Guideline-Informed Reinforcement Learning for Mechanical Ventilation in Critical Care

Floris den Hengst | Martijn Otten | Paul Elbers | Frank van Harmelen | Vincent François-Lavet | Mark Hoogendoorn

Problem: Adoption of RL in Healthcare

RL has recently found many applications in the healthcare domain.

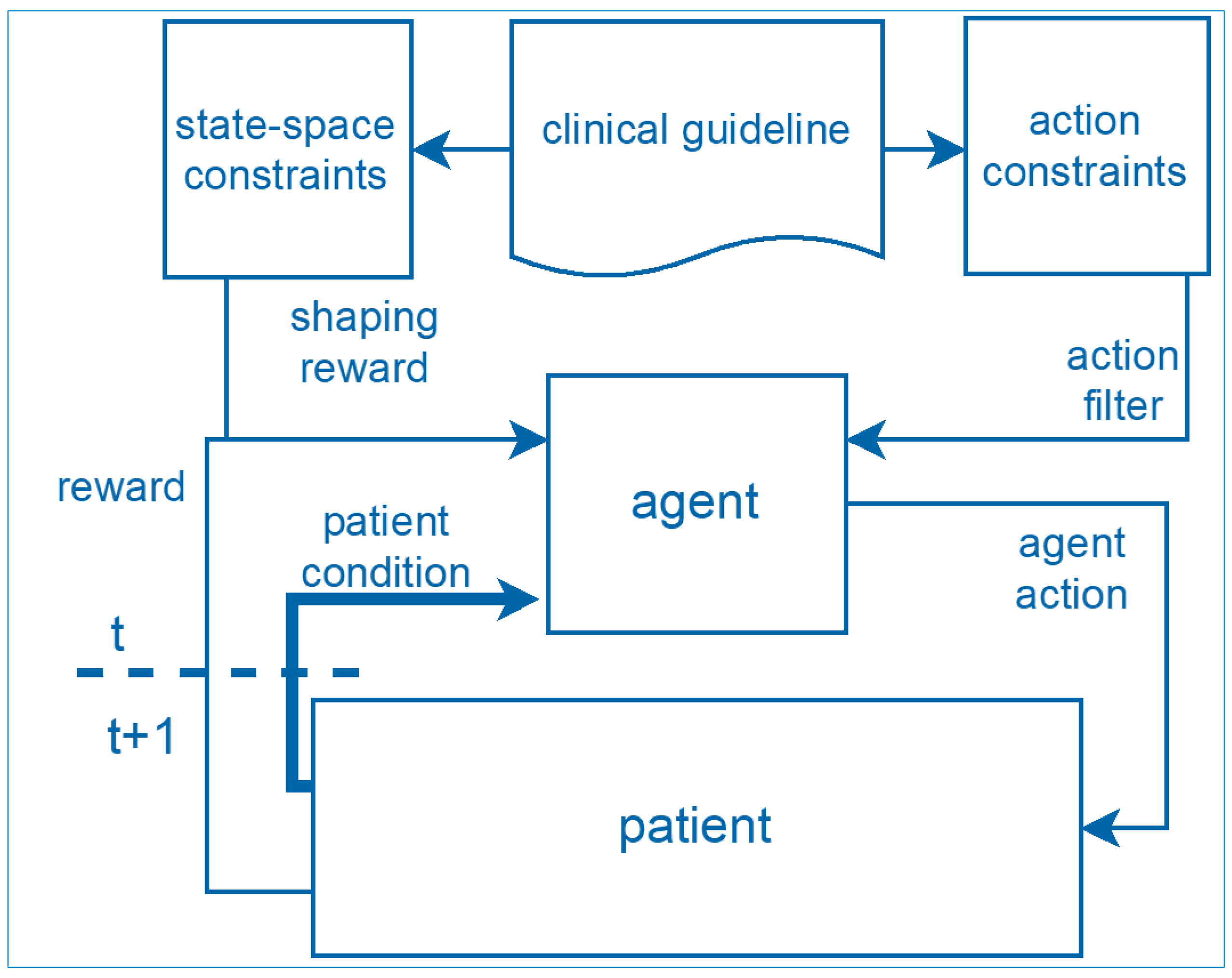
A key challenge in adopting RL-based solution in clinical practice, however, is the inclusion of existing knowledge.

Existing knowledge from medical guidelines may improve safety of solutions, produce a better balance between short- and long-term outcomes for patients, and increase trust and adoption by clinicians.

