# Guideline-informed reinforcement learning for mechanical ventilation in critical care (Extended Abstract)

Floris den Hengst[1] [0000−0001−7139−1203], Martijn Otten[1,2] [0000−0003−3409−3551], Paul Elbers[2] [0000−0003−0447−6893], Frank van Harmelen[1] [0000−0002−7913−0048], Vincent François-Lavet[1] [0000−0002−8593−9740], and Mark Hoogendoorn[1] [0000−0003−3356−3574]

[1] Computer Science dept., Vrije Universiteit Amsterdam
[2] Dept. of Intensive Care Medicine, Amsterdam UMC

## 1 Introduction

Reinforcement learning (RL) is a promising technique to improve decision-making in healthcare because it can handle uncertainty into its sequential decision-making and learns from observational data [4, 10, 7, 9]. However, several challenges related to performance, safety and the alignment of learned solutions with existing knowledge remain before putting RL into clinical practice [8].

In [5], we set out to develop an approach that combines RL with a knowledge-based approach to obtain a hybrid solution that benefits from the best of both worlds [1]. We propose a framework for finding effective and guideline-compliant treatment policies by incorporating treatment guidelines into RL. We evaluate our approach in a case study on mechanical ventilation (MV) optimization using the MIMIC-III database [6].

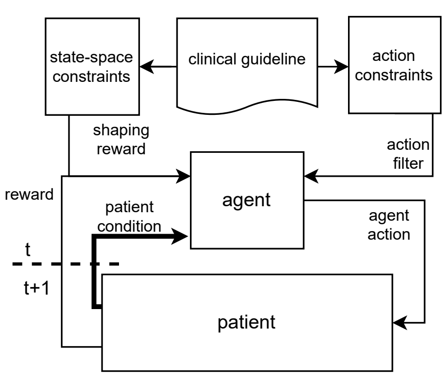
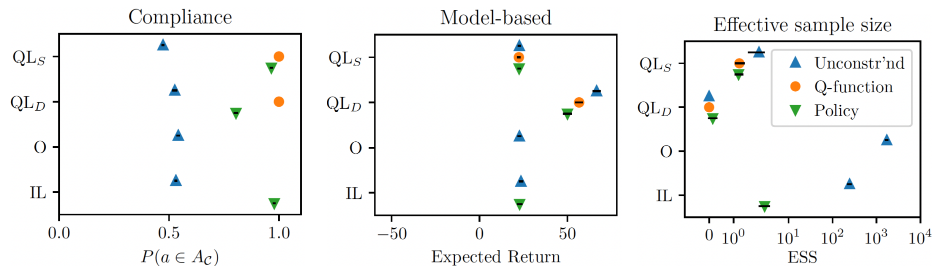



Fig. 1: Guidelines are manually encoded into state- and action constraints. Action constraints are strictly enforced with a filter: (i) after a policy has been learned (Policy), and (ii) during learning (Q-function). State constraints describe desirable properties in the patient and are encoded with shaping rewards. The state representation (patient condition) and the agent actions are learned.




(a) Observed Safety (b) Estimated rewards (c) Effective sample size

Fig. 2: Test set results (mean and 95% CI) obtained using off-policy evaluation.

## 2 Guideline-informed Reinforcement Learning

The proposed framework for RL informed by guidelines generates policies based both on the knowledge encoded in the guidelines and on the experiences obtained in clinical practice. It differs from the standard RL setup in two ways, as visualized and explained in Fig 1. Both extensions of the general RL framework are informed by clinical guidelines.

In our evaluations, we compare these extensions using a model-based off-policy evaluation approach called Fitted Q-Evaluation (FQE) as well as a purely data-drive approach that relies on weighted importance sampling and takes into account variable trajectory lengths (omitted in this abstract) [3, 2]. Figures 2a and 2b respectively report safety expressed as the probability of taking a safe action, and expected returns (associated with probability of survival after 90 days). We see that the two approaches are more safe than the unconstrained approach in both the stochastic (QL$_S$) and deterministic (QL$_D$) policy setting when compared to outcomes observed in clinicians (O) and obtained with Imitation Learning (IL) – at limited cost to estimated reward.

Fig 2c displays the effective sample size (ESS), a metric to assess how many samples contributed to the estimated return during off-policy evaluation (OPE). We see that the ESS is low for learned policies and that this is particularly the case for the deterministic policy (QL$_D$. The full work contains more results an analyses, including the result of reward shaping (no positive effect) and an analysis of the learned policies which are more diverse than those of the clinician's [5].

## 3 Conclusion

We proposed and evaluated a hybrid learning- and knowledge-driven framework for automated clinical sequential decision-making. We found that our approach produces policies that comply with the medical guideline while outperforming clinicians in terms of expected mortality in a model-based evaluation. Our framework can extend existing studies into the use of RL in the medical domain with guideline compliance guarantees and is therefore an important stepping stone in the adoption of RL in clinical practice.

## References


1. van Bekkum, M., de Boer, M., van Harmelen, F., Meyer-Vitali, A., Teije, A.t.: Modular design patterns for hybrid learning and reasoning systems: a taxonomy, patterns and use cases. Applied Intelligence **51**(9), 6528–6546 (2021)
2. Doroudi, S., Thomas, P.S., Brunskill, E.: Importance sampling for fair policy selection. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 5239–5243 (2018)
3. Hao, B., Ji, X., Duan, Y., Lu, H., Szepesvári, C., Wang, M.: Bootstrapping fitted q-evaluation for off-policy inference. In: International Conference on Machine Learning. pp. 4074–4084. PMLR (2021)
4. den Hengst, F., Grua, E.M., el Hassouni, A., Hoogendoorn, M.: Reinforcement learning for personalization: A systematic literature review. Data Science **3**(1), 107–147 (2020)
5. den Hengst, F., Otten, M., Elbers, P., van Harmelen, F., François-Lavet, V., Hoogendoorn, M.: Guideline-informed reinforcement learning for mechanical ventilation in critical care. Artificial Intelligence in Medicine **147**, 102742 (2024)
6. Johnson, A.E.W., Stone, D.J., Celi, L.A., Pollard, T.J.: The mimic code repository: enabling reproducibility in critical care research. Journal of the American Medical Informatics Association **25**(1), 32–39 (2018)
7. Komorowski, M., Celi, L.A., Badawi, O., Gordon, A.C., Faisal, A.A.: The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nature medicine **24**(11), 1716–1720 (2018)
8. Otten, M., Jagesar, A.R., Dam, T.A., Biesheuvel, L.A., den Hengst, F., Ziesemer, K.A., Thoral, P.J., de Grooth, H.J., Girbes, A.R., François-Lavet, V., et al.: Does reinforcement learning improve outcomes for critically ill patients? a systematic review and level-of-readiness assessment. Critical Care Medicine **52**(2), e79–e88 (2024)
9. Peine, A., Hallawa, A., Bickenbach, J., Dartmann, G., Fazlic, L.B., Schmeink, A., Ascheid, G., Thiemermann, C., Schuppert, A., Kindle, R., et al.: Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. NPJ digital medicine **4**(1), 32 (2021)
10. Roggeveen, L., El Hassouni, A., Ahrendt, J., Guo, T., Fleuren, L., Thoral, P., Girbes, A.R., Hoogendoorn, M., Elbers, P.W.: Transatlantic transferability of a new reinforcement learning model for optimizing haemodynamic treatment for critically ill patients with sepsis. Artificial Intelligence in Medicine **112**, 102003 (2021)