GRAPH WORLD MODELS FOR CONSTRAINED EPIDEMIC POLICY PLANNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Epidemic policy planning often requires coordination between geographical regions, taking into account mobility-driven spillovers and how to make use of limited resources. Existing methods either lack action-conditioned models of coupled dynamics or cannot guarantee per-period feasibility. We present EpiMind, a graph world model framework for constrained epidemic policy planning across regions. A graph-factored recurrent state-space model generates joint policy-conditioned rollouts from regional latent beliefs, while graph-temporal ADMM optimizes regional interventions, enforces shared-resource feasibility through projection, and evaluates temporal specifications under the learned model. EpiMind reduces admission RMSE by 29% relative to graph-free dynamics modeling, plans within 1–5% of the best feasible constant policy with guaranteed shared-budget feasibility, and outperforms all deployable baselines across three resource budgets in real-context evaluation. These results demonstrate that graph-structured policy imagination with explicit constrained coordination supports effective epidemic interventions from learned dynamics.

1 INTRODUCTION

The effective control of epidemics requires coordinating non-pharmaceutical, pharmaceutical, and surveillance interventions across multiple regions Kraemer et al. (2020); Ferretti et al. (2020); Hsiang et al. (2020). The COVID-19 epidemic demonstrated how a patchwork of responses created uncertainty, chaos, and ultimately lack of trust in public health authorities Birkland et al. (2021); SteelFisher et al. (2023). Jurisdictions act on observable local conditions but remain coupled through mobility and competition for finite vaccines, hospital capacity, and budgets Emanuel et al. (2020).

Coordinating policy making across regions is difficult, especially under uncertainty and shared resource constraints Kermack & McKendrick (1927); Balcan et al. (2010). A given surveillance trend may reflect transmission changes, altered testing, or voluntary behavioral adaptation across regions. A useful planning model must therefore infer latent epidemic conditions from noisy, delayed, policy-dependent observations and predict their graph-coupled evolution under alternative actions.

Existing methods address only a fragment of this problem. Compartmental models Kermack & McKendrick (1927) and their metapopulation extensions Arino & van den Driessche (2003); Balcan et al. (2010) simulate forward trajectories by modifying mechanistic parameters such as transmission rates or contact matrices. They treat policies as exogenous inputs rather than decision variables. Further, they assume reported cases are direct measurements of true incidence rather than policy-dependent surveillance signals. Graph neural network (GNN)-based spatiotemporal forecasters such as Cola-GNN Deng et al. (2020) and county-level COVID-19 predictors Kapoor et al. (2020) learn flexible dynamics from time series but have no action space; thus, they cannot distinguish whether a forecasted decline reflects an intervention's effect or a confounder such as reduced testing. Reinforcement learning (RL) for epidemic control Kompella et al. (2020); Ohi et al. (2020); Bushaj et al. (2023) reframes the problem as sequential decision-making within a calibrated simulator but typically optimizes a single composite agent over a concatenated multi-region state, with resource limits absorbed into shaped rewards that provide no feasibility guarantee for hard joint constraints such as total vaccine supply summed across regions. Multi-agent RL approaches that extend MAPPO Yu et al. (2022) to regional epidemic control Nayak et al. (2023) train per-region policies under a shared critic, but cross-agent sum constraints are typically incorporated as reward penalties

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.

or adaptive Lagrange multipliers, which guarantee only expected feasibility at convergence rather than per-timestep feasibility for a planner facing fixed inventory, and require a high-fidelity simulator that is unavailable for active outbreak response. Mathematical programming methods such as mixed-integer optimization for vaccination facility location Bertsimas et al. (2022) and vaccine supply-chain optimization Duijzer et al. (2018) enforce hard constraints exactly through branch-and-cut, but require the transmission dynamics (case trajectory, susceptible fraction, reproduction number) to be supplied as a pre-fit input from a separately calibrated SEIR model, so the optimizer cannot adapt as decisions, behavioral response, or variant emergence shifts the trajectory.

World models Hafner et al. (2023); Wang et al. (2025); Memon et al. (2026) address several of these limitations by learning latent transition and observation models that can be rolled forward under candidate action sequences without further interaction with the environment. Conditioning the dynamics on actions allows the model to represent policy-dependent evolution and surveillance. It does not, however, identify causal intervention effects under endogenous historical policies, where regions may adopt stronger interventions precisely when outbreaks worsen. Graph world models (GWMs) Feng et al. (2025) extend recurrent state-space models (RSSMs) Hafner et al. (2023) by representing interacting entities as nodes that exchange information through message passing, providing an appropriate inductive bias for mobility-coupled epidemics. Existing world-model methods nevertheless provide limited machinery for coordinating distinct regional actions under per-period shared-resource constraints, and this is a gap we propose to address. We introduce EpiMind, a graph world-model planning framework whose key contributions are:

- EpiMind helps formulate multi-region epidemic planning on a dynamic policy graph, coupling regional dynamics through mobility and finite shared resources.
- EpiMind couples a parameter-shared GWM with a constrained multi-region planner for joint policy rollout, resource-feasible allocation, and temporal-logic evaluation.
- EpiMind achieves near-oracle performance in real-context evaluation, reduces admissions by 53.4% relative to no intervention while satisfying all shared-resource constraints.

2 METHOD

2.1 PROBLEM FORMULATION

Epidemics on graphs. We formalize the regional epidemic by a time-varying policy graph $G_t = (V_t, E_t)$ (see details in Section A), where nodes V_t represent regions and edges E_t capture mobility and coordination constraints at time t. This formulation makes the planning problem tractable by aligning each computational ingredient with the epidemic's physical structure.

Graph-structured latent dynamics. Let x_t^i denote region i's unobserved epidemic state, and $\mathbf{x}_t = (x_t^1, \ldots, x_t^N)$ denote the joint state. Each region selects a D-dimensional intervention vector:

$$a_t^i \in \mathcal{A}^i \subseteq [0,1]^D, \qquad \mathbf{a}_t = (a_t^1, \ldots, a_t^N), \tag{1}$$

The action comprises local non-pharmaceutical intervention (NPI) intensity and regional allocations of shared vaccine, hospitalization capacity, and fiscal-resource budget. The joint transition is policy-conditioned and graph-coupled:

$$p_\theta(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{a}_t, G_t) = \prod_{i=1}^{N} p_\theta\Big(x_{t+1}^i \,\Big|\, x_t^i, a_t^i, \mathrm{Agg}_{j \in \mathcal{N}_t(i)}\Big(E_t^{ij}, x_t^j, a_t^j\Big)\Big). \tag{2}$$

Here $\mathcal{N}_t(i)$ is region i's mobility neighborhood, and θ is shared across regions. Regional states, actions, observations, and neighborhoods remain distinct. The latent state is not observed directly. Instead, region i receives a surveillance observation, e.g., infections, hospital admissions, or deaths.

$$o_t^i \sim \Omega_\theta^i\big(\cdot \mid x_t^i, \mathbf{a}_{t-1}\big), \tag{3}$$

Candidate policies are evaluated over a multi-step horizon rather than through one-step prediction alone to response to the delayed surveillance and intervention effects.

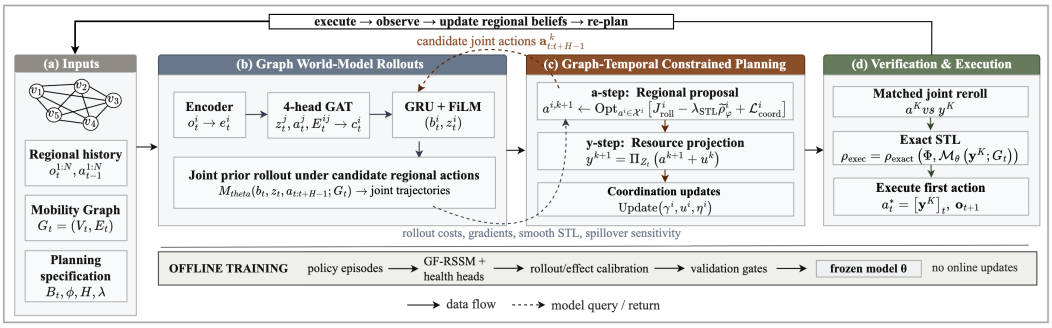


Figure 1: EpiMind framework. A parameter-shared GF-RSSM updates regional beliefs and generates graph-coupled policy rollouts. GT-ADMM coordinates and projects regional actions, rerolls the feasible allocation for model-relative STL evaluation, and executes its first action.

Constrained policy planning. At decision epoch t, the planner evaluates a candidate sequence of joint regional actions $\mathbf{a}_{t:t+H-1}$ over horizon H. Starting from the current regional beliefs, a learned dynamics model M_θ generates the joint policy-conditioned rollout:

$$\hat{\tau}^{1:N}_{t+1:t+H} = M_\theta\left(b_t^{1:N}, z_t^{1:N}, \mathbf{a}_{t:t+H-1}, G_t\right). \tag{4}$$

These rollouts support model-dependent comparisons among candidate policies and, in simulation, are validated against known counterfactual outcomes. The planner minimizes predicted health and intervention costs subject to shared resource budgets:

$$\begin{aligned} \min_{\mathbf{a}_{t:t+H-1}} \quad & \sum_{\tau=t}^{t+H-1} \ell\left(\hat{\tau}^{1:N}_\tau, \mathbf{a}_\tau\right) - \beta\,\widetilde{\rho}_\varphi\left(\hat{\tau}^{1:N}_{t+1:t+H}\right) \\ \text{s.t.} \quad & \sum_{i=1}^{N} a_\tau^{i,d} \le B_\tau^d, \qquad d \in \mathcal{D}_{\mathrm{shared}}, \\ & 0 \le a_\tau^{i,d} \le 1, \qquad i = 1, \dots, N, \quad \tau = t, \dots, t+H-1, \end{aligned} \tag{5}$$

where ℓ balances predicted epidemic burden and intervention cost, $\widetilde{\rho}_\varphi$ is a differentiable robustness score for temporal specification φ, and $\mathcal{D}_{\mathrm{shared}}$ indexes the resource-constrained action dimensions. Projection guarantees that the executed action satisfies the specified resource budgets, and temporal constraints are evaluated relative to the learned model by rerolling the projected action.

