HOW CONSISTENT ARE CLINICIANS? EVALUATING THE PREDICTABILITY OF SEPSIS DISEASE PROGRESSION WITH DYNAMICS MODELS

Unn Seo Park, Venkatesh Sivaraman & Adam Perer
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
upark@andrew.cmu.edu
{venkats,adamperer}@cmu.edu

ABSTRACT

Reinforcement learning (RL) is a promising approach to generate treatment policies for sepsis patients in intensive care. While retrospective evaluation metrics show decreased mortality when these policies are followed, studies with clinicians suggest their recommendations are often spurious. We propose that these shortcomings may be due to lack of diversity in observed actions and outcomes in the training data, and we construct experiments to investigate the feasibility of predicting sepsis disease severity changes due to clinician actions. Preliminary results suggest incorporating action information does not significantly improve model performance, indicating that clinician actions may not be sufficiently variable to yield measurable effects on disease progression. We discuss the implications of these findings for optimizing sepsis treatment.

1 INTRODUCTION

Sepsis is a leading cause of death in hospitals, and there is currently little clinical consensus around best practices for treatment (Centers for Disease Control and Prevention, 2021). Several recent works have applied reinforcement learning (RL) methods in efforts to support clinicians' decision-making on sepsis patients in the intensive care unit (ICU). While these algorithms have shown promise when evaluated using off-policy policy evaluation (OPE) methods, they have also been critiqued for recommending incorrect and even dangerous treatment plans, particularly for more severely ill patients (Jeter et al., 2019; Sivaraman et al., 2023). Due to ethical concerns around prospectively evaluating these models, it is currently an open question whether it is possible to derive policies from public observational datasets that truly improve current clinical practice.

To produce meaningful recommendations with adequate data support, we propose that patient trajectory datasets should exhibit *diversity in observed actions* that correlates with differences in outcomes conditioned on a particular state. In the RL formulation shown in Fig. 1, we assume that for a given state s_t we can estimate not only the cumulative reward of taking observed action a_t, but also the reward for taking a different action a'_t. This would allow the offline-trained RL agent to accurately choose between a_t or a'_t despite having only observed directly the results of the former action.

It is difficult to measure action and outcome diversity conditioned on states directly, since this can depend heavily on how the state is represented. Instead, we constructed an experiment in which we trained transformer-based *dynamics models* to predict future disease severity given a patient's state and optionally the treatment actions that were taken over the subsequent hours. If clinician actions are diverse and have an effect on outcomes, then the action information should improve a model's ability to predict future observed disease severity. Below we present preliminary results from this experiment and discuss their implications for future efforts to optimize sepsis treatment.

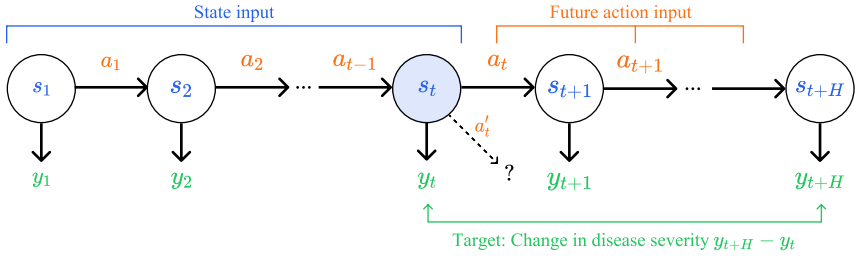

Figure 1: Markov decision process model for patients with sepsis in the ICU. s_t represents the patient state at time t, a_t represents a treatment action, and y_t represents a function of the state that captures the patient's disease severity. Brackets indicate how these values are used in our experiment.

2 Related Work

Several studies (Raghu et al., 2017; Peng et al., 2018; Yu et al., 2019; Liu et al., 2021; Ju et al., 2021) have applied RL to train treatment policies for sepsis patients. For instance, the AI Clinician agent proposed by Komorowski et al. (2018) utilized tabular Q-learning, while subsequent works have proposed combining deep RL with kernel-based RL (Peng et al., 2018), applying deep inverse RL (Yu et al., 2019), integrating physiological models (Nanayakkara et al., 2022), and improving sample efficiency by focusing on important timesteps (Liang et al., 2022; Ju et al., 2021)). Tang et al. (2023) further introduced an approach that leverages factored action spaces to improve the efficiency of offline RL in healthcare settings.