Methodology

A framework for including knowledge in medical guidelines in RL

Components for enforcing safety constraints and reward shaping to balance short- and long-term outcomes



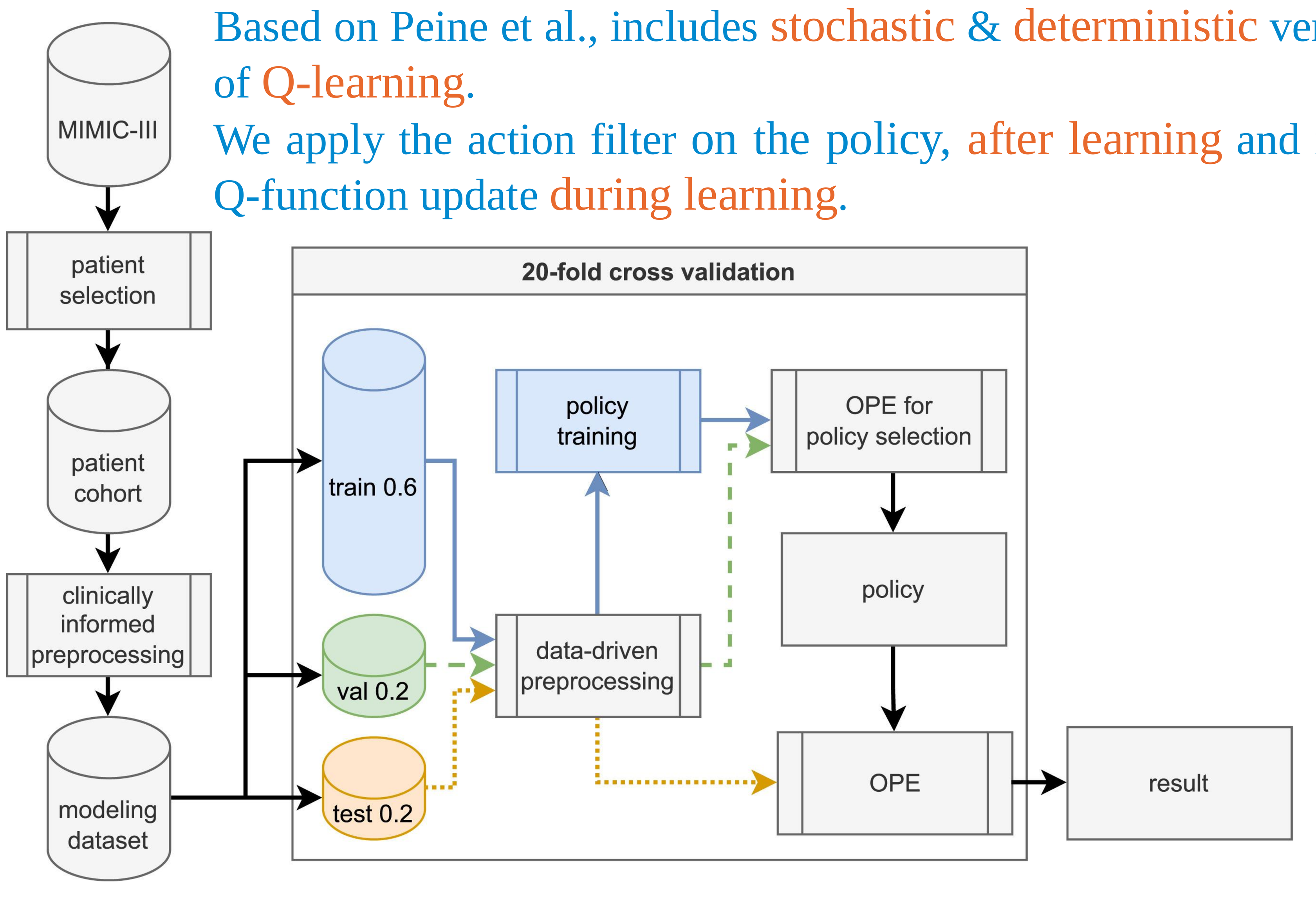
1) Clinical guidelines are manually encoded into state-space constraints and action constraints in collaboration with clinicians.

2) Action constraints describe allowable treatment decisions. These are enforced with a filter that removes all non-compliant treatment actions

3) State-space constraints describe desirable properties in the patient condition. The learning agent is informed of state-space constraints with potential-based reward shaping.