2.2 EpiMind Framework

EpiMind couples a learned GWM with constrained receding-horizon planning as shown in Figure 1. The world model predicts joint epidemic trajectories under candidate regional interventions, and the planner coordinates the interventions subject to the shared resource constraints.

2.2.1 Graph World-Model Rollouts

Graph-factored recurrent state-space model (GF-RSSM). We instantiate the dynamics in Section 2.1 with a graph-factored recurrent state-space model (GF-RSSM). Neural parameters are shared across regions, while each region i maintains its own recurrent belief b_t^i and stochastic state z_t^i. Neighboring states and actions are aggregated through graph attention:

$$c_t^i = \mathrm{GAT}_\theta\left(\{z_t^j, a_t^j, E_t^{ij}\}_{j \in \mathcal{N}(i)}\right), \tag{6}$$

$$b_t^i = f_\theta\left(b_{t-1}^i, z_{t-1}^i, a_{t-1}^i, c_{t-1}^i\right), \tag{7}$$

$$z_t^i \sim q_\theta\left(z_t^i \mid b_t^i, e_\theta(o_t^i)\right), \qquad \hat{o}_t^i \sim p_\theta\left(o_t^i \mid b_t^i, z_t^i\right). \tag{8}$$

Joint policy-conditioned rollout. The graph context c_t^i transmits mobility-weighted information from neighboring regions. Parameter sharing provides a common transition model without imposing identical regional trajectories: beliefs, latent states, observations, actions, and neighborhoods remain region specific. The posterior in (8) assimilates the current observation. Future observations are unavailable during planning, so imagined trajectories use the learned prior recursively. At every rollout step, all regional states advance jointly under the complete action vector and mobility graph:

$$\hat{\mathbf{x}}_{t+1:t+H} = \mathcal{M}_\theta(\mathbf{b}_t, \mathbf{z}_t, \mathbf{a}_{t:t+H-1}; G_{t:t+H-1}) . \tag{9}$$

Thus, each trajectory depends on both local and neighboring interventions. Dedicated admission and occupancy heads decode the health quantities used by the planner.

World-model training. The GF-RSSM is trained offline for predictive and policy-effect fidelity:

$$\mathcal{L}_{\mathrm{WM}} = \mathcal{L}_{\mathrm{obs}} + \mathcal{L}_{\mathrm{reward}} + \mathcal{L}_{\mathrm{KL}} + \mathcal{L}_{\mathrm{rollout}} + \mathcal{L}_{\mathrm{effect}} + \mathcal{L}_{\mathrm{cal}}, \tag{10}$$

where the first three terms form the recurrent state-space objective; the rollout, effect, and calibration terms supervise delayed policy responses. The trained world model is frozen during planning.

2.2.2 Graph-Temporal Constrained Planning

Graph-temporal alternating direction method of multipliers (GT-ADMM). We implement the constrained planner as graph-temporal alternating direction method of multipliers (GT-ADMM), which alternates regional action optimization, shared-resource projection, and coordination updates. At decision epoch t, GT-ADMM optimizes region-specific action sequences over horizon H:

$$\begin{aligned} \min_{\mathbf{a}} \quad & J_\theta(\mathbf{a}) = \sum_{i=1}^{N} f^i\big(\hat{\mathbf{x}}_{t+1:t+H}, \mathbf{a}^i\big) - \lambda_{\mathrm{STL}} \sum_{i=1}^{N} \rho_{\mathrm{sm}}\big(\Phi^i, \hat{\mathbf{x}}^i_{t+1:t+H}\big) , \\ & \hat{\mathbf{x}}_{t+1:t+H} = \mathcal{M}_\theta(\mathbf{a}; G_t), \end{aligned} \tag{11}$$

subject to local action bounds, edge relations, and the shared resource set $\mathcal{Z}_t$. Here f^i is the per-region health–intervention cost:

$$f^i = \sum_{h=1}^{H} \left[\ell^i_{\mathrm{health},t+h} + \lambda_{\mathrm{NPI}} \ell^i_{\mathrm{NPI},t+h} \right] , \tag{12}$$

where λ_{NPI} controls the trade-off between predicted health burden and NPI burden. The coefficient λ_{STL} weights the smooth robustness ρ_{sm} of temporal specification Φ^i. GT-ADMM alternates regional proposal updates, shared-resource projection, and coordination updates.

Regional proposal update. Each region updates its action block while holding the other regions at their latest reference actions:

$$\begin{aligned} a^{i,k+1} = \arg\min_{a^i \in \mathcal{X}^i} \Bigg[& \mathcal{J}^i_{\mathrm{roll}} + \frac{\rho_e}{2} \sum_{j \in \mathcal{N}(i)} \|a^i - a^{j,k}\|^2 + (\gamma^{i,k})^\top a^i \\ & + \frac{\rho_g}{2} \|a^i - y^{i,k} + u^{i,k}\|^2 + (\eta^{i,k})^\top s^i + \frac{\sigma}{2} \|a^i - a^{i,k}\|^2 \Bigg], \end{aligned} \tag{13}$$

where

$$\mathcal{J}^i_{\mathrm{roll}} = f^i\big(a^i, \mathcal{M}_\theta(a^i, \mathbf{a}^{-i,k}; G_t)\big) - \beta\, \rho_{\mathrm{sm}}\big(\Phi^i, \mathcal{M}_\theta(a^i, \mathbf{a}^{-i,k}; G_t)\big) . \tag{14}$$

The rollout remains joint so that changing a^i can alter the predicted outcomes of every connected region. The remaining terms encourage neighboring-policy agreement, consistency with the feasible allocation, spillover awareness, and stable successive updates. Equation (13) is solved by gradient descent through the frozen world model. The spillover signal is

$$s^i = \sum_{j \in \mathcal{N}(i)} E^{ij}_t \nabla_{a^i} g^j_{\mathrm{spill}}(x_t, a^i), \tag{15}$$

where g_{spill} predicts neighboring outcome changes. This signal is predictive, not causally identified.

Resource projection. Regional proposals need not be jointly feasible. GT-ADMM therefore computes the nearest allocation in the shared resource set:

$$y^{k+1} = \arg\min_{z \in \mathcal{Z}_t} \sum_{i=1}^{N} \frac{\rho_g}{2} \left\| a^{i,k+1} - z^i + u^{i,k} \right\|^2 . \tag{16}$$

Because $\mathcal{Z}_t$ contains linear box and sum constraints, the z-step decomposes by resource into capped-simplex projections. The planner executes y^{k+1}, guaranteeing satisfaction of these budgets.

Coordination updates. After projection, the coordination variables are updated as

$$\gamma^{i,k+1} = \gamma^{i,k} + \rho_e \sum_{j \in \mathcal{N}(i)} \left(a^{i,k+1} - a^{j,k+1} \right), \tag{17}$$

$$u^{i,k+1} = u^{i,k} + a^{i,k+1} - y^{i,k+1}, \tag{18}$$

$$\eta^{i,k+1} = \eta^{i,k} + \alpha s^i, \tag{19}$$

where u^i reflects pressure from shared-resource scarcity, γ^i tracks disagreement with mobility-connected regions, and η^i tracks predicted cross-region spillover sensitivity. The node-level γ^i and η^i are approximate accumulators for neighbor disagreement and predicted spillover sensitivity, respectively. These approximations provide interpretable coordination signals.

2.2.3 Projection, Verification, and Execution

After the final iteration, EpiMind retains the unprojected proposal a^K and executes the projected allocation y^K. Both are rerolled through the same joint world-model interface. Smooth STL robustness is used to obtain gradients during optimization, whereas exact nonsmooth robustness is evaluated on the projected trajectory:

$$\rho_{\text{exec}} = \rho_{\text{exact}} \left(\Phi, \mathcal{M}_\theta(y^K; G_t) \right). \tag{20}$$

Resource feasibility is guaranteed for the constraints represented in $\mathcal{Z}_t$. In contrast, STL satisfaction is model relative, i.e., a positive ρ_{exec} certifies the learned rollout, not the unknown true environment. The first action of y^K is executed, and the resulting joint observation updates the regional posterior at decision epoch $t+1$. This produces the closed-loop sequence: *belief update* $\rightarrow$ *joint rollout* $\rightarrow$ *constrained planning* $\rightarrow$ *projection and verification* $\rightarrow$ *execution*.

3 Experiments

We evaluate EpiMind through five research questions spanning predictive fidelity, planning utility, constraint handling, coordination, and real-context transfer:

- **RQ1 Predictive fidelity.** Does the graph world model accurately predict held-out trajectories and policy responses? (Section 3.2.1, Appendix D.1)
- **RQ2 Planning effectiveness.** Does planning through learned joint rollouts improve the matched health–intervention objective? (Section 3.2.2, Appendix D.2)
- **RQ3 Feasibility and verification.** Does projection enforce shared budgets, and do projected rollouts satisfy model-relative temporal specifications? (Section 3.2.3, Appendix D.3)
- **RQ4 Graph coordination.** Does mobility-aware coordination improve allocation under heterogeneous conditions and shared resources (Section 3.2.4, Appendix D.4)
- **RQ5 Real-context transfer.** Does EpiMind support forecasting and allocation in realistic multi-region settings? (Section 3.2.5, Appendices D.5–D.6)

3.1 Experimental Protocol

Evaluation tracks. We evaluate EpiMind in three complementary settings: (1) a mobility-coupled multi-region simulator provides known dynamics and ground-truth outcomes for evaluating policy-conditioned prediction and planning; (2) a retrospective U.S. state-level panel evaluates forecasting and allocation behavior under observed surveillance, intervention, capacity, and mobility data; and (3) a real-context semi-simulated setting which initializes the simulator from real data while retaining known dynamics for realized evaluation of alternative policies.

