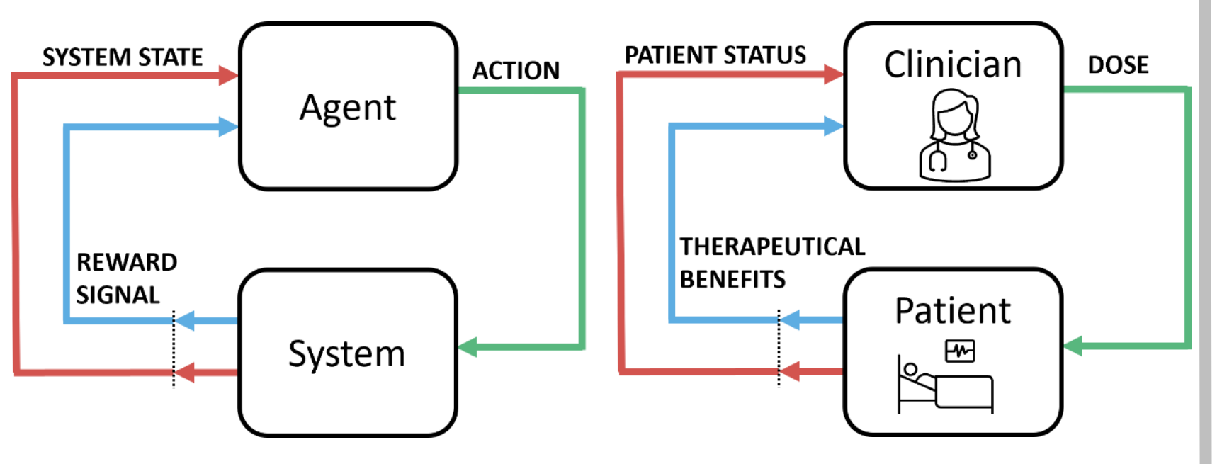
Model-Informed Reinforcement Learning for precision dosing of Carboplatin

Ilaria Gaiffi, Elena Maria Tosca, Paolo Magni

Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

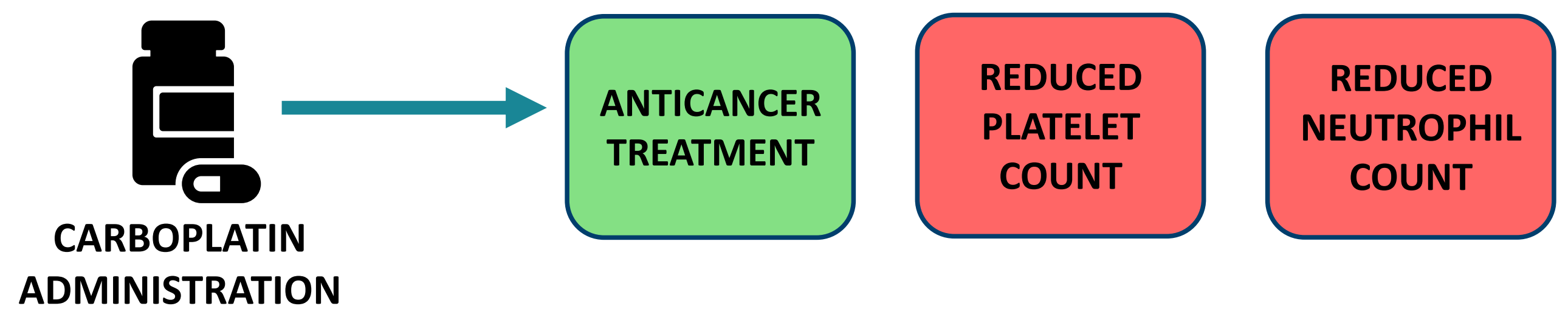
BACKGROUND:

- Precision dosing personalizes drug therapy to maximize efficacy while minimizing toxicity [1,2].
- Model-Informed Precision Dosing (MIPD) relies on PK-PD models and Bayesian estimation approaches to create patient-specific digital twins and simulate the patient therapeutic responses [3, 4, 5].
- Exploring multiple dosing strategies through MIPD may become computationally demanding.
- Reinforcement Learning (RL) is increasingly used to identify optimal adaptive dosing strategies through sequential decision-making algorithms [6, 7].



AIM:

Investigate the potential of a combined MIPD + RL (specifically Q-Learning, QL [8]) framework for the optimization of carboplatin dosing in lung cancer patients.