5) The state representation and the agent actions are learned with offline RL

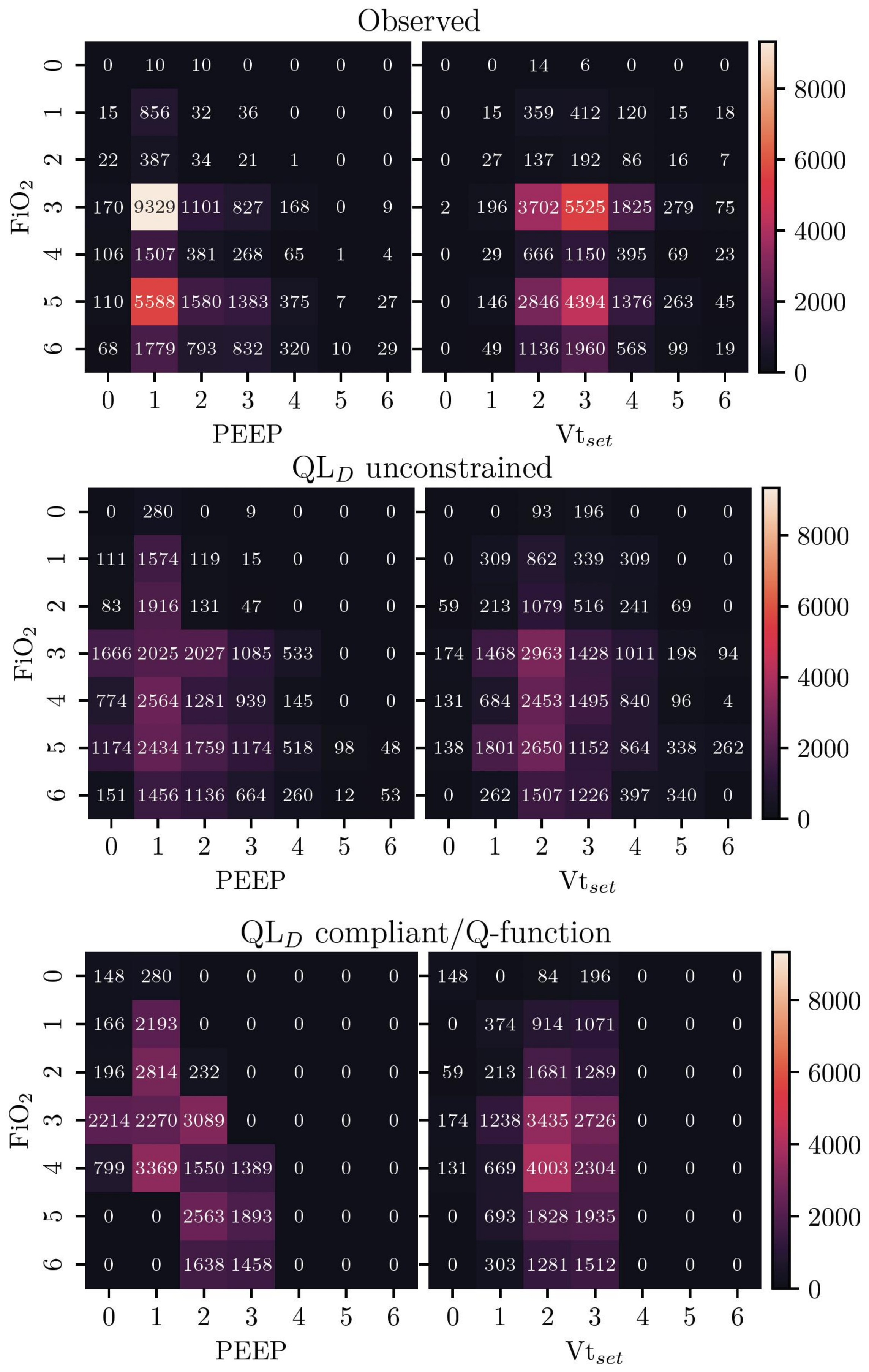
Data & Algorithms



Based on Peine et al., includes stochastic & deterministic versions of Q-learning.

We apply the action filter on the policy, after learning and in the Q-function update during learning.

