-5}
Batch size	256
Maximum epochs	30
Gradient clipping	5.0
Lookback length	12 bins
Forecast horizon	3 or 6 bins (separate models)
Maximum windows per trajectory	8
Early stopping	Patience 12 (MIMIC-III); disabled (MIMIC-IV)
Checkpoint criterion	Validation normalized RMSE

observed variable assigned to that factor exhibits either an extreme normalized level or an extreme change from the history anchor:

$$y_t^s = \mathbf{1}\left[\exists i \in \mathcal{O}_s : m_{t,i} = 1 \wedge \left(|\tilde{x}_{t,i}| \geq q_{0.99}^{i} \vee |\tilde{x}_{t,i} - \tilde{x}_{\text{anchor},i}| \geq q_{0.95}^{\Delta,i}\right)\right]. \tag{7}$$

$\mathcal{L}_{\text{gate}}$ is binary cross entropy against these factor-level event targets, while $\mathcal{L}_{\text{non}}$ penalizes residual magnitude at non-event factor–time positions.

The complete pretraining objective is

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{posterior}} + \lambda_{\text{rollout}}\mathcal{L}_{\text{rollout}} + \lambda_{\text{overshoot}}\mathcal{L}_{\text{overshoot}} + \mathcal{L}_{\text{aux}}, \tag{8}$$

which expands Eq. 1. The coefficients used in the pretraining experiments are listed in Table 9.

D.3 Optimization and Training Configuration

PatientWorld is pretrained separately on MIMIC-III and MIMIC-IV using the same architecture and pretraining objective. The forecasting experiments use a 12-bin observation history, corresponding to 48 h, and H3 and H6 are trained as separate models with forecast horizons of 3 and 6 bins, respectively. Within each configuration, multi-step predictions are generated by recursively propagating the transition prior over the corresponding rollout horizon. Each ICU trajectory contributes at most eight sampled training windows. The checkpoint with the lowest validation normalized RMSE is retained. Optimization settings are summarized in Table 10.

E Task-Specific Context Adaptation Details

E.1 Task-Specific Physiological Relation Graphs

For each downstream task τ, PatientWorld defines a sparse directed physiological relation graph $\mathcal{G}_\tau = (S, E_\tau)$ over the observed physiological factors $S = \{R, C, K, M, G\}$. The graph specifies

which cross-system relations receive additional task-specific adaptation, while the pretrained cross-system dynamics remain fixed.

The relation sets are fixed task-level physiological priors motivated by established cross-system interactions relevant to each setting and are not selected or tuned based on validation or test performance.

Respiratory–cardiovascular interactions are well established during mechanical ventilation (Kreit, 2022); cardiorenal dysfunction involves bidirectional heart–kidney interactions (Rangaswami et al., 2019); and lung–kidney crosstalk links respiratory dysfunction, mechanical ventilation, acute kidney injury, and fluid balance in critical illness (Joannidis et al., 2020). Renal dysfunction is also closely associated with electrolyte and acid–base disturbances (Di Lullo et al., 2019), while severe metabolic acidosis can impair cardiac contractility and vascular responsiveness (Kimmoun et al., 2015). In sepsis, these interactions occur within broader multi-organ crosstalk associated with organ dysfunction (Borges & Bento, 2024).

For sepsis treatment, the directed physiological relation set is

$$E_{\text{sepsis}} = \{R \to C,\ C \to R,\ C \to K,\ K \to C,\ K \to M,\ M \to C\}. \tag{9}$$

For RRT initiation timing, the directed physiological relation set is

$$E_{\text{RRT}} = \{C \to K,\ K \to C,\ R \to K,\ K \to R,\ K \to M,\ M \to C\}. \tag{10}$$

The two graphs share cardiorenal and renal–metabolic–cardiovascular relations, while the sepsis graph additionally includes respiratory–cardiovascular coupling and the RRT graph instead includes lung–kidney coupling.

The RRT task additionally uses action-to-system priors $\text{RRT} \to K$ and $\text{RRT} \to M$ to designate preferred renal and metabolic targets for the task-specific action residual. These entries are separate from the relational graph and do not impose hard routing. Directed graph edges specify information flow within the adapter and should not be interpreted as causal directions between physiological systems. The task graphs and action-system priors are fixed specifications and are not learned during downstream adaptation.

E.2 Relational and Action Residuals

The downstream adapter modifies the pretrained transition context through two complementary components: a graph-conditioned relational residual and a state-conditioned action residual. For target factor s, the adapted context is

$$\tilde{\mathbf{c}}^{\tau,s}_{t-1} = \mathbf{c}^{s}_{t-1} + \mathbf{r}^{s}_{\text{rel}}\left(\{\mathbf{z}^{s'}_{t-1}\}_{s':(s',s)\in E_\tau}\right) + \mathbf{r}^{s}_{\text{act}}\left(\mathbf{z}^{s}_{t-1}, \{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau}\right), \tag{11}$$

where $\mathbf{c}^{s}_{t-1}$ is the frozen pretrained transition context.

Relational residual. For each directed physiological relation $(s', s) \in E_\tau$, an edge-specific mapping transforms the source latent state into the transition-context space:

$$\mathbf{g}^{(\tau)}_{s'\to s,t-1} = F^{(\tau)}_{s'\to s}(\mathbf{z}^{s'}_{t-1}). \tag{12}$$

The relational residual aggregates incoming task-specific messages as

$$\mathbf{r}^{s}_{\text{rel},t-1} = \sum_{(s',s)\in E_\tau} \alpha^{(\tau)}_{s',s}\mathbf{g}^{(\tau)}_{s'\to s,t-1}, \tag{13}$$

where $\alpha^{(\tau)}_{s',s}$ denotes the effective learned soft scale for relation (s', s). The message mappings are lightweight two-layer MLPs with SiLU nonlinearities.