CARBOPLATIN ADMINISTRATION

ANTICANCER TREATMENT

REDUCED PLATELET COUNT

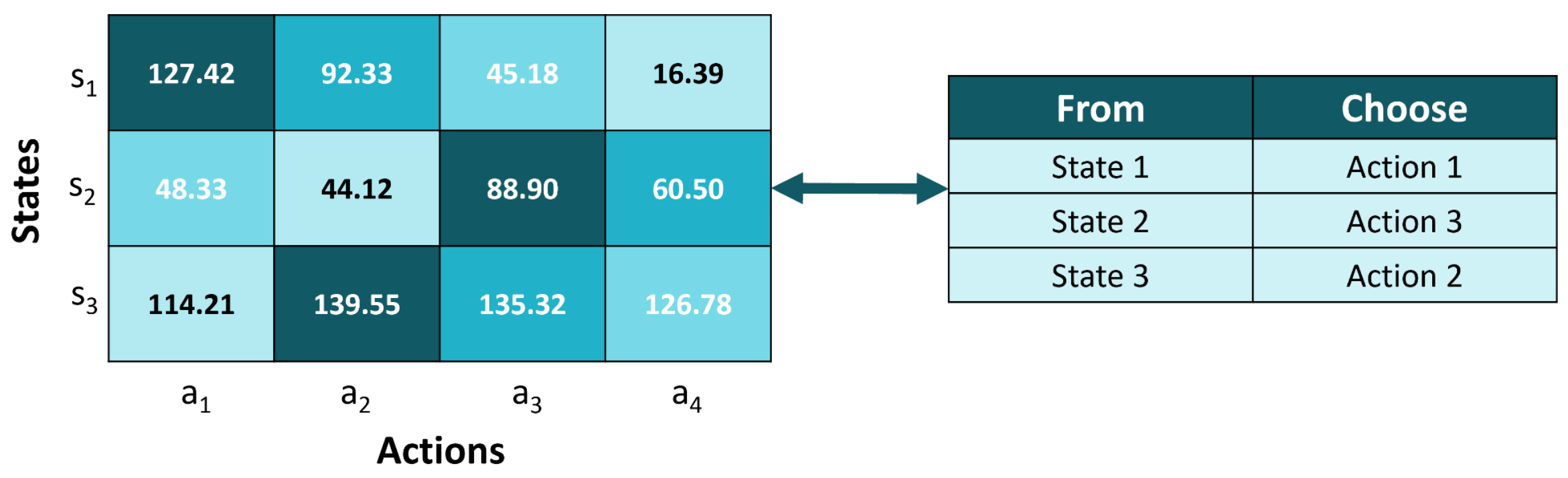
REDUCED NEUTROPHIL COUNT

Therapeutic objective:

Maximize the administered dose while preventing severe myelosuppression of platelets (PLT) and neutrophils (NT).

MATERIALS AND METHODS:

Q-Learning: RL algorithm based on a tabular approximation of the Q-function, called Q-matrix

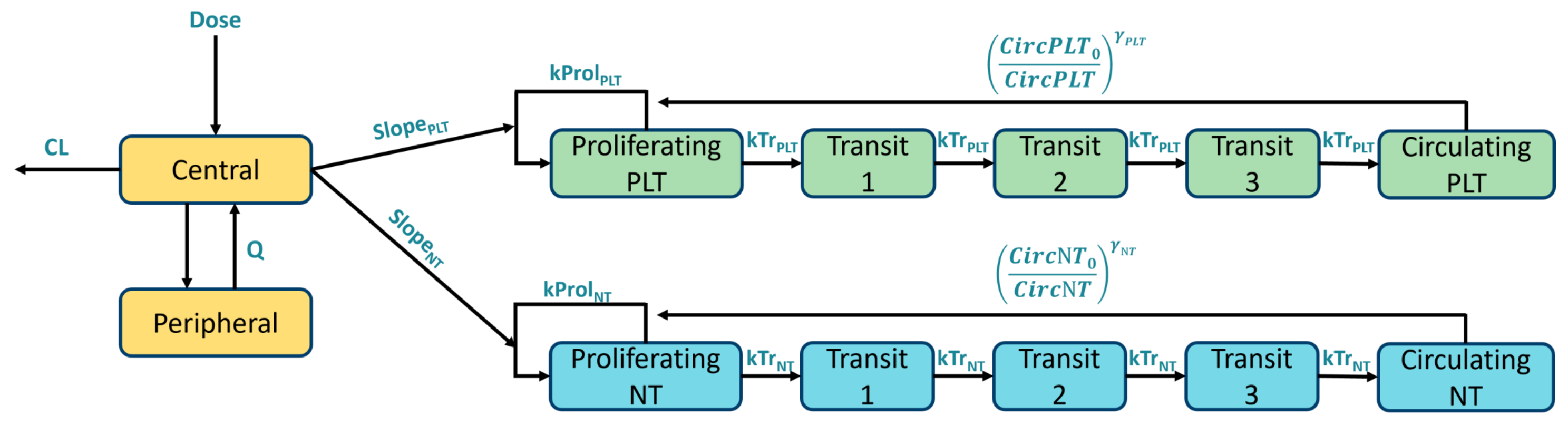


States	a_1	a_2	a_3	a_4
s_1	127.42	92.33	45.18	16.39
s_2	48.33	44.12	88.90	60.50
s_3	114.21	139.55	135.32	126.78