Table 1: World-model fidelity in the synthetic environment. Values are mean ± s.d. Admission errors are measured per 100K; cumulative error is over a five-step rollout. Lower is better.

Model	Adm. MAE ↓	Adm. RMSE ↓	Cum. err. @5 ↓	Params
Statistical baselines				
Persistence	1.563 ± 0.053	1.927 ± 0.056	0.089 ± 0.003	0
Climatology	1.129 ± 0.021	1.283 ± 0.027	0.528 ± 0.011	0
Ridge + action	0.341 ± 0.008	0.437 ± 0.003	0.068 ± 0.006	84
VARX(1) + action	0.329 ± 0.008	0.424 ± 0.001	0.066 ± 0.007	1,960
Learned dynamics models				
Action-LSTM	0.212 ± 0.044	0.310 ± 0.062	0.056 ± 0.007	25,507
GF-RSSM (no graph)	0.211 ± 0.033	0.311 ± 0.046	0.070 ± 0.012	92,972
GF-RSSM (ours)	$\mathbf{0.156 \pm 0.049}$	$\mathbf{0.221 \pm 0.070}$	$\mathbf{0.036 \pm 0.003}$	92,972

Datasets. The synthetic benchmark contains $N{=}5$ mobility-coupled regions over $T{=}26$ weekly decision epochs. Regional actions control NPI intensity and allocations of vaccine, hospitalization-capacity, and fiscal resources under shared budgets. The planner observes delayed, noisy surveillance signals rather than latent SEIR states; Table A2 reports the complete simulator configuration.

For the real-context evaluation, we construct a weekly U.S. state-level panel combining reported cases, deaths, hospital admissions and capacity, vaccination, policy interventions, population, and directed interstate mobility. Track A retrospectively evaluates forecasting and model-relative policy projections at held-out decision origins; outcomes under unexecuted policies are unavailable. Track B initializes a semi-synthetic simulator from the same regional conditions and mobility graph, permitting realized evaluation of alternative policies under known dynamics. Data sources and preprocessing appear in Appendix C.1.

Benchmarking methods. World-model comparisons include statistical predictors, action-conditioned sequence models, and graph ablations. Planning comparisons include constant and heuristic policies, MPC, ADMM, RL-based controllers, graph-free and independent variants, and an oracle-dynamics reference. All policies are evaluated under the same action bounds, resource budgets, projection, and health–intervention objective. Implementation details appear in Appendix C.2.

Metrics. For RQ1, we report held-out admission MAE, RMSE, cumulative rollout error, and policy-conditioned dose response. For RQ2, we report admissions per 100K, NPI burden, matched objective J, and regret relative to the best feasible constant policy. For RQ3, we report budget feasibility, maximum excess, projection displacement, and exact model-relative STL robustness. For RQ4, we compare realized objective values and paired outcomes under shared budgets, supplemented by a stepwise matched-burden ablation. For RQ5, we report retrospective model-relative comparisons and realized outcomes in the real-context benchmark.

3.2 Experimental Results

3.2.1 GF-RSSM supports accurate policy-conditioned prediction.

Table 1 evaluates deterministic prior rollouts on held-out synthetic episodes. GF-RSSM achieves the lowest error on all three admission metrics. Relative to the no-graph ablation, the full model reduces admission MAE by 26% (0.211 to 0.156), RMSE by 29% (0.311 to 0.221), and five-step cumulative error by 49% (0.070 to 0.036). It similarly improves over Action-LSTM by 26%, 29%, and 36%, respectively. Although uncertainty over three checkpoints limits strong statistical conclusions for MAE and RMSE, the cumulative-error improvement is consistent across checkpoints.

Figure 2 examines whether this predictive accuracy extends to policy-conditioned responses. In simulation, GF-RSSM preserves the monotonic NPI dose ordering but overpredicts admissions at low NPI and underpredicts them at high NPI. On retrospective data, predicted admissions decrease with NPI across all four held-out decision origins. These real-data curves demonstrate stable model sensitivity, not causal effects, because counterfactual outcomes are unavailable.

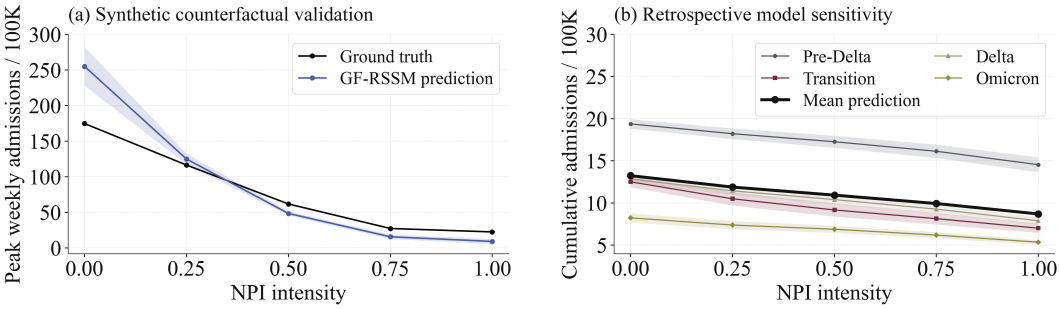


Figure 2: Policy-conditioned admission response. (a) Peak weekly admissions under alternative NPI intensities applied from a common synthetic state. (b) Predicted cumulative admissions under the same sweep at four held-out U.S. decision origins.

3.2.2 Learned rollouts yields effective resources-feasible interventions.

We evaluate all methods under the same health–intervention objective, action bounds, shared budgets, and final resource projection. Table 2 reports simulator-realized objective values across three intervention-cost regimes. The best constant-NPI policy saturates the shared budget, providing a strong non-adaptive comparator. EpiMind remains within 1.1%, 1.5%, and 4.9% of this comparator at $\lambda_{\text{NPI}} \in \{3, 10, 30\}$, respectively, with paired regret 0.93 ± 1.10 at $\lambda_{\text{NPI}} = 10$. It also consistently outperforms PPO, MPC-SEIR, independent MPC, and the remaining planning baselines. All executed allocations have zero post-projection budget excess. These results show that planning through learned joint rollouts produces effective resource-feasible interventions. The learned-versus-oracle decomposition in Appendix D.3 suggests a model contribution to the remaining regret, but the effect is not statistically resolved with four paired cells.

Table 2: Matched-objective planning performance. Values are mean ± s.d. Lower is better.

	Objective $J \downarrow$			**Regret at $\lambda = 10 \downarrow$**
Method	$\lambda = 3$	$\lambda = 10$	$\lambda = 30$	mean ± s.d.
Best feasible constant[†]	**54.7**	**58.9**	**70.9**	**0.00 ± 0.00**
PPO	60.4	64.6	76.6	5.69 ± 3.71
MPC-SEIR	108.6	112.8	124.6	53.9 ± 30.9
Independent MPC	366.3	431.1	394.1	372 ± 307
Greedy	7,186	7,189	7,197	7,130 ± 1,221
D-ADMM	9,300	9,263	9,224	9,204 ± 583
HRL	9,429	9,429	9,429	9,370 ± 478
No intervention	13,853	13,853	13,853	13,795 ± 29
EpiMind	**55.3**	**59.8**	**74.4**	**0.93 ± 1.10**

[†]Constant NPI at the shared-budget cap; no adaptive planning.

3.2.3 Resource projection guarantees feasible execution.

Table 3 evaluates the projected actions that are executed. All allocations satisfy the encoded linear resource constraints, with zero maximum budget excess. Projection modifies synthetic proposals more than real-context proposals, as indicated by their mean displacement (0.476 versus 0.0039). After projection, all evaluated world-model rollouts satisfy the STL specification with positive exact robustness. Resource feasibility is guaranteed for the encoded linear constraints, whereas STL satisfaction is model relative and does not certify the unknown environment. Robustness to operational perturbations and epidemiological model mismatch is reported in Appendix D.2 (Figure A3).

Table 3: Constraint handling and model-relative verification.

Setting	**Feasible** (%)	**Max excess**	**Projection displacement**	**STL satisfaction** (%)	**STL robustness**
Synthetic	100.0	0	0.476 ± 0.201	100.00	0.0021 ± 0.0004
Real context	100.0	0	0.0039 ± 0.0400	100.00	0.0016 ± 0.0003

Table 4: Realized policy performance in the real-context semi-synthetic evaluation.

Method	**Adm./100K** ↓	**NPI burden**	$J^{\dagger}$ ↓	ΔJ (%)	**EpiMind wins** (p)
Standard ADMM	30.24	0.423	31.51	+1.14	23/36 (0.132)
Centralized MPC	30.65	0.424	31.92	+2.45	32/36 (<0.001)
Independent MPC	30.94	0.424	32.21	+3.40	26/36 (0.011)
Uniform allocation	31.23	0.479	32.67	+4.87	30/36 (<0.001)
EpidRLearn	35.64	0.500	37.14	+19.23	36/36 (<0.001)
PPO	35.69	0.499	37.19	+19.37	36/36 (<0.001)
Incidence-proportional	36.61	0.352	37.67	+20.92	30/36 (<0.001)
No intervention	64.14	0.000	64.14	+105.90	36/36 (<0.001)
EpiMind	**29.89**	0.423	**31.15**	—	—

†Policies are evaluated by $J = \text{Adm.}/100\text{K} + \lambda_{\text{NPI}}\text{NPI}$; ΔJ is the percentage change relative to EpiMind.