State-conditioned action residual. Let $\{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau}$ denote the task-specific action representation provided to the adapter. For each physiological factor, a learned scalar gate modulates a vector-valued correction conditioned on the current factor state and task action:

$$\rho^{\tau,s}_{\text{act}} = \sigma\left(\Gamma^{\tau,s}_{\text{act}}(\mathbf{z}^{s}_{t-1}, \{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau})\right), \qquad \boldsymbol{\epsilon}^{\tau,s}_{\text{act}} = \Psi^{\tau,s}_{\text{act}}\left(\mathbf{z}^{s}_{t-1}, \{\mathbf{a}^{(j)}_{t-1}\}_{j\in J_\tau}\right). \tag{14}$$

The resulting residual is $\mathbf{r}^{s}_{\text{act},t-1} = \rho^{\tau,s}_{\text{act}}\boldsymbol{\epsilon}^{\tau,s}_{\text{act}}$. The gate is scalar-valued for each physiological factor, whereas the correction has the same dimensionality as the deterministic state. Global Context X does not receive this residual.

E.3 Adaptation Objective and Optimization

During adaptation, all inherited pretrained parameters Θ_{pre} are frozen, and only the task-specific relational and treatment-residual parameters, collectively denoted by η_τ, are optimized. Adaptation remains a predictive world-model objective and does not directly optimize reward, planning, or policy parameters.

Let $\hat{\mathbf{x}}_{i,t+k}$ and $\hat{\mathbf{m}}_{i,t+k}$ denote the predicted normalized physiology values and future observation-mask logits for patient window i at horizon step k. Let $\mathbf{x}_{i,t+k}$ be the normalized target physiology, $\mathbf{m}_{i,t+k} \in \{0,1\}^{D_x}$ the observed-variable mask, and $v_{i,k} \in \{0,1\}$ the valid-horizon mask. We define $\omega_{i,k,d} = v_{i,k} m_{i,t+k,d}$, so value losses are evaluated only on observed target variables. The one-step and multi-step prediction losses are

$$\mathcal{L}_{\text{1step}} = \frac{\sum_{i,d} \omega_{i,1,d} \left(\hat{x}_{i,t+1,d} - x_{i,t+1,d}\right)^2}{\sum_{i,d} \omega_{i,1,d} + \epsilon}, \tag{15}$$

$$\mathcal{L}_{\text{multi}} = \frac{\sum_{i,k,d} \omega_{i,k,d} \left(\hat{x}_{i,t+k,d} - x_{i,t+k,d}\right)^2}{\sum_{i,k,d} \omega_{i,k,d} + \epsilon}. \tag{16}$$

The future-mask prediction loss is the horizon-masked binary cross-entropy:

$$\mathcal{L}_{\text{mask}} = \frac{\sum_{i,k,d} v_{i,k} \ \text{BCEWithLogits}\left(\hat{m}_{i,t+k,d}, m_{i,t+k,d}\right)}{D_x \sum_{i,k} v_{i,k} + \epsilon}. \tag{17}$$

The adaptation objective is

$$\mathcal{L}_{\text{adapt}} = \mathcal{L}_{\text{1step}} + \lambda_{\text{multi}} \mathcal{L}_{\text{multi}} + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}}. \tag{18}$$

For both sepsis and RRT adaptation, we set $\lambda_{\text{multi}} = 1.0$ and $\lambda_{\text{mask}} = 0.1$. Sepsis adaptation uses Adam with learning rate 10^{-3} and batch size 128, with checkpoint selection based on the lowest validation normalized RMSE at horizon 6. RRT adaptation uses Adam with learning rate 10^{-4} and training batch size 64 for 10 epochs, with checkpoint selection based on validation balanced renal/metabolic RMSE at horizon 6. No early stopping is used in the reported configurations. The task graphs and treatment priors remain fixed throughout adaptation; validation is used only for checkpoint selection under these fixed task specifications.

F Baseline, Ablation, and Implementation Details

This appendix documents the implementation choices specific to the comparison methods in the main experiments. We retain the original method names in the main tables for readability and describe here the adaptations required to place them under a common longitudinal EHR forecasting and downstream adaptation interface.

F.1 Common Trajectory Forecasting Protocol

All trajectory forecasting methods use the same processed MIMIC-III and MIMIC-IV patient-level partitions and the 4-hour trajectory representation described in Appendix B. Models observe the preceding 12 bins and are trained separately for the H3 and H6 forecasting settings. All methods are evaluated with the same normalized RMSE and MAE computation over observed future values. In the conditional setting, recorded future treatment information is provided through the corresponding model-specific conditioning interface, whereas future treatment information is withheld in the blind setting. Thus, the trajectory forecasting comparisons in Table 1 use the same cohort construction, forecasting horizons, and evaluation protocol across methods.