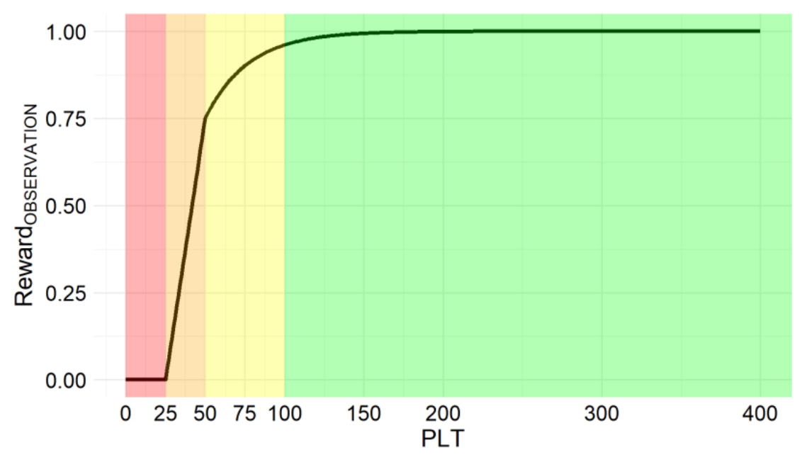
Actions

From	Choose
State 1	Action 1
State 2	Action 3
State 3	Action 2

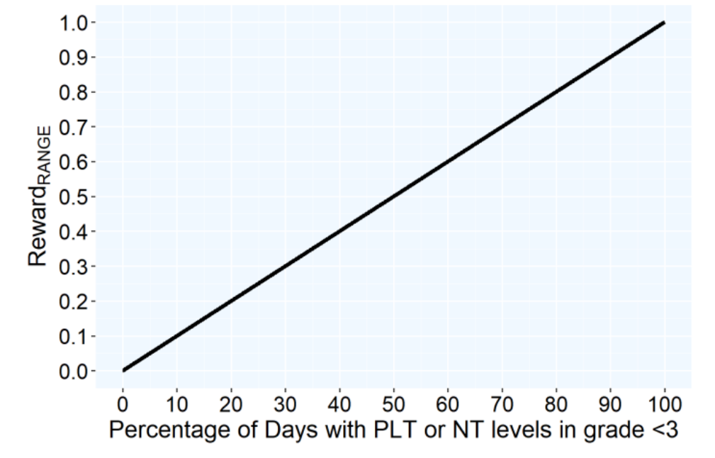
- **PK-PD model** of carboplatin-induced myelosuppression in platelets (PLT) and neutrophils (NT) ⟶ already available [9].



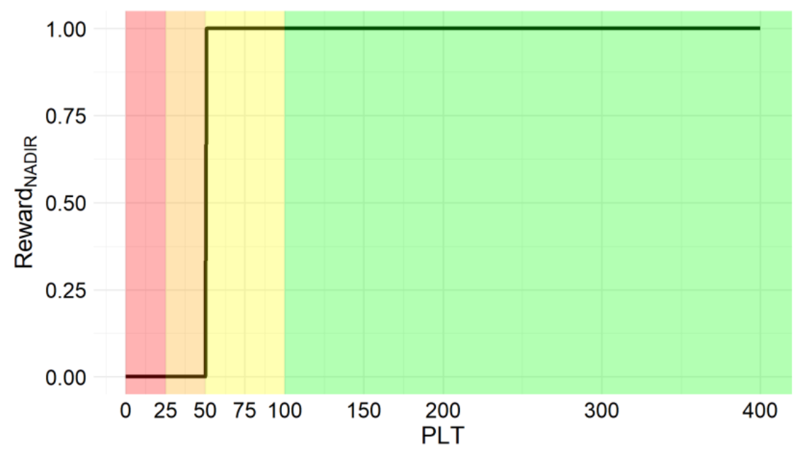
- **Actions** ⟶ expressed as fractions of the standard individual carboplatin dose: [0, 0.5, 0.8, 0.9, 1, 1.1, 1.2, 1.5] x standard dose.
- **System States** ⟶ maximum toxicity level observed in PLT and NT at the end of the cycle, along with the previously administered dose.
- **Reward function** ⟶ defined by:
  - value of PLT/NT at the end of the cycle