3.2.4 Coordination benefits are modest under matched intervention burden.

Table 4 evaluates realized outcomes under a common objective and shared resource constraints. Among the directly comparable planning methods, which incur nearly identical NPI burden (0.423–0.424), EpiMind reduces J by 1.14% relative to global-only ADMM, 2.13% relative to graph-free planning, 2.21% relative to time-shuffled planning, and 3.40% relative to independent MPC. These results suggest benefits from graph structure and temporal allocation, although their magnitude is small. A stricter matched-burden ablation in Appendix Table A5 holds the NPI trajectory fixed step by step, isolating where interventions are allocated from how much is spent. Under this control, the coordination gains fall below 1%, indicating that much of the uncontrolled difference arises from intervention burden rather than allocation alone. Region shuffling is reported only as a sensitivity diagnostic because shuffling after projection breaks the population-weighted budget constraints. A matched-burden component ablation further isolates the coordination mechanism (Appendix Table A5). The results indicate that the coordination signals provide consistent but incremental gains once intervention burden is controlled.

3.2.5 EpiMind transfers to real-context multi-region planning.

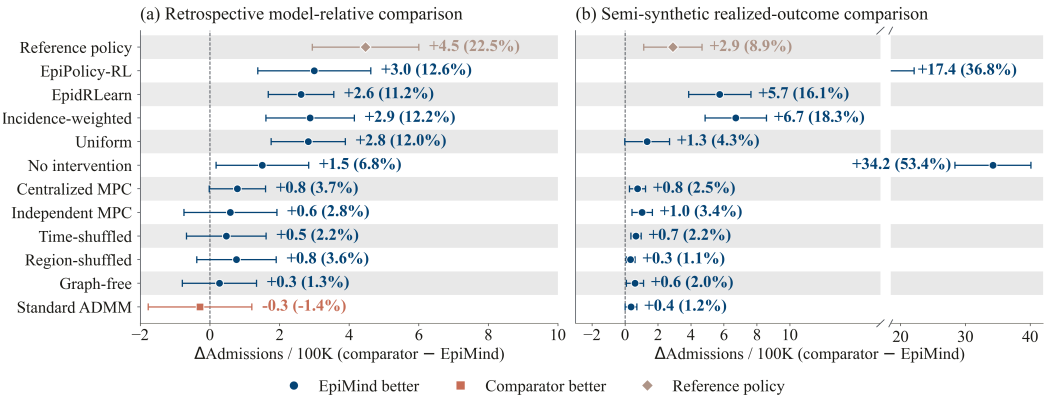


Figure 3: Real-context policy comparisons. Each point reports the mean paired admission difference $\Delta\text{Adm} = \text{Adm}_{\text{comparator}} - \text{Adm}_{EpiMind}$; positive values (blue circles) favor EpiMind; horizontal bars denote 95% confidence intervals. (a) Retrospective Track A reports model-relative projections at held-out U.S. state-level decision origins. (b) Semi-synthetic Track B reports realized outcomes under known simulation dynamics. Diamonds denote the track-specific reference policy.

Figure 3 summarizes paired comparisons across both real-context tracks. In retrospective Track A, EpiMind projects fewer admissions than most comparators, although Standard ADMM is marginally better on average. These comparisons are model relative because the candidate policies were not executed. In semi-synthetic Track B, where counterfactual outcomes are known, EpiMind outperforms every deployable comparator. Its largest gains are over no intervention, EpiPolicy-RL, incidence-weighted allocation, and EpidRLearn. Smaller differences from Standard ADMM, graph-free planning, and shuffled controls indicate that constrained optimization provides most of the improvement, with graph and temporal coordination contributing incrementally.

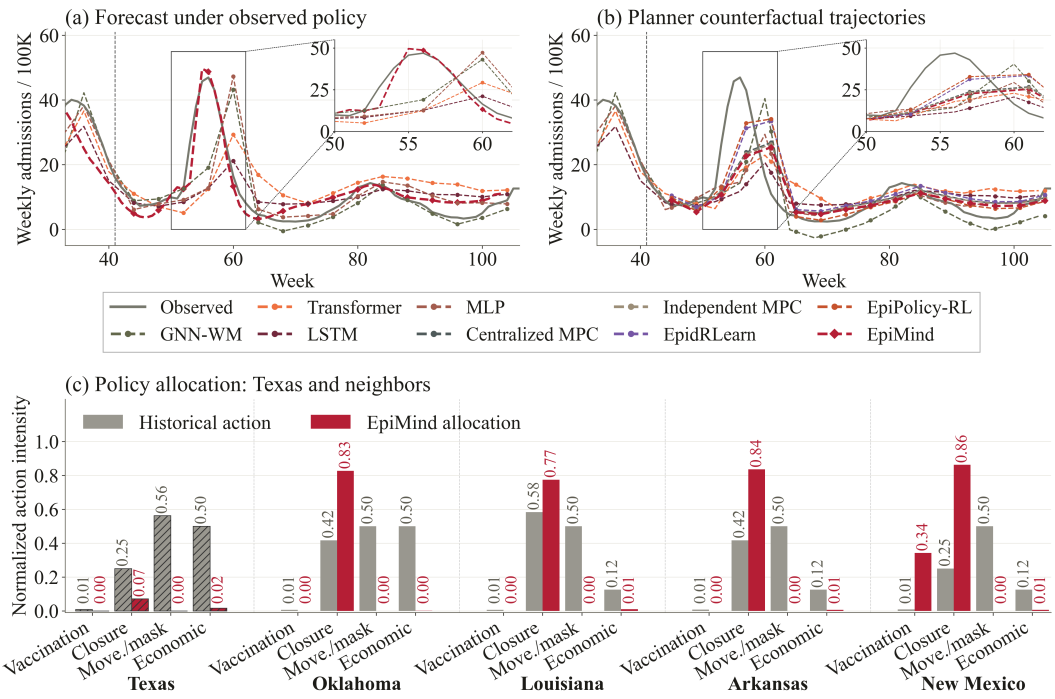


Figure 4: Texas case study. (a) Admission forecasts under the observed policy; the inset enlarges weeks 50–60. (b) Model-relative counterfactual trajectories under matched resource constraints; the observed trajectory provides context but is not an outcome of the unexecuted policies. (c) EpiMind's first-step allocation compared with historical actions in Texas and neighboring states. The dashed vertical line marks the training cutoff. MPC and RL baselines appear only in panel (b).

We next use Texas as a representative decision origin to illustrate the retrospective forecasting and planning workflow (Figure 4, Figure A5). Compared with the forecasting baselines, GF-RSSM more closely tracks the principal admission peak, while EpiMind and the planning baselines produce distinct model-relative trajectories under matched constraints. EpiMind assigns distinct actions to Texas and its mobility-connected neighbors under shared constraints (Figure 4c)). The empirical mobility graph and additional state-level rollouts are shown in Appendix Figures A4 and A6.

4 Conclusion and Limitations

We presented EpiMind, a graph world-model framework for constrained epidemic planning across regions. Its parameter-shared GF-RSSM maintains region-specific beliefs and predicts joint trajectories under candidate policies. GT-ADMM uses these rollouts to coordinate regional interventions and project shared allocations onto the feasible set. The experiments demonstrate accurate policy-conditioned prediction, effective planning through learned dynamics, and exact enforcement of the specified linear resource budgets. Under matched intervention burden, however, graph coordination provides a measurable but incremental benefit.

Several limitations remain. Projection guarantees only the constraints encoded in the feasible set, and STL satisfaction applies to learned trajectories rather than the unknown environment. Planning quality depends on world-model calibration, and the nonconvex GT-ADMM procedure has no global convergence guarantee. Because real-world outcomes under alternative policies are unobserved, the predicted trajectories and spillover effects represent model-based sensitivities rather than causally identified counterfactuals Hernán & Robins (2020). Finally, intervention costs and allocation priorities must reflect local economic, ethical, and public-health considerations. EpiMind is therefore intended to support policy comparison and resource allocation, not to make decisions autonomously.

5 Software and Data

We release the full implementation at `https://anonymous.4open.science/r/epimind-9706/README.md`.

AI USE DISCLOSURE

Generative AI tools were used to assist with literature retrieval and discovery and to improve the clarity and readability of the manuscript. All AI-assisted text was reviewed and revised by the authors, and all citations and literature-derived statements were verified against their original sources. The authors take full responsibility for the final content of this work.

REFERENCES

Julien Arino and Pauline van den Driessche. A multi-city epidemic model. *Mathematical Population Studies*, 10(3):175–193, 2003.

Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The GLobal epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145, 2010.

Dimitris Bertsimas, Vassilis Digalakis Jr, Alexandre Jacquillat, Michael Lingzhi Li, and Alessandro Previero. Where to locate COVID-19 mass vaccination facilities? *Naval Research Logistics*, 69 (2):179–200, 2022.

Alyssa Bilinski, Joshua A. Salomon, John Giardina, Andrea Ciaranello, and Meagan C. Fitzpatrick. Passing the test: A model-based analysis of safe school-reopening strategies. *Annals of Internal Medicine*, 174(8):1090–1100, 2021. doi: 10.7326/M21-0600. URL https://doi.org/10.7326/M21-0600. PMID: 34097433.

Thomas A. Birkland, Kristin Taylor, Deserai A. Crow, and Rob A. DeLeo. Governing in a polarized era: Federalism and the response of U.S. state and federal governments to the COVID-19 pandemic. *Publius: The Journal of Federalism*, 51(4):650–672, 2021. doi: 10.1093/publius/pjab024.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011a.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011b. doi: 10.1561/2200000016.