F.2 Forecasting Baseline Implementations

Table 11 summarizes the implementation provenance and the nontrivial adaptations used for the trajectory forecasting comparisons. For standard forecasting architectures, we reimplement the corresponding forecasting backbone under the common EHR interface, adapting only the input, output,

Method	Initialization	Implementation and adaptation
GRU (Cho et al., 2014), Transformer (Vaswani et al., 2017), DLinear (Zeng et al., 2023), PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024)	From scratch	We reimplement the corresponding forecasting architectures under the common EHR trajectory interface. Inputs and prediction heads are adapted to the shared multivariate representation, and conditional forecasting incorporates recorded future-treatment context through a lightweight conditioning pathway.
MOMENT (Goswami et al., 2024)	Pretrained frozen backbone	We use the pretrained MOMENT backbone as a frozen representation model and train an ICU-specific forecasting head on the common trajectory representation.
Moirai 2.0 (Liu et al., 2025)	Pretrained frozen backbone	We keep the pretrained Moirai base frozen and learn a residual forecasting and mask adapter for the common EHR output space, with future-treatment conditioning supplied through the adapted forecasting interface.
TFT-multi (He & Chiang, 2025)	From scratch	We adapt the public implementation to the common historical and known-future EHR inputs. Its output is converted to point forecasting over the shared multivariate target space for evaluation with normalized RMSE and MAE.
TDSTF (Chang et al., 2024)	From scratch	We adapt the public implementation by converting dense EHR windows into its sparse feature–time–value representation. In the conditional setting, future-treatment events are appended to the conditioning input, while evaluation uses the shared forecasting metrics.
Phys-JEPA (Nie et al., 2026)	From scratch	We reimplement the published latent-predictive architecture, retaining its context/target encoding and latent prediction structure, and add the supervised forecasting interface required by the common EHR benchmark.
MedDreamer (Xu et al., 2026)	From scratch	We retain the DenseRSSM world-model implementation and train it from scratch on the same fixed-bin ICU trajectory representation used by the other forecasting methods. Trajectory forecasting uses the dynamics model only; reward modeling, actor–critic learning, and downstream imagination-based policy optimization are excluded.

Table 11: **Implementation of trajectory forecasting baselines.** The table reports initialization and the main adaptations required to evaluate each method under the common longitudinal EHR forecasting protocol. Routine optimization details are omitted when they do not alter the method-specific comparison.

and treatment-conditioning interfaces required by our multivariate trajectory setting. For methods with released pretrained models or task-specific input formats, we preserve their characteristic modeling components while using the common preprocessed trajectory representation and the minimal model-specific interface required by the shared forecasting protocol. Accordingly, the comparison standardizes the cohort, forecasting horizons, target space, treatment-availability setting, and evaluation metrics, while preserving each method's initialization regime and model-specific forecasting interface.

MedDreamer world-model baseline. For trajectory forecasting, we compare against the world-model component of MedDreamer rather than its downstream policy-learning pipeline. Specifically, we retain the DenseRSSM dynamics implementation and train it from scratch on the same fixed-bin ICU trajectories used by the other forecasting methods. Recorded treatments enter through the DenseRSSM action interface in conditional forecasting. Reward modeling, actor–critic learning, and latent-imagination policy optimization are not used in this comparison. This construction evaluates MedDreamer's latent dynamics under the same trajectory-prediction setting used for PatientWorld.

F.3 Baseline Hyperparameters and Optimization

All trainable trajectory forecasting baselines are trained on the same fixed patient-level train, validation, and test splits as PatientWorld. Sequence forecasting baselines are optimized using AdamW and selected based on the lowest validation total loss, which combines normalized value prediction loss and future observation-mask prediction loss. Table 12 summarizes the exact hyperparameter configurations applied to the standard sequence forecasting architectures (e.g., Transformer, PatchTST, GRU, iTransformer, and DLinear) to ensure a fair and consistent comparison.

Table 12: **Optimization configurations for trajectory forecasting baselines.** These represent the actual training hyperparameters and architectural bounds used to evaluate the sequence forecasting baselines.

Hyperparameter	Configuration
Optimizer	AdamW
Learning rate	3×10^{-4} (default); 10^{-4} for source-specific sensitivity runs
Weight decay	10^{-5}
Batch size	128
Maximum epochs	30
Early stopping	Disabled (best checkpoint saved during training)
Checkpoint criterion	Lowest validation total loss
Hidden dimension	256
Number of layers	3
Number of attention heads	8 (for attention-based baselines)
Dropout	0.1
Patch length / stride	3 / 1 (for PatchTST)
Gradient clipping	5.0

F.4 Pretraining Ablation Definitions

The pretraining ablations examine the structural controls evaluated in Fig. 3. Table 13 summarizes the corresponding architectural variants. *Base WM* provides a monolithic architectural reference using a dense RSSM without physiological factorization, physiology-specific treatment routing, or the state-dependent local residual. To ensure a fair comparison and isolate the effect of physiological factorization, *Base WM* is scaled to maintain a comparable total parameter count and latent capacity to the factorized variants. *PW w/o Treat. & Local* replaces this monolithic reference with physiology-factorized latent states and transitions while removing both the physiology-specific treatment pathways and the local residual. *PW w/o Local* additionally introduces the physiology-aligned physiology-specific treatment pathways but omits the state-dependent local residual. *Shuffled Treat.* retains the complete PatientWorld architecture, including the state-dependent local residual, but permutes the primary target-system sets S_j across physiology-specific treatment families. This preserves all other model components while breaking the intended treatment–physiology correspondence. *PatientWorld* retains the physiology-aligned assignments and all components. The cross-system context mixer is unchanged across the physiology-factorized variants.

F.5 Global-Context Analysis

Because the non-decoding global context state X aggregates the full physiological observation for cross-system transition modeling, we separately examine the role of this shared context in trajectory forecasting. As shown in Fig. 5, removing X produces only small increases in mean forecasting error across both datasets and conditioning settings, while the full PatientWorld model achieves lower RMSE and MAE in all eight H3 comparisons. These results indicate that X provides a modest but consistent complementary benefit beyond the physiology-specific latent factors.