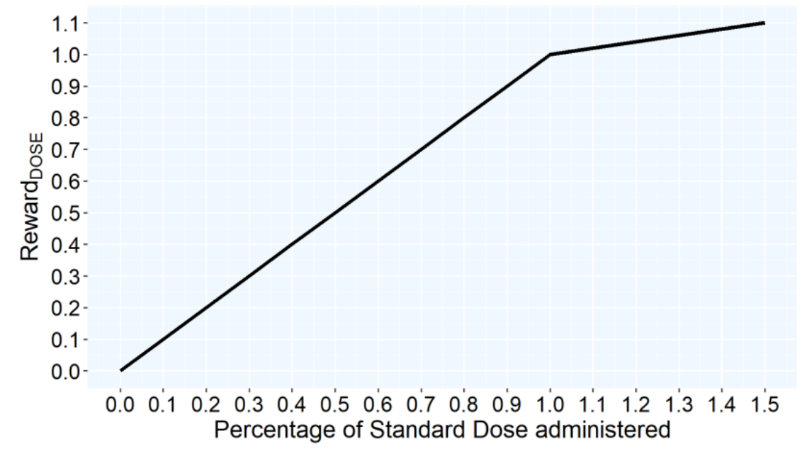
  - percentage of days without PLT/NT toxicity



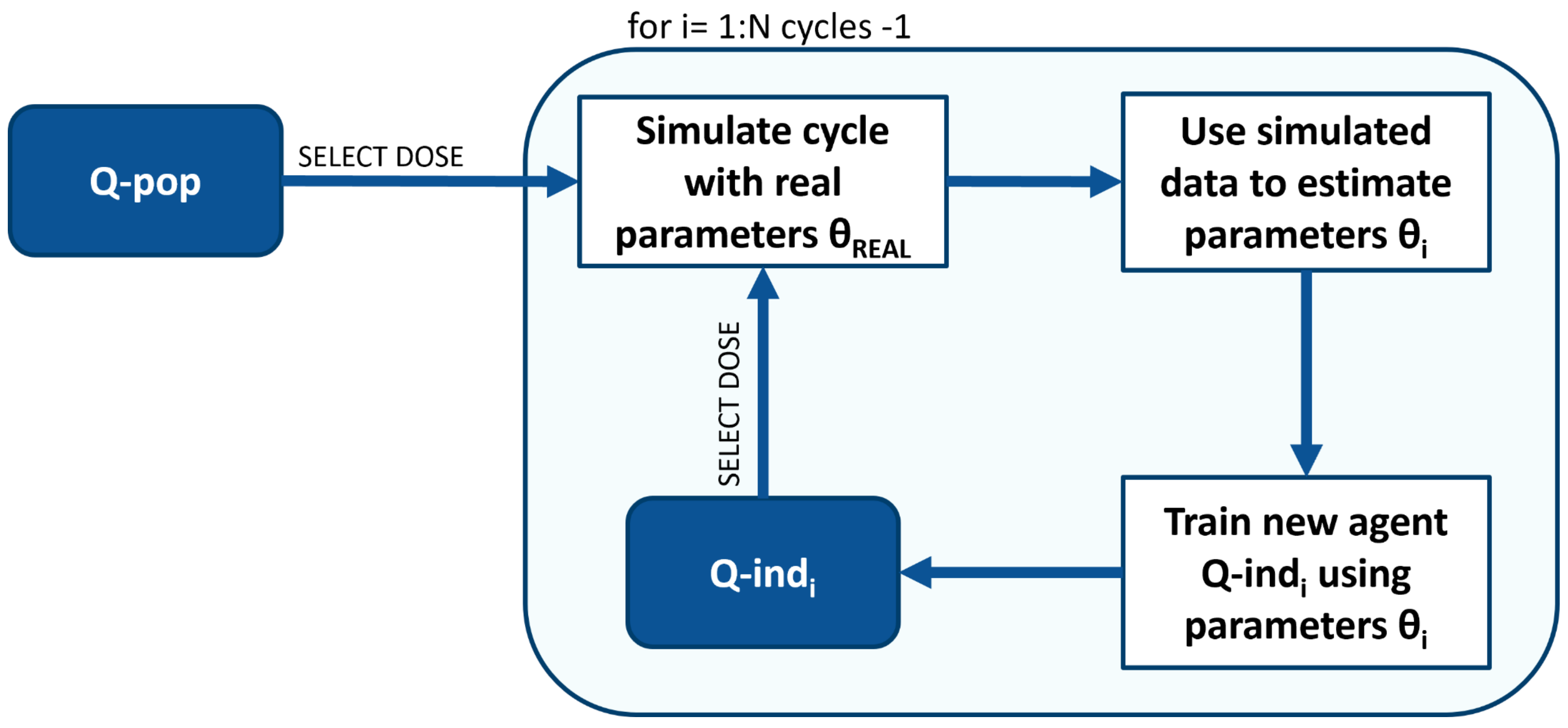
  - nadir value of PLT/NT



  - previously administered dose



Bayesian framework: implemented to simulate real clinical conditions, where individual parameters are not available at the beginning of the treatment.