Jan M. Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, et al. Inferring the effectiveness of government interventions against covid-19. *Science*, 371(6531):eabd9338, 2021. doi: 10.1126/science.abd9338.

Sabah Bushaj, Xuecheng Yin, Arjeta Beqiri, Donald Andrews, and İ. Esra Büyüktahtakın. A simulation-deep reinforcement learning (DRL) approach for epidemic control optimization. *Annals of Operations Research*, 328:245–277, 2023.

Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. Cola-GNN: Cross-location attention based graph neural networks for long-term ILI prediction. In *CIKM*, pp. 245–254, 2020.

Lotty Evertje Duijzer, Willem van Jaarsveld, and Rommert Dekker. Literature review: The vaccine supply chain. *European Journal of Operational Research*, 268(1):174–192, 2018.

Ezekiel J. Emanuel, Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, and James P. Phillips. Fair allocation of scarce medical resources in the time of COVID-19. *New England Journal of Medicine*, 382(21):2049–2055, 2020. doi: 10.1056/NEJMsb2005114.

Tao Feng, Yexin Wu, Guanyu Lin, and Jiaxuan You. Graph world model. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=xjTrTlBbrc.

Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, 2020. doi: 10.1126/science.abb6936.

Seth Flaxman, Swapnil Mishra, Axel Gandy, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

Sebastian Funk, Marcel Salathé, and Vincent A A Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7 (50):1247–1256, 2010. doi: 10.1098/rsif.2010.0142.

Shikha Garg, Lindsay Kim, Michael Whitaker, Alissa O'Halloran, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 — covid-net, 14 states, march 1–30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(15): 458–464, 2020. doi: 10.15585/mmwr.mm6915e3.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025. doi: 10.1038/s41586-025-08744-2.

Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, et al. A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour*, 5 (4):529–538, 2021. doi: 10.1038/s41562-021-01079-8.

Xi He, Eric H. Y. Lau, Peng Wu, Xilong Deng, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature Medicine*, 26(5):672–675, 2020. doi: 10.1038/s41591-020-0869-5.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, FL, 2020.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Emily Howerton, Lucie Contamin, Luke C. Mullany, Michelle Qin, et al. Evaluation of the us covid-19 scenario modeling hub for informing pandemic response under uncertainty. *Nature Communications*, 14(1):7260, 2023. doi: 10.1038/s41467-023-42680-x.

Solomon Hsiang, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, 584:262–267, 2020. doi: 10.1038/s41586-020-2404-8.

Amol Kapoor, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, Martin Blais, and Shawn O'Banion. Examining COVID-19 forecasting using spatio-temporal graph neural networks. In *NeurIPS Workshop on Machine Learning in Public Health*, 2020.

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115(772):700–721, 1927.

Varun Kompella, Roberto Capobianco, Stacy Jong, Jonathan Browne, Spencer Fox, Lauren Meyers, Peter Wurman, and Peter Stone. Reinforcement learning for optimization of COVID-19 mitigation policies. In *AAAI Fall Symposium on AI for Social Good*, 2020.

Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, et al. The effect of human mobility and control measures on the COVID-19 epidemic in china. *Science*, 368 (6490):493–497, 2020. doi: 10.1126/science.abb4218.

Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, et al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases. *Annals of Internal Medicine*, 172(9):577–582, 2020. doi: 10.7326/M20-0504.

Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020. doi: 10.1056/NEJMoa2001316.

Zeeshan Memon, Yiqi Su, Christo Kurisummoottil Thomas, Walid Saad, Liang Zhao, and Naren Ramakrishnan. Toward world models for epidemiology. In *ICLR 2026 the 2nd Workshop on World Models: Understanding, Modelling and Scaling*, 2026. URL https://openreview.net/forum?id=T5ACq6FQqh.

Gideon Meyerowitz-Katz and Lea Merone. A systematic review and meta-analysis of published research data on covid-19 infection fatality rates. *International Journal of Infectious Diseases*, 101:138–148, 2020. doi: 10.1016/j.ijid.2020.09.1464.

Siddharth Nayak et al. Multi-agent reinforcement learning for decentralized epidemic control. *arXiv preprint arXiv:2301.11367*, 2023.

Abu Quwsar Ohi, M F Mridha, Muhammad Mostafa Monowar, and Md Abdul Hamid. Exploring optimal control of epidemic spread using reinforcement learning. *Scientific Reports*, 10(1):22106, 2020.

Minah Park, Alex R. Cook, Jue Tao Lim, Yinxiaohe Sun, and Borame L. Dickens. Reproduction numbers of covid-19: a systematic review. *Journal of Clinical Medicine*, 9(4):967, 2020. doi: 10.3390/jcm9040967.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347.

Gillian K. SteelFisher, Mary G. Findling, Hannah L. Caporello, Keri M. Lubell, Kathleen G. Vidoloff Melville, Lindsay Lane, Alyssa A. Boyea, Thomas J. Schafer, and Eran N. Ben-Porath. Trust in US federal, state, and local public health agencies during COVID-19: Responses and policy implications. *Health Affairs*, 42(3), 2023. doi: 10.1377/hlthaff.2022.01204.

Yiqi Su, Ray Lee, Jiaming Cui, and Naren Ramakrishnan. How (not) to hybridize neural and mechanistic models for epidemiological forecasting. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=otBTuq5AKB.

Huan Wang and Arindam Banerjee. Randomized smoothing variance reduction for stochastic admm. In *Proceedings of the SIAM International Conference on Data Mining*, 2019.

Yifei Wang et al. DMWM: Dual-mind world model with long-term imagination. *arXiv preprint*, 2025.

Joseph T. Wu, Kathy Leung, Mary Bushman, Nishant Kishore, et al. Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china. *Nature Medicine*, 26(4):506–510, 2020. doi: 10.1038/s41591-020-0822-7.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Baez, Siddharth Bhatt, Pieter Abbeel, and Thanasis Karagiannis. The surprising effectiveness of PPO in cooperative multi-agent games. *NeurIPS*, 2022.

A EPIDEMIOLOGICAL FOUNDATIONS

This section introduces the epidemiological structure underlying EpiMind, including its latent dynamics, observations, and planning constraints.

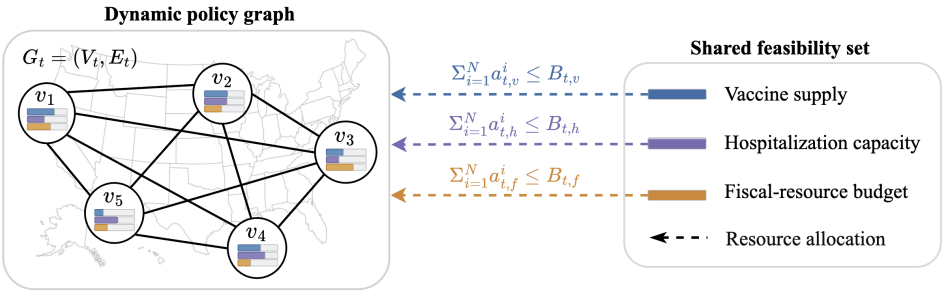


Figure A1: Multi-region policy graph under shared resource constraints. Regions are represented as nodes v_i in a dynamic policy graph $G_t = (V_t, E_t)$, with edges encoding interregional mobility coupling. Each region receives allocations of vaccine, hospitalization-capacity, and fiscal resources. The joint regional allocations must satisfy graph-level resource budgets, $\sum_{i=1}^N a_{t,k}^i \le B_{t,k}$, for $k \in \{v, h, f\}$.

A.1 EPIDEMICS AS GRAPHS.

The spatiotemporal policy graph enables reasoning at three levels:

- **Node level.** Each node represents a region and encodes its latent state, healthcare capacity, population characteristics, and interventions through policy actions.
- **Edge level.** Each edge encodes inter-regional coupling, including mobility flows, spatial proximity, or shared infrastructure that mediates epidemic spillovers and policy coordination between connected regions.
- **Graph level.** Global resource constraints (vaccine supply, hospitalization capacity, fiscal-resource budget) and equity requirements operate over the entire graph, coupling all regions' feasible action sets.

Latent states. As the true compartmental state is *never directly observed*, surveillance systems report cases (a function of testing), hospitalizations (a function of severity and care-seeking), and deaths (a lagged indicator). Each of these is a noisy, delayed, and policy-dependent projection of the underlying state. For instance, expanding testing increases reported cases without changing true cases, while reducing testing has the opposite effect. This means that an observed decline in cases may reflect either genuine transmission reduction or reduced testing coverage, a fundamental ambiguity that forecasting models cannot resolve without separating latent state from observation. This motivates the use of a *latent state representation* that captures the true epidemic reality, including infection burden, immunity, behavioral compliance, and variant fitness behind the noisy observations Memon et al. (2026). By separating latent dynamics from a policy-dependent observation model, this representation disentangles genuine transmission changes from surveillance artifacts, enabling policy reasoning based on inferred reality rather than distorted measurements.

Interventions, behavioral mediation, and delayed effects. Epidemic interventions include NPIs, PIs, and surveillance interventions. A critical feature of these interventions is *behavioral mediation* Funk et al. (2010). Mandates do not directly reduce transmission. Instead, they alter human behavior, which in turn changes contact patterns and infection risk. Compliance with interventions is partial, heterogeneous, and time-varying, depending on perceived risk, fatigue, trust, and economic pressure. In addition, all interventions operate with *temporal delays*. An NPI enacted today cannot reduce transmission until behavioral changes propagate through the population (typically 1–2 weeks, corresponding to the generation interval τ_g). Vaccination requires weeks to build immunity. These delays mean that the effect of an action taken at time t does not appear in surveillance

data until $t + \tau_g$ or later, so a planner cannot validate or course-correct an active intervention against current observations. By the time an intervention's effect becomes visible, the epidemic has already evolved into a different state, which is precisely why planning under epidemic dynamics must rely on forward simulation of latent state under candidate interventions rather than on real-time feedback from surveillance alone.