F.6 Rollout-Loss Weight Sensitivity

The pretraining objective in Eq. 1 uses λ_{rollout} to weight recursive open-loop rollout supervision. We evaluate $\lambda_{\text{rollout}} \in \{0.25, 0.5, 1.0\}$ on MIMIC-III and MIMIC-IV at H3, with all other training settings unchanged. As shown in Fig. 6, conditional and blind Global RMSE and MAE vary only

Variant	Physio. factorization	Physiology-specific routing	Primary mapping	Local residual
Base WM	No	No	–	No
PW w/o Treat. & Local	Yes	No	–	No
PW w/o Local	Yes	Yes	Physiology-aligned	No
Shuffled Treat.	Yes	Yes	Shuffled $\mathcal{T}_j$	Yes
PatientWorld	Yes	Yes	Physiology-aligned	Yes

Table 13: **Definitions of the pretraining ablations.** The variants examine physiological state factorization, the selective treatment pathway, treatment-to-system alignment, and the state-dependent local residual.

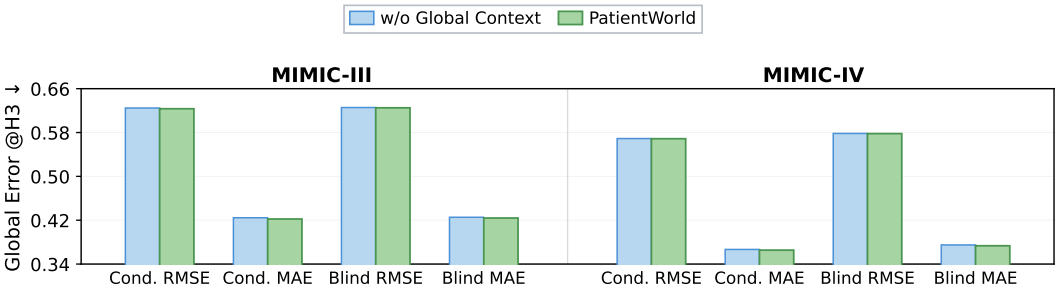


Figure 5: **Global-context analysis for trajectory forecasting.** Global RMSE and MAE at H3 are reported on MIMIC-III and MIMIC-IV under conditional and blind forecasting for PatientWorld with and without the global context state X. Bars show three-seed means; lower is better.

slightly across the tested range. The default $\lambda_{\text{rollout}} = 0.5$ yields the lowest mean error in six of eight comparisons; in the two Blind RMSE exceptions, the best alternative differs from the default by only 0.0001. These results indicate that H3 forecasting shows limited sensitivity to the rollout-loss weight within the tested range.

F.7 Multi-Horizon Evaluation of the H6 Checkpoint

To examine whether the longer-horizon pretrained model remains usable at shorter rollout horizons, we evaluate the pretrained H6 PatientWorld configuration at cumulative horizons H1, H3, and H6 without retraining. Table 14 reports normalized RMSE pooled over rollout steps $1{:}H$. Across both datasets and conditioning settings, the H6-trained configuration maintains accurate shorter-horizon forecasts. At H3, it matches or improves upon the separately trained H3 configuration in all four dataset–conditioning settings. These results show that, within the evaluated H1–H6 range, the H6-trained model can be used directly for shorter-horizon prediction without horizon-specific retraining.

F.8 Downstream Reuse Baselines

We compare PatientWorld Adapt with four alternative downstream training and reuse strategies. *From Scratch* uses the same downstream task architecture but initializes it randomly and trains only on the corresponding downstream data, without loading pretrained weights. *Zero-shot* loads the pretrained PatientWorld checkpoint, keeps the pretrained parameters fixed, and performs no task-specific adaptation. *Full Fine-tuning* initializes from the pretrained checkpoint and jointly optimizes the pretrained backbone and the downstream task-specific modules. *Random Graph Adapt* freezes the pretrained backbone and uses the same relational and state-conditioned action-residual adaptation mechanism as PatientWorld Adapt, but replaces the task-specific physiological relation structure with randomized controls. PatientWorld Adapt instead uses the task-specific physiological relations defined in Appendix E. All trained variants use the same task-specific predictive supervision.

For each downstream task and adaptation seed, Random Graph Adapt samples a randomized directed relation structure before training and keeps it fixed throughout the corresponding run. The random graph is sampled without replacement from the directed non-self edge space over the same

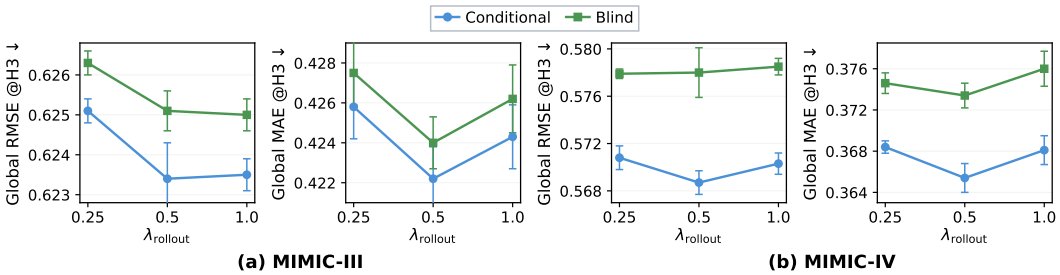


Figure 6: **Sensitivity to the rollout-loss weight.** Conditional and blind Global RMSE and MAE at H3 are reported on MIMIC-III and MIMIC-IV for $\lambda_{\text{rollout}} \in \{0.25, 0.5, 1.0\}$. Points show three-seed means and error bars denote standard deviations; lower is better. Tested values are displayed at equal spacing for readability.