Comparing Action Preferences



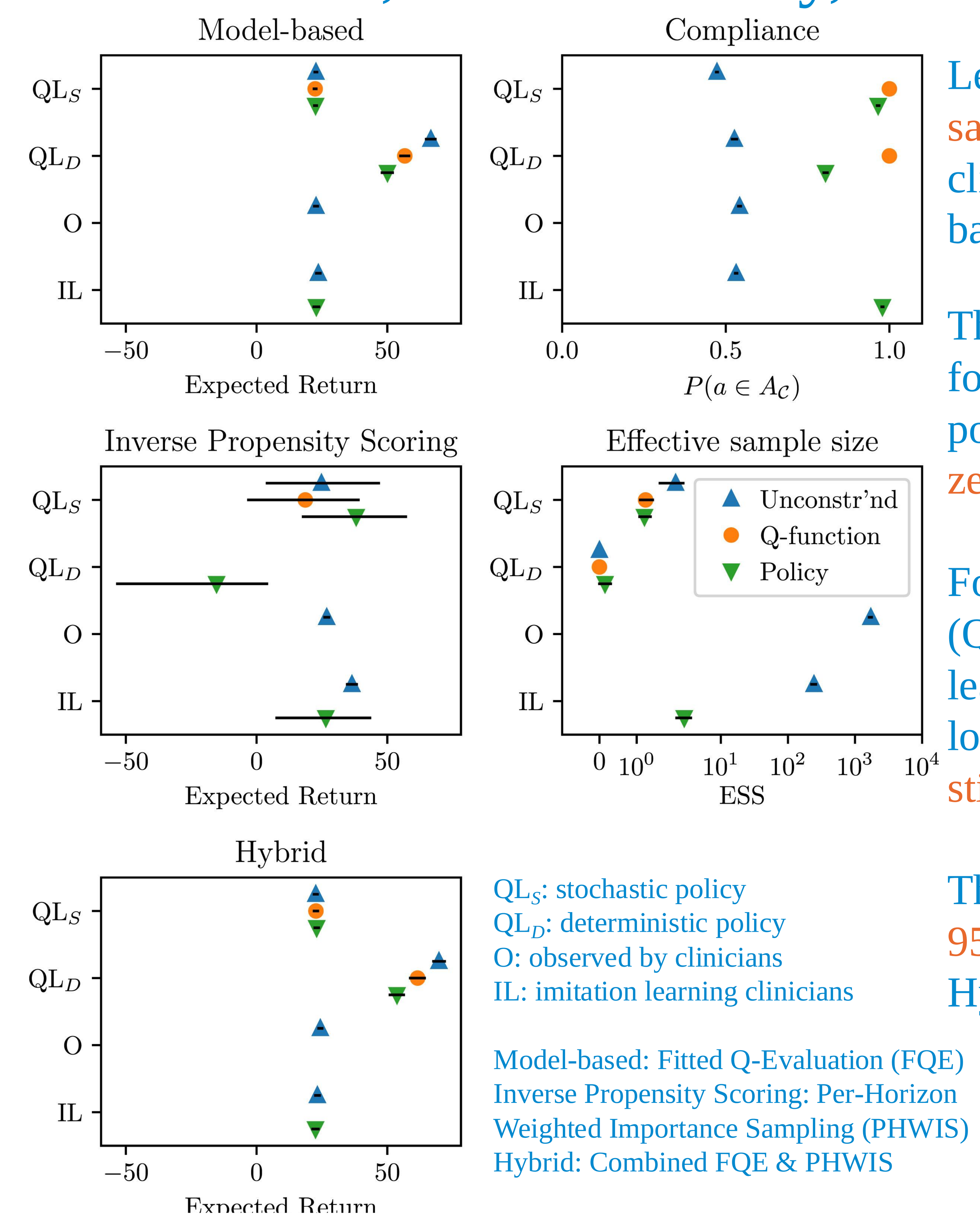
Learned & unconstrained, unsafe policies select more varied actions then those observed in clinicians

This may indicate better adaptation to patient conditions

Learned & constrained, safe policies avoid extreme actions that are not according to guideline

This may result in more trust by clinicians and hence better adoption

Results: Safe, Low Mortality, Small Samples



Learned policies are more safe and better (QL_D) than clinicians in a model-based evaluation

The effective sample size for deterministic safe policies (QL_D) may be zero

For stochastic policies (QL_S) and imitation learning (IL), the ESS looks much better, but is still small

This explain the large 95% CIs for IPS and Hybrid evaluations

Conclusions

Guideline-compliant RL to move RL to closer to clinical practice.

Varied policies that comply with the medical guideline while outperforming clinicians in terms of expected mortality in a model-based evaluation

No benefits in reward shaping: the training data were sufficiently rich for minimizing 90-day mortality.

Amsterdam UMC

VU