Resource constraints. Since the resources available for intervention are *shared across regions and physically finite*, epidemic policy can be formulated as a constrained optimization, with two properties that distinguish it from standard constrained programs. First, feasibility must hold at every decision step, since vaccine doses cannot be administered beyond the available stock at time t, hospital and ICU occupancy cannot exceed bed capacity, and per-period public-health expenditures cannot exceed the appropriated budget. Formulations that incorporate constraints as additive cost terms drive expected violation to zero only asymptotically. They are insufficient here, because an over-allocation at any single step cannot be implemented in practice. Second, each constraint is a sum across regions rather than a per-region limit, as the underlying resources are pooled at the national or system-wide level. Therefore, each region's feasible action set depends on what the others request, and the optimization cannot be decomposed into independent single-region problems.

A.2 From Policy-Conditioned Prediction to Constrained Coordination

The epidemiological properties described in Section A create the following unique computational challenges that current methods only partially address:

(A) **Policy-conditioned prediction under uncertainty.** Epidemic dynamics are nonlinear, partially observed, and shaped by latent behavioral responses Su et al. (2026). Planning therefore requires inferring regional latent states from noisy surveillance and generating joint trajectories under candidate interventions.

(B) **Coordination under shared constraints.** Regional decisions are coupled by mobility-driven spillovers, finite shared resources, and delayed intervention effects Flaxman et al. (2020). Effective planning must account for these interactions while enforcing resource limits and temporal requirements explicitly.

(C) **Verification after projection.** Resource projection can alter the planner's unconstrained proposal. Specifications must therefore be evaluated on a new joint rollout under the projected action that will be executed. This separates exact resource feasibility from model-relative temporal verification.

Table A1: Epidemiological challenges and their computational treatment in EpiMind.

Challenge	Epidemiological property	Component	Computational response
(A)	Latent disease burden	GF-RSSM	Latent belief inference
	Policy-dependent surveillance	GF-RSSM	Action-conditioned observation model
	Delayed intervention effects	GF-RSSM	Multi-horizon policy-conditioned rollout
	Cross-region spillovers	GF-RSSM	Graph attention over neighbors
(B)	Shared resource limits	GT-ADMM	Capped-simplex projection
	Cross-border coordination	GT-ADMM	Neighbor-consensus signal
	Spillover externalities	GT-ADMM	Spillover-sensitivity accumulator
	Temporal specifications	GT-ADMM/STL	Differentiable robustness objective
(C)	Proposal–execution mismatch	Verification	Joint reroll under projected action

These challenges motivate the pipeline in Figure 1. The learned graph world model maps the current posterior beliefs and a candidate joint action sequence to a differentiable joint trajectory. GT-ADMM uses these trajectories as its predictive objective, updates region-specific action sequences, and projects the joint allocation onto the shared-resource polytope. Finally, the framework rerolls the projected action and evaluates exact STL robustness under the learned model. Dynamics learning remains predictive: constraint satisfaction is not inserted into the world-model training loss. The

interaction between learning and planning occurs operationally, because executed actions determine the observations used for the next posterior update.

Table A1 summarizes the division of computational responsibilities. The GF-RSSM answers how the coupled regional epidemic is predicted to evolve under a candidate joint intervention. GT-ADMM answers which region-specific intervention sequence minimizes the predicted objective while respecting the specified shared budgets. Post-projection reroll distinguishes the proposed trajectory from the model-predicted trajectory under the action that will actually be executed. This modular separation permits the dynamics model to be trained for predictive fidelity and the planner to enforce operational constraints explicitly, without claiming that optimization constraints reshape the learned dynamics.

B EpiMind Framework

Algorithm 1 EpiMind constrained receding-horizon planning

Require: Frozen world model $\mathcal{M}_\theta$, graph G_t, observation $\mathbf{o}_t$, previous plan $\mathbf{a}^\star_{t-1}$, feasible set $\mathcal{Z}_t$, temporal specifications Φ, horizon H
Ensure: Executed joint action $\mathbf{a}^\star_t$

Update regional beliefs: $(\mathbf{b}_t, \mathbf{z}_t) \leftarrow \mathrm{Posterior}_\theta(\mathbf{o}_t, \mathbf{a}_{t-1}, G_t)$
Warm-start action sequences $\mathbf{a}^0 \leftarrow \mathrm{Shift}(\mathbf{a}^\star_{t-1})$; set $\mathbf{y}^0 \leftarrow \mathbf{a}^0$ and $\boldsymbol{\gamma}^0, \mathbf{u}^0, \boldsymbol{\eta}^0 \leftarrow 0$
for $k = 0, \ldots, K_{\max} - 1$ **do**
  for all regions i **in parallel do**
    Update $a^{i,k+1}$ by minimizing (13) through a joint prior rollout

$$\hat{\mathbf{x}}_{t+1:t+H} = \mathcal{M}_\theta\big(\mathbf{b}_t, \mathbf{z}_t, a^i, \mathbf{a}^{-i,k}; G_t\big)$$

  end for
  Project shared allocations:

$$\mathbf{y}^{k+1} \leftarrow \mathrm{Proj}_{\mathcal{Z}_t}\big(\mathbf{a}^{k+1} + \mathbf{u}^k\big)$$

  Update $\boldsymbol{\gamma}^{k+1}, \mathbf{u}^{k+1}, \boldsymbol{\eta}^{k+1}$ using (17)–(19)
  if primal residuals satisfy their tolerances **then**
    break
  end if
end for
Reroll the proposal $\mathbf{a}^K$ and projected allocation $\mathbf{y}^K$ through the same joint world model
Evaluate exact model-relative robustness:

$$\rho_{\mathrm{exec}} = \rho_{\mathrm{exact}}\big(\Phi, \mathcal{M}_\theta(\mathbf{y}^K; G_t)\big)$$

Execute the first projected action: $\mathbf{a}^\star_t = \mathbf{y}^K_t$
return $\mathbf{a}^\star_t, \rho_{\mathrm{exec}}$

B.1 Signal Temporal Logic for Epidemiological Rules

We use differentiable STL robustness during optimization and exact robustness for model-relative verification of the projected rollout. In the experiments, the specification requires predicted hospital occupancy to remain below regional capacity over the planning horizon:

$$\Phi^i = \square_{[0,H]}\left(\widehat{H}^i_{t+h} \le H^i_{\mathrm{cap}}\right). \tag{21}$$

A positive exact robustness score certifies satisfaction only on the learned rollout; it is not a guarantee for the unknown environment.

C EXPERIMENTAL SETTINGS

C.1 REAL-WORLD DATA SOURCES AND PREPROCESSING

We construct a weekly U.S. state-level panel from the following sources: *(1) Epidemic burden and hospital resources* are obtained from the HHS COVID-19 Reported Patient Impact and Hospital Capacity by State Timeseries, which provides new COVID-19 admissions, inpatient and ICU occupancy, staffed beds, and ICU capacity. *(2) Reported infections and deaths* are obtained from the CDC COVID-19 state surveillance datasets; these measure reported cases rather than latent infections and are treated as noisy observations. *(3) Vaccination* is drawn from the CDC COVID-19 Vaccinations in the United States, Jurisdiction dataset, including doses delivered and administered, primary-series completion, and booster coverage. *(4) Policy interventions* are taken from the COVID-19 U.S. State Policy Database (CUSP) (GitHub), which records state-level mask requirements, gathering limits, stay-at-home orders, school and business closures, emergency declarations, and economic-support policies. We aggregate active mitigation policies into a normalized NPI-intensity index and use the economic-policy fields as fiscal-support indicators. *(5) Interregional mobility* is obtained from Advan Patterns+ (data dictionary), whose visitor origins are aggregated into directed state-to-state flows defining the time-varying graph G_t. *(6) State populations* are obtained from the U.S. Census Bureau's 2021 Population Estimates Program and are used to calculate per-100K outcomes and population-weighted allocations.

C.2 BASELINE METHODS

We organize the baselines by the capability being evaluated.

World-model baselines. *Persistence* repeats the latest observation, whereas *climatology* predicts the training-set mean. *Ridge+action* and *VARX(1)+action* are deterministic linear predictors conditioned on regional interventions. *Action-LSTM* is a recurrent sequence model conditioned on the joint action vector Hochreiter & Schmidhuber (1997). *GF-RSSM (no graph)* retains the recurrent latent-state architecture but removes graph attention, isolating the contribution of mobility-based message passing. All forecasting models use the same train/test split and are evaluated in the same observation units and rollout protocol.

Planning baselines. *No intervention* applies zero intervention throughout the horizon. *Best feasible constant* selects a time-invariant NPI level by grid search under the same budget and evaluation objective. *Uniform* divides each shared resource equally across regions, while *incidence-weighted* allocates resources in proportion to current reported incidence. *Greedy* assigns resources according to the immediate predicted health benefit without multi-step optimization. *Centralized MPC* jointly optimizes all regional actions through a common finite-horizon objective, whereas *independent MPC* optimizes each region without cross-region coordination. *MPC-SEIR* plans with access to the simulator's compartmental state and fixed epidemiological parameters and is therefore an informed simulator-based comparator rather than a deployable real-data method. *Standard ADMM* retains only the global resource-consensus variable and projection of conventional distributed ADMM Boyd et al. (2011a). *D-ADMM* provides a distributed optimization baseline without EpiMind's learned graph coordination. The *graph-free* ablation removes mobility coupling from EpiMind, while *independent planning* removes both graph coupling and cross-region coordination. Region- and time-shuffled controls preserve intervention burden while disrupting spatial or temporal allocation.