Table 14: **Multi-horizon evaluation of the pretrained H6 PatientWorld model.** H6-trained PatientWorld checkpoints are evaluated without retraining at cumulative horizons H1, H3, and H6, where normalized RMSE pools observed targets over rollout steps $1{:}H$. H3-trained @ H3 denotes the separately trained H3 configuration reported in Table 1. Values are mean ± standard deviation over three seeds; lower is better.

Dataset	Setting	H3-trained @ H3	H6-trained @ H1	H6-trained @ H3	H6-trained @ H6
MIMIC-III	Conditional	0.6234 ± .0009	0.5296 ± .0010	0.5911 ± .0012	0.6546 ± .0009
MIMIC-III	Blind	0.6251 ± .0005	0.5305 ± .0013	0.5935 ± .0018	0.6591 ± .0020
MIMIC-IV	Conditional	0.5687 ± .0010	0.5039 ± .0004	0.5577 ± .0005	0.6052 ± .0005
MIMIC-IV	Blind	0.5780 ± .0021	0.5078 ± .0006	0.5669 ± .0006	0.6233 ± .0012

physiological factors, matching the number and relation types of the corresponding physiological structure. For RRT, randomized action-system relations retain the RRT action source and sample alternative system targets, avoiding the physiological K/M targets when sufficient alternatives are available. All other adapter components, predictive supervision, and optimization settings are matched to PatientWorld Adapt within the same task. The randomized structures are not selected based on validation or test performance.

G Evaluation Metrics and Offline Policy Evaluation

G.1 Forecasting and Adaptation Metrics

Let $\hat{y}_{i,h,d}$ and $y_{i,h,d}$ denote the normalized prediction and target for example i, forecast step h, and variable d, and let $m_{i,h,d} \in \{0,1\}$ indicate a valid observed target. Trajectory forecasting uses pooled normalized RMSE and MAE over forecast steps $h = 1, \ldots, H$:

$$\text{RMSE}_H = \sqrt{\frac{\sum_{i,h\le H,d} m_{i,h,d}(\hat{y}_{i,h,d} - y_{i,h,d})^2}{\sum_{i,h\le H,d} m_{i,h,d}}}, \qquad \text{MAE}_H = \frac{\sum_{i,h\le H,d} m_{i,h,d}|\hat{y}_{i,h,d} - y_{i,h,d}|}{\sum_{i,h\le H,d} m_{i,h,d}}. \tag{19}$$

Thus, H3 and H6 pool all observed targets over forecast steps 1:3 and 1:6, respectively. Here, $m_{i,h,d}$ combines the future observation mask with the valid-horizon mask, so each observed variable–time cell contributes once.

Downstream *Global RMSE@H* uses the same pooled normalized squared error over all valid examples, variables, and forecast steps $1{:}H$, for $H \in \{1, 3, 6\}$. *Macro RMSE@H6* first computes normalized RMSE separately for each task variable over the six-step horizon and then averages the variable-level RMSEs equally:

$$\text{MacroRMSE@H6} = \frac{1}{|\mathcal{V}_{\text{task}}|} \sum_{d \in \mathcal{V}_{\text{task}}} \sqrt{\frac{\sum_{i,h\le 6} m_{i,h,d}\,(\hat{y}_{i,h,d} - y_{i,h,d})^2}{\sum_{i,h\le 6} m_{i,h,d}}}. \tag{20}$$

Table 15: **Action/No-action subgroup forecasting on MIMIC-III.** Global RMSE is reported separately for forecasting windows containing at least one future treatment event (Act.) and those containing none (No-act.) under conditional and blind forecasting. Values are mean ± standard deviation over three seeds; lower is better. Light rose and blue denote the best and second-best means.

	Global H3				Global H6			
	Conditional		Blind		Conditional		Blind	
Model	Act.	No-act.	Act.	No-act.	Act.	No-act.	Act.	No-act.
GRU	0.6576±.0029	0.5936±.0004	0.7158±.0155	0.6675±.0246	0.6805±.0019	0.6359±.0066	0.7210±.0155	0.6998±.0147
Transformer	**0.6497**±.0030	**0.5861**±.0021	0.6644±.0047	0.6016±.0036	0.6700±.0025	0.6302±.0039	0.6977±.0115	0.6616±.0050
DLinear	0.6771±.0040	0.6026±.0028	0.7033±.0235	0.6501±.0410	0.6922±.0018	0.6365±.0015	0.7101±.0118	0.6657±.0289
PatchTST	0.6587±.0015	0.5948±.0030	0.7318±.0297	0.7002±.0393	0.6805±.0026	0.6352±.0068	0.7225±.0177	0.7026±.0280
iTransformer	0.6571±.0008	0.5915±.0018	0.7069±.0375	0.6484±.0446	0.6795±.0007	0.6424±.0028	0.7450±.0282	0.7086±.0270
MOMENT	0.7167±.0009	0.6671±.0017	0.7307±.0028	0.6846±.0041	0.7298±.0009	0.6991±.0011	0.7337±.0036	0.7022±.0024
Moirai 2.0	0.6605±.0002	0.5922±.0010	0.6684±.0031	0.6046±.0076	0.6795±.0005	**0.6271**±.0011	0.6923±.0033	0.6464±.0099
TFT-multi	0.6564±.0009	0.5897±.0011	0.6683±.0047	0.6071±.0086	0.6783±.0013	0.6364±.0024	0.6971±.0001	0.6646±.0157
TDSTF	0.9192±.0091	0.8787±.0088	0.9192±.0091	0.8787±.0088	0.9104±.0031	0.8868±.0017	0.9104±.0031	0.8868±.0017
Phys-JEPA	0.6580±.0027	0.5923±.0030	1.1198±.0748	1.0978±.0378	0.6752±.0019	0.6355±.0029	0.9608±.0647	0.9576±.0683
MedDreamer	0.6718±.0019	0.6048±.0025	0.7005±.0035	0.6429±.0129	0.6865±.0012	0.6348±.0020	0.7293±.0067	0.6983±.0144
PatientWorld	0.6499±.0015	0.5863±.0011	**0.6527**±.0011	**0.5863**±.0010	**0.6666**±.0003	0.6303±.0028	**0.6731**±.0017	**0.6304**±.0027