Other research highlights the challenges and shortcomings of these algorithmic approaches. For example, Killian et al. (2020) found that even powerful sequential models were unable to accurately separate patient state representation by mortality. Jeter et al. (2019) suggested that the Q-learning approach may learn to recommend dangerous treatments for severely ill patients because the most common clinician actions have consistently low rewards, whereas rarer actions may have high but noisy estimated values. Gottesman et al. (2020) proposed improving RL policy evaluation by identifying timepoints with a high OPE weight, while Ji et al. (2021) selected and visualized trajectories that may explain policy behavior. To our knowledge, however, predicting future disease outcomes from actions has not been examined as a way to evaluate the feasibility of off-policy RL in sepsis.

3 Methods

3.1 Data and Preprocessing

Patient trajectory data was extracted following Komorowski et al. (2018) and Killian et al. (2020) from MIMIC-IV (Johnson et al., 2020) and the eICU Collaborative Research Database (Pollard et al., 2018).[1][2] Data was aggregated at one-hour intervals, and patients with more than 14 days in the ICU were excluded. Missing data was imputed using a transformer-based autoencoder model. This resulted in a total of 2,060,446 timesteps from 33,779 patients.

The state space for our models consisted of 60 normalized observation variables (vitals, labs, prior treatments, and fluid balances) and 35 demographic variables (age, gender, and Elixhauser comorbidities). The action space comprised log-transformed continuous-valued dosages of IV fluids and vasopressors. Three widely-used severity metrics were used as outcomes: the Sequential Organ Failure Assessment (SOFA) score, the Systemic Inflammatory Response Syndrome (SIRS) score, and Shock Index. Actions and disease severity were z-transformed for model input and output.

[1] While previous work has generally used MIMIC-III, the AI Clinician modeling procedure has been shown to yield consistent results in the two versions of MIMIC (Sivaraman et al., 2023).

[2] Preprocessing and modeling code available at `https://github.com/cmudig/AI-Clinician-MIMICIV`.

3.2 MODELS

Dynamics models Our experiment utilized decoder-only transformer models, where each input "token" comprised embeddings of the patient's observed state, demographics, and actions.[3] The model consisted of two transformer blocks, each comprising 4 self-attention layers, each with 16 attention heads and a total dimension of 1024. The first transformer block took the state and demographic embeddings as input, while the second transformer block added embedded clinician actions. Models were trained on the future disease severity task along with three other proxy tasks: (1) predicting the current state of the patient, (2) predicting whether the current state is the last step in the patient's trajectory, and (3) predicting whether two embeddings correspond to states that are adjacent in time. The proxy tasks were included only to improve the model's convergence and generalizability, and results for these tasks are not shown.

Behavior cloning While the dynamics models described above aimed to predict the difference in disease severity as a function of states and actions, we also trained behavior cloning models to predict clinician actions as a function of states. These models utilized the first transformer block from above to encode the state observations and demographics, then applied a two-layer feedforward network to simultaneously predict fluid and vasopressor dosages at one-hour intervals up to 6 hours ahead.

4 EXPERIMENT RESULTS

4.1 INFLUENCE OF ACTION INPUTS ON DISEASE SEVERITY PREDICTIONS

Three groups of models, totaling 81 dynamics models, were trained to predict changes in future disease severity. The first group was trained with both future action information and state information. In contrast, the second group, featuring identical architectures, had all future actions set to the mean action values (effectively removing them from training). The last group also shared identical architectures, but was trained without the information about states. Furthermore, disease severity changes were measured according to the three metrics described above at 6 hours, 12 hours, and 18 hours ahead. Each model configuration was trained and evaluated across three random weight initializations. We then conducted four evaluations for each model by generating predictions on variants of the test dataset: **True** (original treatment actions), **Zero** (all dosage values set to zero), **Shuffled** (real but randomly-permuted dosages), and **Mean** (all actions replaced with the mean dosages). Fig. 2 shows the root mean squared error (RMSE) of these predictions in z-scaled space, as well as two examples comparing model predictions to ground-truth.

Overall, the RMSE was almost constant across training conditions and action input types except for the **Mean** condition, which generally showed higher error and variance across initializations when actions were used in training (likely because consistently receiving nonzero fluids and vasopressors is a highly unusual input). Among the other three conditions, the range of RMSEs was within 0.05 for SIRS and Shock Index, and within 0.1 for SOFA. Furthermore, performance in the **True** condition was highly similar whether or not actions were provided during training. This null result suggests that actions did not substantially improve the model fit, consistent with our hypothesis that they are not diverse enough for policy learning. In addition, models trained without state information showed similar trends, indicating that action information is largely redundant with the states.