Policy-learning baselines. We compare with PPO Schulman et al. (2017), MAPPO Yu et al. (2022), hierarchical reinforcement learning (HRL), and adapted implementations of EpidRLearn and EpiPolicy-RL. These methods are trained under the same action bounds and resource budgets; their outputs are passed through the same final projection used for EpiMind. Because the adapted epidemic-policy baselines do not use the authors' original implementations, we treat them as representative algorithmic adaptations rather than exact reproductions of published results.

Oracle reference. *Oracle dynamics* uses the same planning interface as EpiMind but replaces the learned rollout with the hidden simulator dynamics. It measures the effect of dynamics-model error and is reported as an idealized reference, not as a deployable baseline.

C.3 PARAMETER TABLE

Table A2 summarizes the epidemiological, intervention, observation, world-model, planner, budget, objective, training, and evaluation parameters used in the experiments, together with their values and supporting references or implementation rationale.

Table A2: Parameter inventory.

Symbol	Description	Value	Reference / Justification
Epidemiological			
β	Transmission rate	0.25/day	Park et al. (2020); $R_0 = \beta/\gamma = 2.5$
σ	E→I rate	0.25/day (4 d)	Li et al. (2020); Lauer et al. (2020)
γ	I→R rate	0.1/day (10 d)	He et al. (2020)
δ_H	Hospitalization rate	0.15	Garg et al. (2020) (pre-Omicron)
δ_D	Base death rate	0.01	Meyerowitz-Katz & Merone (2020)
Intervention effects			
NPI β-red.	Max local NPI effect	60%	Brauner et al. (2021): 13–77%
NPI import red.	Max effect on imported force	40%	Travel restriction is partial
FOI split	Local : imported weight	0.7 : 0.3	Mobility coupling strength
Vacc. uptake	$S \to R$ rate at full allocation	0.01/day of S	Vaccinated S move directly to R
Fiscal scale	Funding→compliance	0.3	Hale et al. (2021)
Capacity and overflow			
LOS	Hospital length of stay	7 days	Converts incidence to occupancy
$m_{\max}$	Max overflow mortality mult.	5.0	$\delta_D^{\text{eff}} = \min(\delta_D\, m, 1)$
$c_{\text{base}}, c_{\text{surge}}$	Capacity fractions of pop.	0.001, 0.002	Baseline plus surge at full allocation
Observation model			
σ_{obs}	Measurement noise std	0.005	Additive Gaussian on all channels
p_{detect}	Base detection prob.	0.7	Wu et al. (2020): 14–86%
NPI det. boost	Testing scales with NPI	$+0.2$, $[0.3, 1]$	Bilinski et al. (2021)
τ_{delay}	Reporting delay	1 week	Infection-to-confirmation lag
Graph world model (GF-RSSM)			
d_b, d_z, d_c	Belief / state / context	64 / 16 / 32	Hafner et al. (2025); 5-region problem
n_{heads}	Graph attention heads	4	
d_o	Observation channels	7	
$\sigma_{\min}$	Min prior/posterior std	0.01	
GT-ADMM coordinator			
ρ_e, ρ_g	Edge / global penalties	1.0	Boyd et al. (2011b) (ADMM default)
σ_{prox}	Proximal weight	0.1	Wang & Banerjee (2019)
$K_{\max}$	Max ADMM iterations	15	
inner steps	Gradient steps per x-step	3	
η_x	x-step learning rate	0.02	
β_{STL}	Smooth-STL temperature	0.5	Smooth for gradients
Shared budgets			
$B_{\text{vac}}, B_{\text{hosp}}, B_{\text{fisc}}$	Resource budgets	3.0, 2.5, 3.5	0.6 / 0.5 / 0.7 per region
B_{NPI}	Shared NPI budget	3.0	0.6 per region
Objective			
λ_{NPI}	NPI cost weight	0.1/3/{3,10,30}	Track A / Track B / synthetic sweep
λ_{res}	Resource cost weight	0.02	Prices the budget channels
Training and evaluation			
H_{rollout}	Planning rollout horizon	4 steps	CDC 4–6 week (Howerton et al., 2023)
T	Episode length	26 weeks	
Episodes	Train / val / test	128 / 32 / 32	Chronological split
Epochs	Max, with early stopping	1000	Validation-based selection
Seeds	Training / evaluation	3 / 5	

D RESULTS

D.1 WORLD MODEL EVALUATION

Figure A2 separates one-step admission accuracy from cumulative open-loop error. The full GF-RSSM achieves the lowest admission MAE and the lowest cumulative error at $H = 5$. Its advantage narrows by $H = 10$, and the action-conditioned LSTM performs better at $H = 20$, indicating greater long-horizon drift in the GF-RSSM. Removing graph attention consistently increases rollout error, while all learned models outperform the persistence and climatology references.

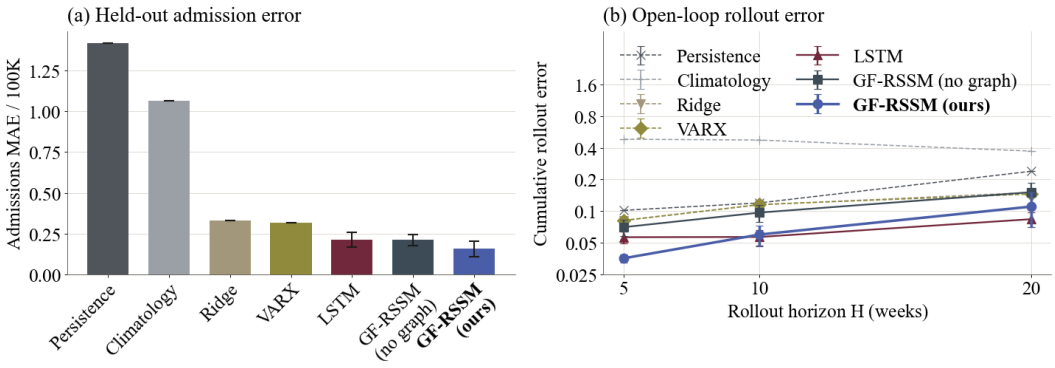


Figure A2: Extended world-model evaluation on the synthetic benchmark. (a) Held-out admission MAE for learned architectures and statistical baselines under the same evaluation protocol. Error bars for learned models show mean ± s.d. over three training seeds. (b) Relative cumulative admission error over open-loop rollout horizons $H \in \{5, 10, 20\}$, shown on a logarithmic scale.

D.2 ROBUSTNESS ANALYSIS

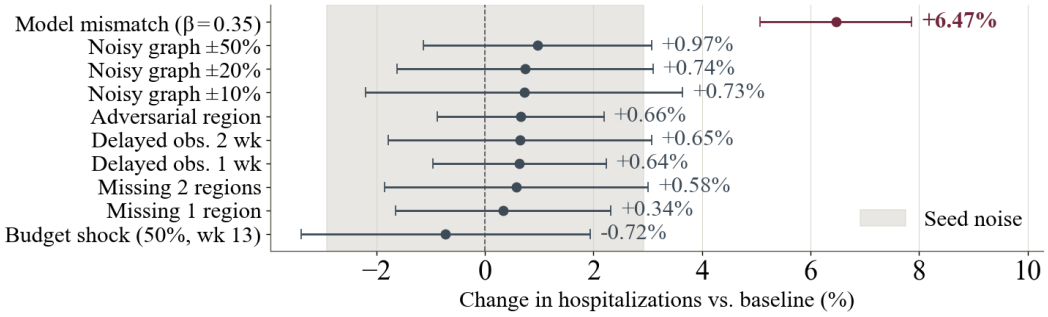


Figure A3: Sensitivity to operational and model perturbations. Changes in cumulative hospitalizations relative to each seed-matched baseline, reported as mean ± s.d. The shaded band shows the largest within-condition standard deviation among the operational perturbations. Graph noise, masked regional observations, reporting delays, a mid-horizon budget cut, and regional non-compliance remain within this descriptive variability band. Epidemiological model mismatch produces the only substantially larger mean degradation.

Figure A3 evaluates sensitivity to graph noise, missing or delayed observations, a mid-horizon budget cut, regional non-compliance, and epidemiological model mismatch. Across five paired seeds, the operational perturbations change cumulative hospitalizations by less than 1% on average and remain within seed-level variability. Model mismatch produces the largest degradation (+6.47%), identifying misspecified epidemic dynamics as the dominant tested failure mode. Given the limited number of seeds, these results are descriptive rather than evidence of statistical equivalence.

D.3 Planning and Calibration Analysis

Table A3 expands the RQ2 regret analysis across intervention costs λ and planning horizons H. Learned- and oracle-dynamics planners use the same optimizer and per-cell best-constant comparator. The model effect is therefore the paired regret difference $R_{\text{learned}} - R_{\text{oracle}}$, isolating the change produced by replacing the learned rollout with the simulator dynamics.

Table A3: Learned- versus oracle-dynamics planning regret. Values are mean ± s.d.

λ	H	**Regret ↓**		**Model effect ↓**	$p^\dagger$
		Learned	Oracle		
3	4	32.5 ± 32.8	27.8 ± 3.2	4.7 ± 34.6	1.000
3	8	38.2 ± 50.0	17.6 ± 0.7	20.5 ± 50.3	1.000
3	12	32.5 ± 34.2	20.6 ± 0.2	11.9 ± 34.2	1.000
10	4	217.0 ± 229.3	103.2 ± 3.8	113.9 ± 231.3	1.000
10	8	184.9 ± 201.1	79.3 ± 0.4	105.6 ± 200.9	1.000
10	12	186.4 ± 198.0	78.2 ± 3.1	108.2 ± 199.7	0.625
30	4	593.9 ± 517.6	328.6 ± 6.8	265.4 ± 513.3	0.625
30	8	457.9 ± 401.8	232.7 ± 15.0	225.2 ± 392.6	0.625
30	12	459.1 ± 440.5	235.9 ± 27.9	223.2 ± 424.9	0.625

† p is an exact two-sided sign test.