Table 16: **Action/No-action subgroup forecasting on MIMIC-IV.** Global RMSE is reported separately for forecasting windows containing at least one future treatment event (Act.) and those containing none (No-act.) under conditional and blind forecasting. Values are mean ± standard deviation over three seeds; lower is better. Light rose and blue denote the best and second-best means.

	Global H3				Global H6			
	Conditional		Blind		Conditional		Blind	
Model	Act.	No-act.	Act.	No-act.	Act.	No-act.	Act.	No-act.
GRU	0.5851±.0004	0.5310±.0004	0.6517±.0048	0.6022±.0068	0.6223±.0009	0.5503±.0004	0.6951±.0072	0.6479±.0256
Transformer	**0.5735**±.0006	0.5264±.0016	0.6248±.0069	0.5939±.0152	**0.6038**±.0006	0.5385±.0010	0.6786±.0100	0.6357±.0108
DLinear	0.6116±.0001	0.5577±.0006	0.7152±.0176	0.7252±.0297	0.6455±.0001	0.5655±.0003	0.7525±.0332	0.7536±.0559
PatchTST	0.5859±.0007	0.5333±.0008	0.6418±.0131	0.6112±.0236	0.6238±.0007	0.5512±.0021	0.6868±.0122	0.6435±.0228
iTransformer	0.5847± < .0001	0.5325±.0012	0.6045±.0075	0.5433±.0065	0.6288±.0007	0.5432±.0005	0.6923±.0256	0.6447±.0390
MOMENT	0.6488±.0014	0.5912±.0009	0.6541±.0036	0.5908±.0023	0.6828±.0014	0.6014±.0018	0.6910±.0043	0.6190±.0079
Moirai 2.0	0.5909±.0069	0.5418±.0057	0.6130±.0043	0.5749±.0107	0.6225±.0003	0.5510±.0003	0.6689±.0093	0.6489±.0353
TFT-multi	0.5875±.0016	0.5302±.0020	0.6272±.0052	0.5797±.0219	0.6204± < .0001	0.5424±.0021	0.6749±.0019	0.5880±.0160
TDSTF	0.8476±.0070	0.7352±.0090	0.8476±.0070	0.7352±.0090	0.8523±.0092	0.7286±.0050	0.8523±.0092	0.7286±.0050
Phys-JEPA	0.5847±.0018	0.5346±.0018	1.1854±.1412	1.2062±.1891	0.6194±.0020	0.5502±.0030	1.1020±.1658	1.1015±.2502
MedDreamer	0.5994±.0005	0.5414±.0006	0.6527±.0124	0.5835±.0107	0.6318±.0008	0.5514±.0009	0.7080±.0040	0.6102±.0077
PatientWorld	0.5744±.0010	**0.5208**±.0062	**0.5834**±.0020	**0.5250**±.0030	0.6088±.0005	**0.5354**±.0011	**0.6276**±.0014	**0.5385**±.0026

The task-variable set is task-specific. For sepsis, it includes `MeanBP`, `Lactate`, `Creatinine`, `UrineOutput`, and `ShockIndex`, intersected with the variables available in the corresponding task vocabulary. For RRT initiation timing, it is the renal/metabolic KM set: `Creatinine`, `BUN`, `UrineOutput`, `Potassium`, `Bicarbonate`, and `pH`.

For the Action/No-action analysis in Sec. 5.4, a forecasting window is assigned to the *Action* subset if at least one treatment/action event occurs anywhere in its valid forecast horizon. A valid forecasting window with no future action event in the horizon is assigned to the *No-action* subset. This partition therefore depends on future treatment occurrence within the forecast horizon rather than on the current treatment alone.

G.2 Additional Forecasting Analyses

To complement the relative comparison in Fig. 4(a), Tables 15 and 16 report absolute Global RMSE for Action and No-action subsets across all trajectory forecasting methods. Subset definitions follow Appendix G.1. Across both datasets and forecasting horizons, PatientWorld achieves the lowest blind RMSE in all eight Action/No-action subgroup settings, while conditional performance remains closer to the strongest baselines.

G.3 Action-Response Diagnostic

The action-response analysis in Sec. 5.4 measures whether the learned dynamics produce distinguishable trajectories under alternative admissible actions while holding patient history fixed. It is used only as a model-sensitivity diagnostic and is not interpreted as a causal treatment-effect estimate.

For a fixed history i, let $\hat{y}^{(a)}_{i,h,d}$ denote the normalized prediction under admissible action or action schedule $a \in \mathcal{A}_i$. The primary action-response statistic is the standard deviation of predictions across admissible actions, averaged over evaluated examples, forecast steps, and variables:

$$\mathrm{AR} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{|\mathcal{H}_i||\mathcal{D}_i|} \sum_{h\in\mathcal{H}_i}\sum_{d\in\mathcal{D}_i} \mathrm{Std}_{a\in\mathcal{A}_i}\left[\hat{y}^{(a)}_{i,h,d}\right]. \tag{21}$$

The implementation additionally records the corresponding max–min action range as a secondary sensitivity diagnostic.