for i= 1:N cycles -1

Q-pop

SELECT DOSE

Simulate cycle with real parameters θ_{REAL}

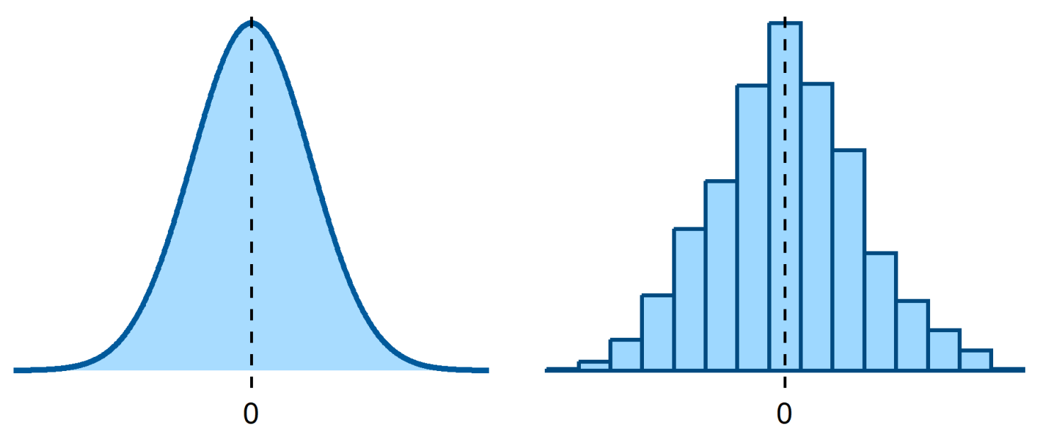
Use simulated data to estimate parameters θ_i

Train new agent $Q\text{-}ind_i$ using parameters θ_i

$Q\text{-}ind_i$

SELECT DOSE

To estimate individual parameters: MCMC algorithm where in the first cycle, the population prior was used, while in the subsequent cycles, an empirical prior based on the samples obtained in the previous cycle was adopted.