4.2 PREDICTION OF FUTURE ACTIONS WITH BEHAVIOR CLONING

To evaluate the predictability of actions from states more directly, we trained 3 replicates of the behavior cloning model with different random weight initializations. If these models showed a strong fit to the data, one could infer that actions were fully consistent and predictable across clinicians.

Fig. 3 shows that the average R^2 correlations between the true and predicted actions (in log-transformed and z-scaled units) were overall low, particularly after several hours. IV fluid predictions were markedly less correlated with the true values than vasopressor predictions, perhaps because (1) vasopressors are more commonly zero than fluids, increasing the overall predictability

[3] We conducted the same experiments with linear and recurrent networks as well as XGBoost models, but found that transformers yielded the best performance.

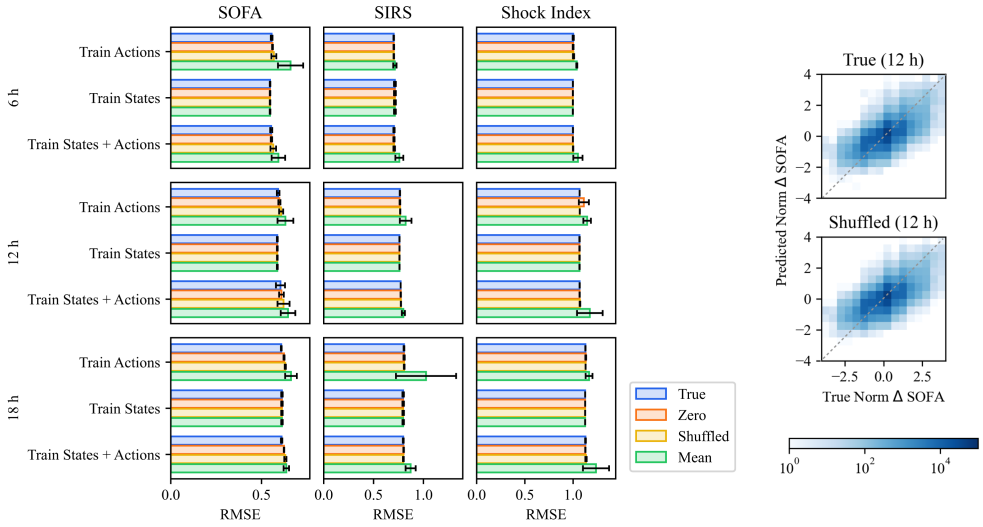

Figure 2: Left: RMSE (lower is better) of the predicted change in disease severity across training schemes ("Train Actions", "Train States," and "Train States + Actions") and action inputs at test time (**True**, **Zero**, **Shuffled**, and **Mean**). Error bars indicate the standard deviation across three random weight initializations. Note that all units are in z-scaled space, so an RMSE of 1 corresponds to 1 standard deviation in the severity metric. Right: example histograms comparing true and predicted changes in SOFA score at 12 hours ahead, in the **True** and **Shuffled** evaluation conditions.

of vasopressor use, or because (2) the amount of IV fluid used is generally more clinician-dependent. The regression models also appeared to struggle with the wide range of fluid dosage values, and tended to predict values within a more constrained range (Fig. 3, third panel).

Aside from the possible modeling issues in the IV fluid predictions, the low correlations across both treatments suggest there is in fact some diversity in clinician actions that could benefit policy learning. However, action diversity does not necessarily correspond to observable differences in outcomes, since there is likely a range of treatment dosages that correspond to similar effects for a given patient state. The results in the preceding section suggest that even when dosage differences exist, they may not yield sufficient differences in outcomes to provide a useful signal to an RL agent.

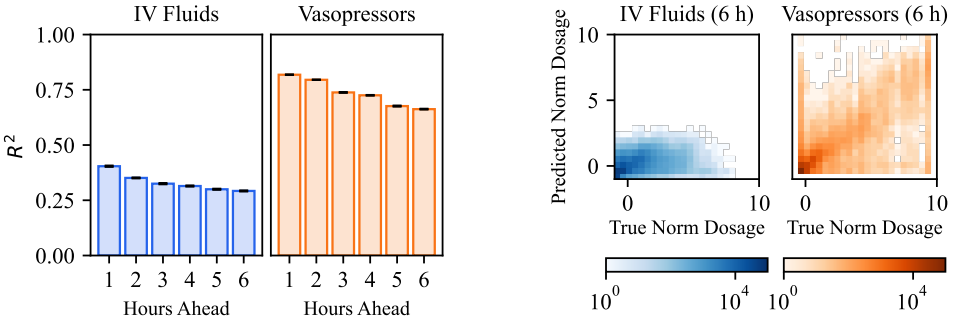

Figure 3: Left: correlations between true and predicted normalized actions from 1 to 6 hours ahead. Right: example histograms of correlations between true and predicted normalized actions at 6 hours.