The learned planner has higher mean regret in every configuration, but its variation across cells is large: the model-effect standard deviation exceeds its mean in every row, and no sign test is significant. The grid therefore suggests a rollout-model contribution to regret but does not establish its magnitude at this sample size.

Table A4 tests whether the dedicated hospitalization head and its calibration terms improve planning while holding the planner and comparator fixed. Because all variants are scored against the same per-cell best-constant policy, comparisons are paired.

Table A4: Calibration ablation. Values are mean ± s.d. over n=9 cells.

	Regret ↓			**vs. decoder**		**vs. uncalibrated**	
λ	Calibrated	Shared dec.	Uncalib.	cells	p	cells	p
3	56.5 ± 50.5	72.8 ± 85.4	68.8 ± 42.2	5/9	1.000	7/9	0.180
10	252.8 ± 242.3	705.3 ± 586.5	364.9 ± 309.3	8/9	*0.039*	9/9	*0.004*
30	641.4 ± 474.1	2401.9 ± 906.1	882.1 ± 634.7	9/9	*0.004*	7/9	0.180

At λ=10, the calibrated head reduces mean regret by 64.2% relative to the shared decoder and by 30.7% relative to the uncalibrated head, with both paired tests significant. At λ=30, it reduces mean regret by 73.3% relative to the shared decoder, but the additional benefit over the uncalibrated head is not significant. Neither contrast is established at λ=3. These results support the dedicated head and calibration terms at intermediate intervention costs, while showing that aggregate mean ratios should not be interpreted as uniform per-cell gains.

D.4 Coordination Ablation

We isolate the contributions of neighbor consensus γ, predicted spillover sensitivity η, and global resource consensus μ under a matched-burden protocol. Every variant follows the full model's stepwise NPI-burden trajectory, so differences reflect where and when interventions are allocated rather than total NPI use.

Full EpiMind achieves the lowest cumulative admissions (656.81/100K), but the matched-burden gains are limited. Removing global resource consensus (μ) produces the largest individual degradation ($+0.94\%$), followed by removing spillover sensitivity (η, $+0.54\%$) and neighbor consensus (γ, $+0.47\%$). Disabling all three signals increases admissions by 0.88%. Thus, the coordination components provide complementary but incremental improvements once intervention burden is controlled;

Table A5: Matched-burden coordination ablation. Simulator-realized outcomes on the synthetic benchmark. All variants follow the reference configuration's stepwise NPI-burden trajectory. Here, γ denotes neighbor consensus, η spillover sensitivity, and μ global resource consensus. Δ is the percentage change relative to the full model; lower is better.

Variant	γ	η	μ	**Cum. adm./100K** $\downarrow$	Δ (%)
Graph-free	–	–	–	662.57	+0.88
No edge (γ off)	–	✓	✓	659.87	+0.47
No spillover (η off)	✓	–	✓	660.35	+0.54
No global (μ off)	✓	✓	–	662.99	+0.94
EpiMind	✓	✓	✓	**656.81**	—

the substantially larger differences observed without burden matching partly reflect variation in total intervention effort rather than coordination alone.

D.5 Real Mobility Graph

Figure A4 shows a January 2021 snapshot of the row-normalized Advan mobility graph for ten selected high-flow U.S. states. Rows denote origins and columns denote destinations. The matrix exhibits directed, heterogeneous connectivity, with several dominant interstate links and long-range flows involving California, Florida, and Texas. EpiMind uses these flows as graph-edge weights, allowing the GF-RSSM to learn nonuniform neighbor contributions rather than assuming homogeneous regional mixing.

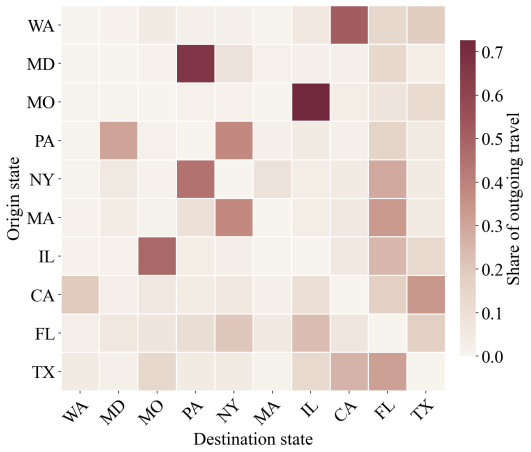


Figure A4: Real interstate mobility graph. Row-normalized Advan device-mobility flows among ten selected high-flow U.S. states in January 2021. Rows denote origins, columns denote destinations, and color indicates each destination's share of an origin's outgoing travel among the displayed states. The asymmetric, nonuniform matrix provides mobility-edge weights to the graph world model.

D.6 Real-Context Prediction and Projection

Figure A5 examines whether forecast-capable models provide both accurate predictions and useful planning signals. EpiMind attains the lowest forecast RMSE (4.1) and projects 9.9 weekly admissions per 100K, compared with 11.8 under the historical-policy reference. Several baselines also project admissions below this reference, but with substantially larger forecast errors. These results distinguish predictive fidelity from projected policy quality; because the proposed policies were not executed, the vertical axis represents model-relative outcomes rather than realized policy effects.

Figure A6 illustrates EpiMind's shared-model behavior across heterogeneous regional trajectories. The historical-policy forecast tracks the timing of major admission waves, although peak magnitude is imperfectly calibrated in several states. The planner rollouts produce state-specific trajectories

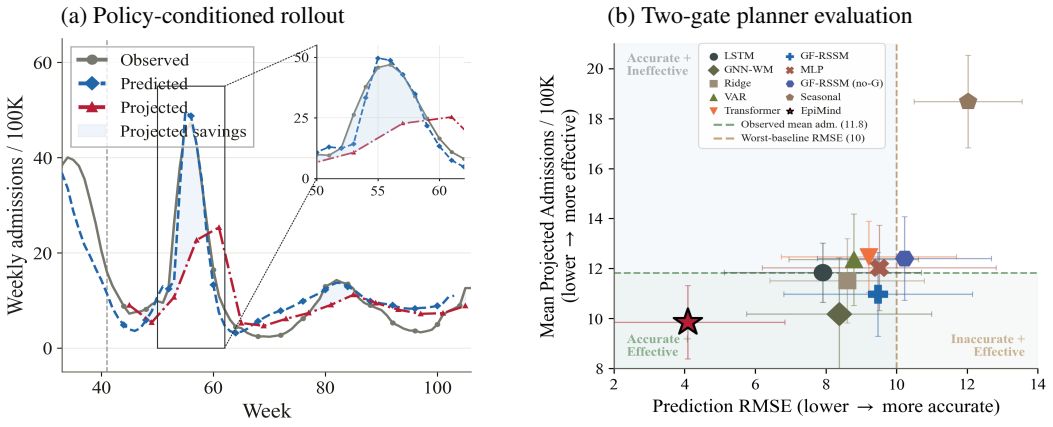


Figure A5: Texas real-context evaluation. (a) Observed admissions, the held-out forecast under historical actions, and the model-relative rollout under EpiMind's proposed policy. The inset enlarges weeks 50–62; shading denotes the predicted difference between the historical- and proposed-policy rollouts. (b) Forecast RMSE versus model-relative admissions projected under each model's optimized policy. Dashed lines mark the observed-policy admission mean and the selected forecast-error reference. Lower values are preferable on both axes.

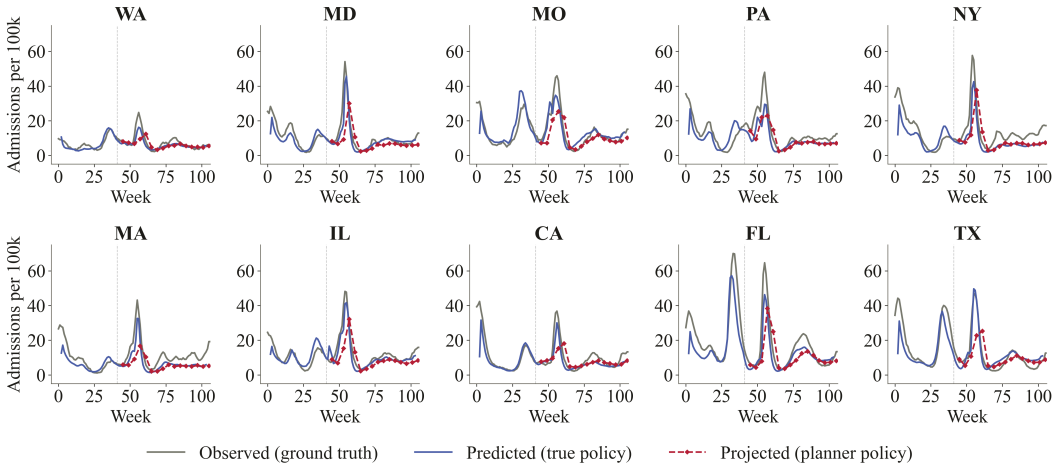


Figure A6: Multi-state forecasting and planner projections. Observed weekly admissions (gray), forecasts under the recorded policy (blue), and model-relative rollouts under EpiMind's proposed policy (red) for ten mobility-connected U.S. states. The vertical dashed line marks the training cutoff. Planner projections represent unexecuted counterfactuals, not observed outcomes.

from the same mobility-coupled model. Because these policies were not executed, the projected reductions are interpreted as model-relative policy comparisons rather than causal effects.