For sepsis, we enumerate all 25 fluid–vasopressor actions and compare one-step predictions. Fluid-only and vasopressor-only action slices are additionally recorded. For RRT, we compare seven six-step schedules: continued deferral or initiation beginning at one of steps $0, \ldots, 5$. Sensitivity is computed at H1, H3, and H6, with the reported headline magnitude taken at H6. For both tasks, the diagnostic is computed over all eligible factual forecasting rows in the held-out test set. Overall sensitivity is computed across all modeled variables; for RRT, the renal/metabolic subset contains Creatinine, BUN, UrineOutput, Potassium, Bicarbonate, and pH.

G.4 Downstream Policy Construction and Selection

Sepsis treatment. The sepsis action space contains 25 discrete actions formed by five fluid bins and five vasopressor bins. Motivated by prior sepsis RL reward designs combining physiological shaping with terminal survival outcomes (Raghu et al., 2018), we define the task-specific reward as

$$r_t = \mathrm{clip}\left(\mathrm{SOFA}_t - \mathrm{SOFA}_{t+1}, -4, 4\right) + \tanh\left(\frac{\mathrm{Lactate}_t - \mathrm{Lactate}_{t+1}}{2}\right) + r_t^{\mathrm{terminal}}, \tag{22}$$

where

$$r_t^{\mathrm{terminal}} = \begin{cases} +10, & \text{hospital survival,} \\ -10, & \text{hospital death,} \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

The terminal component is additive rather than replacing the physiological shaping terms: terminal transitions retain both the SOFA- and lactate-based components whenever they are available. If either endpoint required to compute a SOFA or lactate change is unavailable, the corresponding reward component is set to zero rather than imputed.

Dynamics-based sepsis policies use a common six-step planning horizon, corresponding to 24 hours. Candidate action sequences are generated using the same CQL-supported candidate mechanism and are scored using the frozen world model together with a common planner-value model. The highest-scoring sequence is selected and its first action is executed, yielding a receding-horizon policy. All dynamics-based sepsis methods therefore use the same action space, reward, candidate-generation procedure, planning horizon, and planner.

RRT initiation timing. The formal RRT action space is binary: Defer or Initiate within the next 24 hours. Planning uses three-stage backward induction over the 0-, 24-, and 48-hour decision points. Initiation branches are represented by six-step within-24-hour start schedules. At each decision point, the planner compares the estimated values Q_D and Q_I and selects Initiate when $Q_I \geq Q_D$; otherwise it selects Defer. Sixty-day hospital-free days (HFD60) is the planning utility and primary downstream endpoint, with higher values preferred. All dynamics-based RRT methods use the same binary action definition, endpoint, and backward-induction planner.

Validation and held-out evaluation. Model, policy, and evaluator selection is performed using validation data, and held-out test data are used only for final evaluation. For sepsis CQL, checkpoint and hyperparameter selection uses the validation FQE initial-value estimate. For fixed world-model policies, the fitted FQE evaluator checkpoint is selected by validation Bellman loss. For RRT FQE, the ridge penalty is selected by validation RMSE, while the model-free RRT CQL baseline is selected by validation CQL loss.

G.5 Offline Policy Evaluation

Sepsis Reward-FQE. For a fixed target policy π, Reward-FQE fits a Q-function using Bellman targets

$$y_t^{\mathrm{FQE}} = r_t + \gamma(1 - d_t)Q_{\mathrm{targ}}\left(s_{t+1}, \pi(s_{t+1})\right), \qquad \gamma = 0.99, \tag{24}$$

where d_t indicates a terminal transition. States use the normalized replay representation. World-model policies supply deterministic target actions, while CQL uses its greedy action. The reported Reward-FQE estimate is the mean fitted value at the first transition of each test episode.

Estimated mortality and behavior-policy support. Estimated hospital mortality is computed with a separate probability-constrained FQE. Its immediate reward is 1 only on terminal death and 0 otherwise, with $\gamma = 1$; a sigmoid output constrains the fitted value to $[0, 1]$. The reported estimate is the mean mortality probability at the initial transition of each episode.

To characterize logged-data support, we fit a 25-action behavior-policy classifier on the training replay. For state s, the support of the target-policy decision is

$$\mathrm{Supp}(s) = \hat{\mu}\left(\pi(s) \mid s\right), \tag{25}$$

where $\hat{\mu}$ is the learned behavior policy. *Mean Support* averages this probability over evaluated target-policy decisions, while *Low Support* is the proportion of selected actions satisfying

$$\hat{\mu}\left(\pi(s) \mid s\right) < 0.01. \tag{26}$$

Sepsis uses the formal test replay directly; no additional OPE-specific cohort is constructed, and initial-value reporting uses the first transition of each episode.

RRT HFD60 FQE. RRT HFD60 FQE uses stage-wise linear ridge Q-functions over the normalized state, decision-stage indicator, and binary action indicator. The estimator is fit backward over the 48-, 24-, and 0-hour decision stages. The target is the observed HFD60 endpoint after initiation, at the final decision stage, or at a clinical terminal transition; otherwise, it bootstraps from the next-stage value under the target policy. The reported HFD60 FQE estimate is the mean fitted value at stage 0. The same stage-wise construction is used for the hospital-outcome evaluator with hospital survival as the endpoint; Table 3 reports the corresponding hospital mortality percentage.