5 Discussion

This work explored the impact of clinician actions on the predictability of future changes in sepsis disease severity, in order to gain insight into whether actions have sufficient diversity to support

accurate RL-based policies. We found that action information does *not* confer substantive improvements in dynamics model fit, as our transformer models could predict future disease severity almost equally well with or without true actions as input. Taken alone, the dynamics model results in Sec. 4.1 might suggest that actions are fully predictable from the states and there was no need to learn from the action inputs. This echoes results from Beaulieu-Jones et al. (2021), who critique patient risk predictions as "looking over the shoulders of clinicians." However, action prediction (Sec. 4.2) was still fairly noisy, indicating that while variation in actions exists, it is not enough to cause measurable differences in outcomes in our sepsis cohort. Rather, the outcome differences we observe may be more driven by unobserved patient variables or natural random variation.

Some of the observed lack of diversity in actions on MIMIC data may be due to inherent challenges in working with patient trajectories. For instance, there may only be a small number of treatment possibilities that are clinically feasible and safe, limiting the space of actions that clinicians could take. Clinicians may also tend to choose actions in predefined patterns, such as monotonically increasing or decreasing dosages, that appear diverse yet lead to consistent outcomes. Alternatively, missing data imputation could have caused patient states and actions to appear more consistent than they really are. These obstacles are likely to exist in any patient treatment dataset, underscoring the importance of using learning methods that are robust to missingness and a constrained action space.

Another possible explanation for our results is that our models simply didn't learn to use actions effectively, and a better model formulation might yield more pronounced differences between the "Train States" and "Train States + Actions" models. It is impossible to determine *a priori* whether there exists a more effective way to use actions, but we conjecture that if such a method exists, it would likely require more clinically-informed descriptions of actions than what has currently been explored in the literature. For instance, models could use other treatments such as antibiotics and mechanical ventilation, contextualize actions using the patient's physiological state, or limit the training data to only the most important decision points. Future work should incorporate clinician guidance on how to meaningfully encode treatments to further test the effects of action information.

This work highlights the importance of diversity in data sources when building medical recommendation models. While it has been extremely valuable in developing and exploring ways to improve sepsis treatment, the MIMIC dataset is sourced from a single well-resourced hospital in Boston (Johnson et al., 2020), where clinicians are likely to be consistent and compliant with existing practice guidelines. Human-centered ML efforts undertaken in collaboration with clinicians and medical data experts can also inspire more clinically-relevant and performant model formulations, such as focusing on the emergency department (a higher-stress environment that is less specialized towards sepsis than the ICU) or building smaller models that are relevant to specific subgroups of patients (Sivaraman et al., 2023). Through these research directions, applied ML efforts may be able to better utilize available observational data to improve sepsis treatment recommendation.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Jeremy Kahn, Andrew King, Jason Kennedy, and the anonymous reviewers for feedback on the manuscript. This work was supported by a National Science Foundation Graduate Research Fellowship (DGE2140739), and by the Carnegie Mellon University Center of Machine Learning and Health.

REFERENCES

Brett K. Beaulieu-Jones, William Yuan, Gabriel A. Brat, Andrew L. Beam, Griffin Weber, Marshall Ruffin, and Isaac S. Kohane. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *npj Digital Medicine*, 4(1):1–6, March 2021. ISSN 2398-6352. doi: 10.1038/s41746-021-00426-3. URL https://www.nature.com/articles/s41746-021-00426-3. Publisher: Nature Publishing Group.

Centers for Disease Control and Prevention. What is sepsis?, 2021.

Omer Gottesman, Joseph Futoma, Yao Liu, Sonali Parbhoo, Leo Anthony Celi, Emma Brunskill, and Finale Doshi-Velez. Interpretable off-policy evaluation in reinforcement learning by highlighting influential transitions. *37th International Conference on Machine Learning, ICML 2020*, PartF16814:3616–3625, 2020. arXiv: 2002.03478.