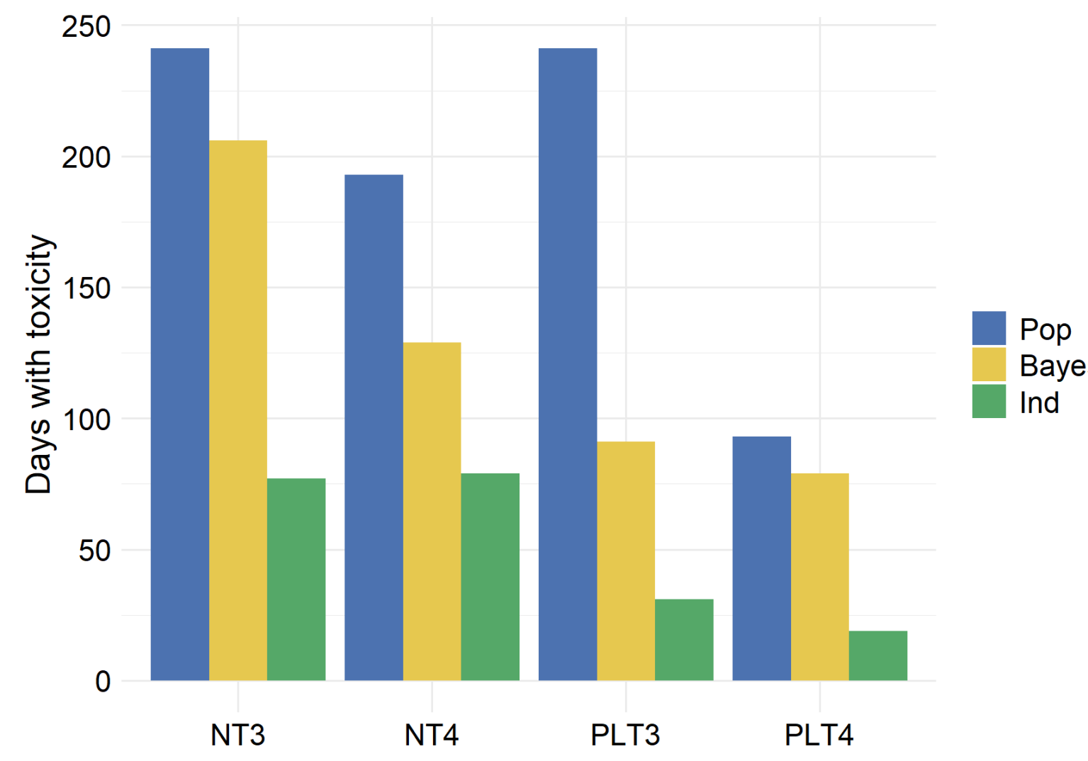
Prior cycle = 1

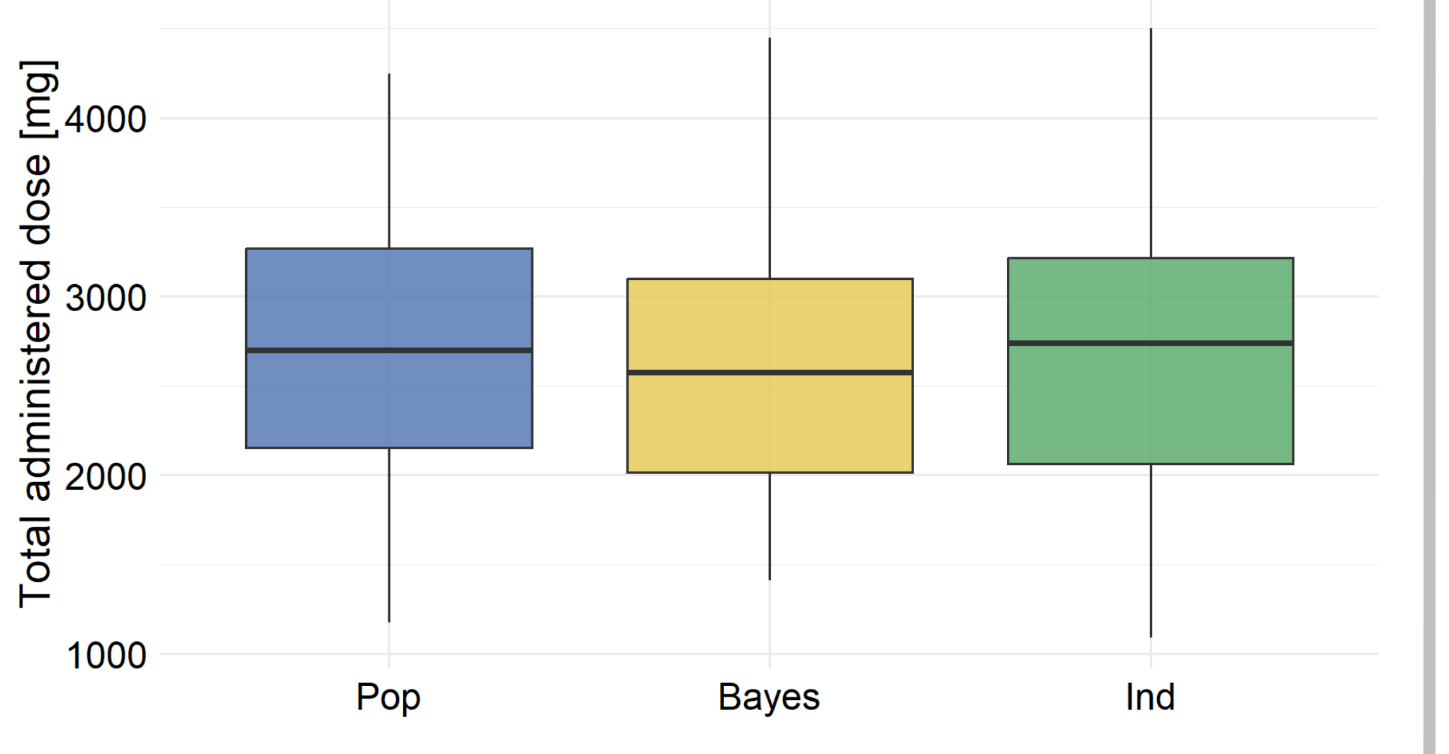
Prior cycle >1

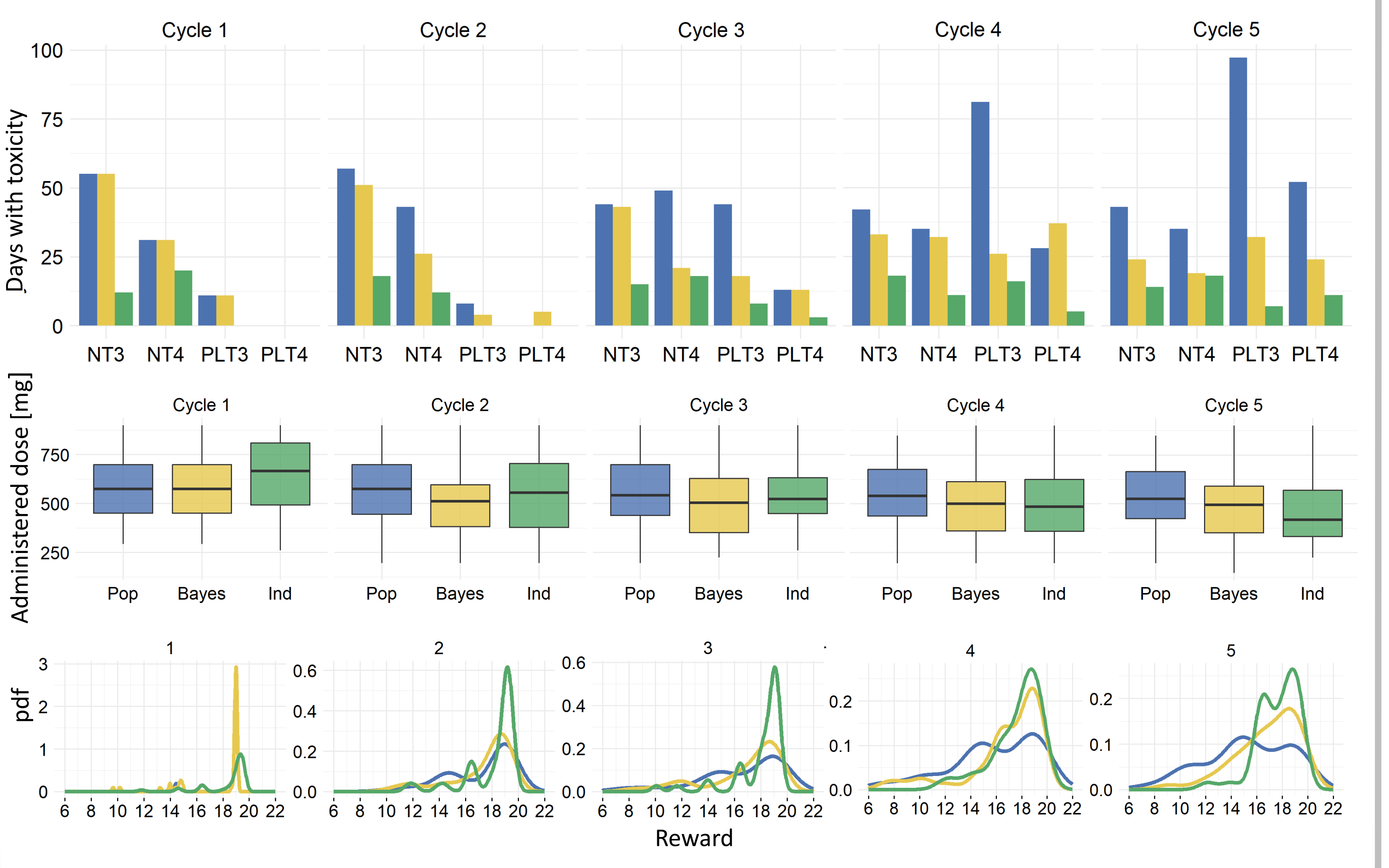
RESULTS:

To test the proposed framework, a population of 41 virtual patients was created. For each patient, outcomes were evaluated using three different Q-learning agents:

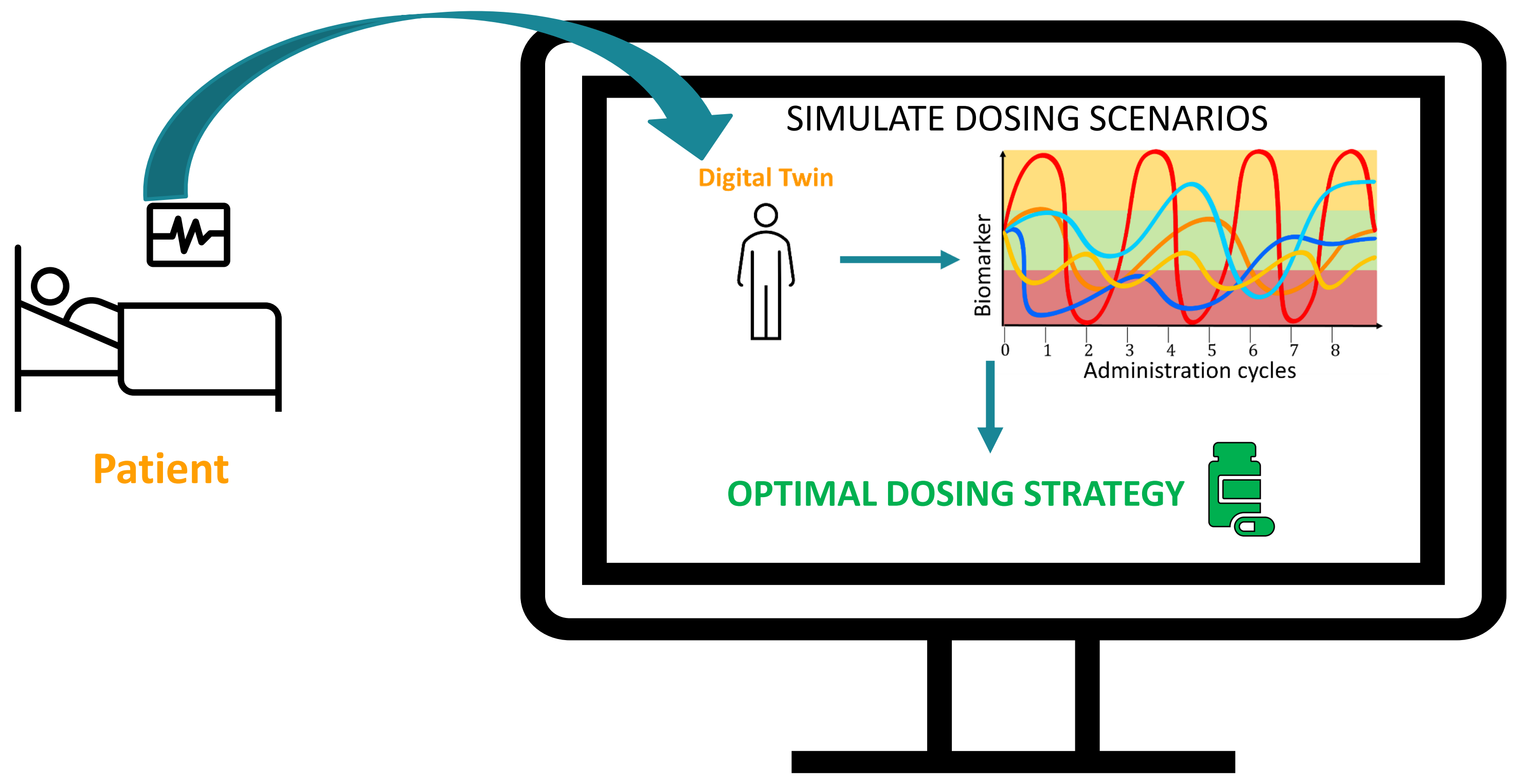
- **Population agent:** trained on a large external virtual population. It represents a general protocol intended to work across the entire population, but it is not optimized for any specific patient.
- **Bayesian agent:** trained using the proposed framework. It aims to replicate a realistic clinical scenario in which patient-specific parameters are not known a priori but are progressively inferred.
- **Individual agent:** patient-specific agent trained directly using the true parameters of each patient from the beginning.







Clinician oriented GUI: a graphical user interface was developed through which clinicians can simulate different scenarios and observe their potential effects on the patient. From the application, it is possible to input patient data, obtain a digital twin by identifying individual parameters, and simulate both the recommendation provided by the QL algorithm and any alternative decision made by the clinician. The GUI was implemented in Python using the library PyQt6.



Patient

SIMULATE DOSING SCENARIOS

Digital Twin

OPTIMAL DOSING STRATEGY

CONCLUSIONS: The proposed framework shows that the Bayesian agent outperforms the population protocol, but, as expected, it performs worse than the individual agent, which represents an unreachable benchmark. This supports the potential clinical application of the framework, further facilitated by the development of a GUI that enables easy use by clinicians.

[1] Peck RW. Precision Dosing: An Industry Perspective. *Clin Pharmacol Ther*. 2021;109(1):47-50.

[2] Polasek TM, Shakib S, Rostami-Hodjegan A. Precision dosing in clinical medicine: present and future. *Expert Rev Clin Pharmacol*. 2018;11(8):743-746.

[3] Maier C, de Wiljes J, Hartung N, Kloft C, Huisinga W. A continued learning approach for model-informed precision dosing: Updating models in clinical practice. *CPT Pharmacometrics Syst Pharmacol*. 2022;11(2):185-198.

[4] Minichmayr IK, Dreesen E, Centanni M, et al. Model-informed precision dosing: State of the art and future perspectives. *Adv Drug Deliv Rev*. 2024;215:115421.

[5] Tosca EM, De Carlo A, Ronchi D, Magni P. Model-Informed Reinforcement Learning for Enabling Precision Dosing Via Adaptive Dosing. *Clin Pharmacol Ther*. 2024;116(3):619-636.

[6] Ribba B. Reinforcement learning as an innovative model-based approach: Examples from precision dosing, digital health and computational psychiatry. *Front Pharmacol*. 2023;13:1094281. Published 2023 Feb 17.

[7] Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT press, 2018.

[8] Clifton, Jesse and Laber, Eric B., Q-Learning: Theory and Applications (March 2020). Annual Review of Statistics and Its Application, Vol. 7, Issue 1, pp. 279-301, 2020.

[9] A. De Carlo, M. Crul, T. Schutte, L. van Zuylen, I. Bahce, Chi Fong Loo, D. van Valkengoed, H. Huls, E. M. Tosca, P. Magni, and I. Bartelink. Poppk-pd modelling of carboplatin-induced myelosuppression to support therapeutic drug monitoring and dose individualization in cancer patients using electronic health record data. PAGE 32.