RRT WIS and effective sample size. Let $\hat{\mu}(a_t \mid s_t)$ denote the binary behavior policy fitted on the training data and $\pi(a_t \mid s_t)$ the target-policy probability. For episode i, the trajectory importance weight is

$$w_i = \prod_{t \in \mathcal{T}_i} \frac{\pi(a_{i,t} \mid s_{i,t})}{\hat{\mu}(a_{i,t} \mid s_{i,t})}. \tag{27}$$

The weighted HFD60 estimate is

$$\widehat{V}_{\mathrm{WIS}} = \frac{\sum_{i=1}^{N} w_i G_i}{\sum_{i=1}^{N} w_i}, \tag{28}$$

where G_i is the observed HFD60 outcome. The target policy is deterministic, so a trajectory containing a decision inconsistent with the target policy receives zero target-policy probability. We report the corresponding effective sample size

$$\mathrm{ESS} = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}. \tag{29}$$

When reported as a percentage, ESS is divided by the number of episodes entering the WIS estimate.

Table 17: **Complementary weighted doubly robust (WDR) offline policy evaluation.** Results are mean ± standard deviation over three seeds; higher is better. WDR is reported as a complementary OPE estimator and is not used as the primary model-selection metric.

Method	Sepsis WDR	RRT WDR
CQL	4.605 ± 3.008	36.94 ± 0.08
MedDreamer	11.625 ± 1.060	37.88 ± 0.15
Transformer	9.838 ± 6.510	37.85 ± 1.31
Moirai 2.0	12.192 ± 8.386	36.96 ± 1.28
TFT-multi	$\mathbf{19.158 \pm 3.222}$	36.35 ± 1.16
Phys-JEPA	10.437 ± 4.494	35.59 ± 1.36
PatientWorld	$\underline{15.319 \pm 0.210}$	$\mathbf{38.50 \pm 0.11}$

RRT OPE cohort. The formal downstream RRT test split contains 3,777 episodes from 2,995 patients and 8,007 decision rows, comprising 3,711 stage-0 rows, 2,585 24-hour rows, and 1,711 48-hour rows. OPE is anchored at the initial stage-0 decision point. Accordingly, the 3,711 episodes containing a valid stage-0 row form the OPE cohort, while the remaining 66 formal test episodes are excluded because they lack an initial stage-0 decision row. HFD60 is fully observed in the frozen formal test cohort and therefore introduces no additional episode-level exclusion. All RRT OPE estimators operate on this same episode-ID set, and no additional clinical eligibility criteria are reapplied during OPE.

G.6 Weighted Doubly Robust Evaluation

As a complementary offline policy-evaluation estimator, we additionally report weighted doubly robust (WDR) estimates for the downstream policies. WDR combines importance weighting with fitted value estimates and is used only for evaluation; it is not used for model, policy, or checkpoint selection. Within each task, all methods are evaluated using the same replay splits, behavior-support estimation procedure, WDR formulation, and held-out test protocol. The fitted value-function nuisance model is policy-specific where required and is trained or selected without using test outcomes.

Sepsis. The behavior policy is modeled by a 25-action neural classifier, $\pi_b(a \mid s)$, trained on the TRAIN replay split by cross-entropy against the logged action identifiers. The checkpoint with the lowest VALID cross-entropy is retained, and the resulting behavior model is used to estimate the logged-action probabilities entering importance weighting. TEST data are used only for final reporting.

The value-function nuisance model is the same neural FQE architecture used in the primary sepsis OPE. For a target policy π, FQE is trained using Bellman targets

$$r_t + \gamma(1 - d_t)Q_{\text{target}}(s_{t+1}, \pi(s_{t+1})),$$

with optimization on TRAIN and evaluator selection by VALID Bellman loss. For world-model planning policies, a separate FQE nuisance model is fitted for each fixed target policy under this common training and selection procedure; the greedy CQL policy is evaluated through the same behavior-policy and FQE machinery.

Sepsis WDR starts from the estimated initial FQE value and adds per-time-step weighted Bellman-residual corrections. At each time step, the residual contribution is normalized by the sum of the nonzero importance weights. Thus, all sepsis policies share the same frozen replay splits, train-fitted behavior-policy estimator, and WDR/FQE implementation, while the FQE nuisance model is fitted under the corresponding target policy.

RRT initiation timing. For RRT, the behavior policy is a binary logistic-regression model for the probability of INITIATE versus DEFER. It is fitted using TRAIN rows only, with features consisting of the policy state and stage one-hot representation, without class weighting. The corresponding factual-action propensity is used in the importance ratios.

The value-function nuisance model consists of stage-wise linear ridge Q-functions over the policy state, stage one-hot encoding, and action one-hot encoding. These models are fitted backward over

the 48-h, 24-h, and 0-h decision stages using TRAIN data only. When a ridge-penalty grid is used, the regularization coefficient is selected by VALID RMSE; TEST data are reserved for final evaluation. The same RRT OPE implementation is used for PatientWorld-planning policies and the RRT CQL baseline.

RRT WDR combines the mean estimated initial value with stage-wise weighted Bellman-residual corrections. Weights are formed from cumulative target-to-behavior propensity ratios, together with the same propensity floor and importance-weight clipping used by the task-specific OPE implementation. The same evaluator is applied to the available RRT endpoints, including HFD60, hospital survival, 60-day survival, and 90-day survival; HFD60 is the primary endpoint reported in the main results.

For both tasks, behavior propensities are estimated only from the training split and nuisance value models are fitted without test outcomes. All reported methods within a task are evaluated through the same task-specific WDR/OPE implementation and held-out test split; target policies may use separately fitted FQE nuisance models under the same evaluator protocol.