Russell Jeter, Christopher Josef, Supreeth Shashikumar, and Shamim Nemati. Does the "Artificial Intelligence Clinician" learn optimal treatment strategies for sepsis in intensive care? *arXiv*, November 2019. ISSN 1078-8956. doi: 10.1038/s41591-018-0213-5. arXiv: 1902.03271.

Christina X. Ji, Michael Oberst, Sanjat Kanjilal, and David Sontag. Trajectory Inspection: A Method for Iterative Clinician-Driven Design of Reinforcement Learning Studies. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2021(i):305–314, 2021. ISSN 1942597X. arXiv: 2010.04279.

A Johnson, L Bulgarelli, T Pollard, S Horng, L A Celi, and R Mark. MIMIC-IV (version 1.0), 2020.

Song Ju, Yeo Jin Kim, Markel Sanz Ausin, Maria E. Mayorga, and Min Chi. To Reduce Healthcare Workload: Identify Critical Sepsis Progression Moments through Deep Reinforcement Learning. *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, pp. 1640–1646, 2021. doi: 10.1109/BigData52589.2021.9671407. Publisher: IEEE.

Taylor W. Killian, Haoran Zhang, Jayakumar Subramanian, Mehdi Fatemi, and Marzyeh Ghassemi. An Empirical Study of Representation Learning for Reinforcement Learning in Healthcare. pp. 1–22, 2020. URL http://arxiv.org/abs/2011.11235. arXiv: 2011.11235.

Matthieu Komorowski, Leo A. Celi, Omar Badawi, Anthony C. Gordon, and A. Aldo Faisal. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24(11):1716–1720, 2018. ISSN 1546170X. doi: 10.1038/s41591-018-0213-5. URL http://dx.doi.org/10.1038/s41591-018-0213-5. arXiv: 1902.03271 Publisher: Springer US.

Dayang Liang, Huiyi Deng, and Yunlong Liu. The treatment of sepsis: an episodic memory-assisted deep reinforcement learning approach. *Applied Intelligence*, 2022. ISSN 15737497. doi: 10.1007/s10489-022-04099-7. Publisher: Applied Intelligence.

Ran Liu, Joseph Greenstein, James Fackler, Jules Bergmann, Melania Bembea, and Raimond Winslow. Offline reinforcement learning with uncertainty for treatment strategies in sepsis. 2021.

Thesath Nanayakkara, Gilles Clermont, Christopher James Langmead, and David Swigon. Unifying cardiovascular modelling with deep reinforcement learning for uncertainty aware control of sepsis treatment. *PLOS Digital Health*, 1(2):e0000012, 2022. doi: 10.1371/journal.pdig.0000012. URL http://dx.doi.org/10.1371/journal.pdig.0000012. arXiv: 2101.08477.

Xuefeng Peng, Yi Ding, David Wihl, Omer Gottesman, Matthieu Komorowski, Li Wei H. Lehman, Andrew Ross, Aldo Faisal, and Finale Doshi-Velez. Improving Sepsis Treatment Strategies by Combining Deep and Kernel-Based Reinforcement Learning. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2018:887–896, 2018. ISSN 1942597X. arXiv: 1901.04670.

Tom J Pollard, Alistair E W Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data*, 5(1):1–13, 2018.

Aniruddh Raghu, Matthieu Komorowski, Leo Anthony Celi, Peter Szolovits, and Marzyeh Ghassemi. Continuous State-Space Models for Optimal Sepsis Treatment - a Deep Reinforcement Learning Approach. 68, 2017. URL http://arxiv.org/abs/1705.08422. arXiv: 1705.08422.

Venkatesh Sivaraman, Leigh A. Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. volume 1. Association for Computing Machinery, 2023. doi: 10.1145/3544548.3581075. arXiv: 2302.00096 Publication Title: Conference on Human Factors in Computing Systems - Proceedings Issue: 1.

Shengpu Tang, Maggie Makar, Michael W. Sjoding, Finale Doshi-Velez, and Jenna Wiens. Leveraging Factored Action Spaces for Efficient Offline Reinforcement Learning in Healthcare, May 2023. URL http://arxiv.org/abs/2305.01738. arXiv:2305.01738 [cs].

Chao Yu, Guoqi Ren, and Jiming Liu. Deep inverse reinforcement learning for sepsis treatment. *2019 IEEE International Conference on Healthcare Informatics, ICHI 2019*, pp. 31–33, 2019. doi: 10.1109/ICHI.2019.8904645. Publisher: IEEE.