UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS
Department of Applied Mathematics

Brydon Eastman (b2eastma@uwaterloo.ca)
Michelle Przedborski
Mohammad Kohandel

Template: Felix Breur, 2010

# Reinforcement learning derived chemotherapeutic schedules for robust patient-specific therapy given unknown patient response parameters

**Brydon Eastman**
University of Waterloo

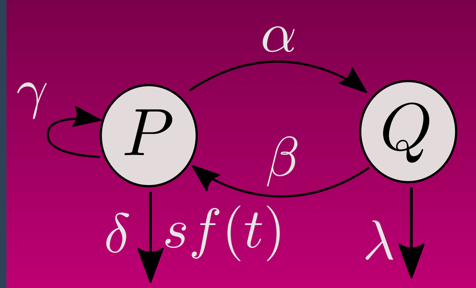
## Motivation

Personalized therapy often requires identifiability of hard to measure parameters.

## Research Question

Can we personalize therapy without directly measuring these parameters?

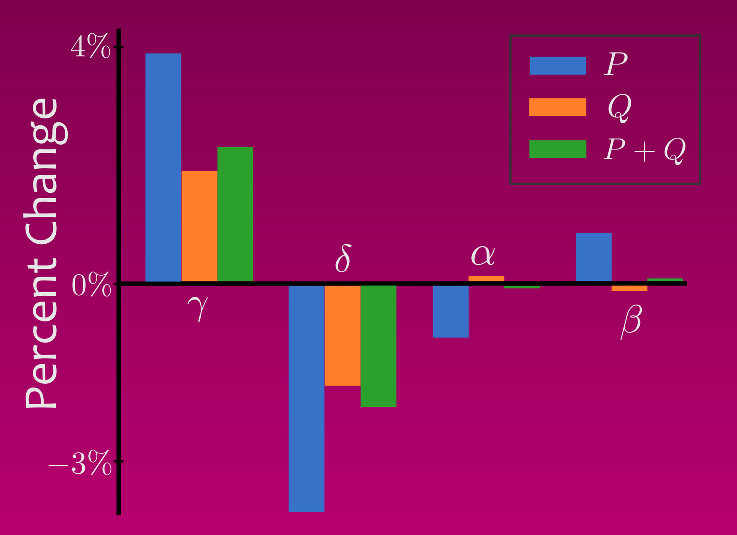
## Tumor Growth Inhibition Model




Where $P$ and $Q$ are proliferative and quiescent (resp.) [1]

$$\dot{P} = (\gamma - \delta - \alpha - s\,f(t))\,P + \beta\,Q$$
$$\dot{Q} = \alpha\,P - (\lambda + \beta)\,Q$$

## Local Sensitivity Analysis




Perturbing parameters above the mean by 1% produces large changes in the predicted trajectories.

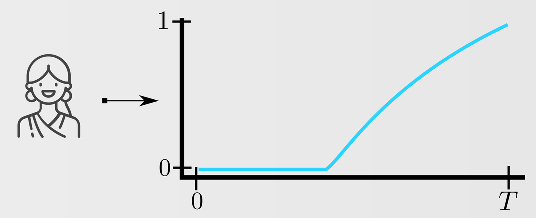
Sensitivity in the parameters suggests need for accurate identification of patient specific parameters.

We can define an **optimal treatment** by optimizing an objective functional

$$J(f) = \int_0^T \underbrace{P_{\rm bm}(t) + Q_{\rm bm}(t)}_{\text{Preserve Healthy Cells}} - \underbrace{\frac{b}{2}(1 - f(t))^2}_{\text{Deliver Drug}}\,\mathrm{d}t$$

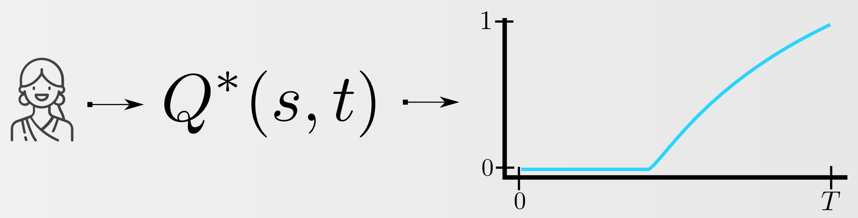
Assuming we know the patient specific parameters, we can find the optimum via optimal control theory [2] or by reinforcement learning (or other methods...)

In optimal control theory, we derive the treatment directly




In RL, we approximate a value function from which we derive the treatment schedule [3]

$\rightarrow Q^*(s,t) \rightarrow$




Suppose we only know some nominal patient parameters

Which we use to train our learner

$$Q^*(s,t) \equiv Q^*(s,t;\ \cdot\ )$$

We want to choose the form of $\mathcal{S}$ such that

$$Q^*(s,t;\ \cdot\ )$$

Produces schedules near the optimum for all virtual patients

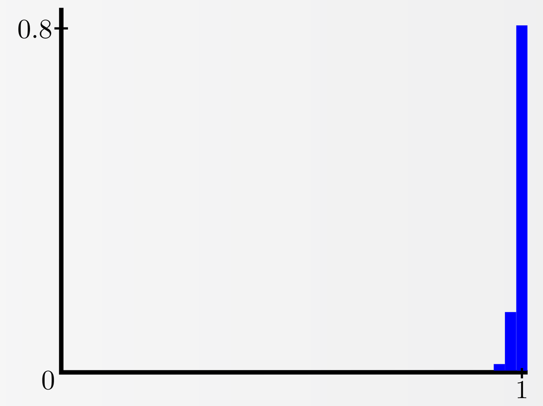
Our form of $\mathcal{S}$ should

- Be readily measurable
- Depend on personal parameters

We use a window of daily relative bone marrow densities:

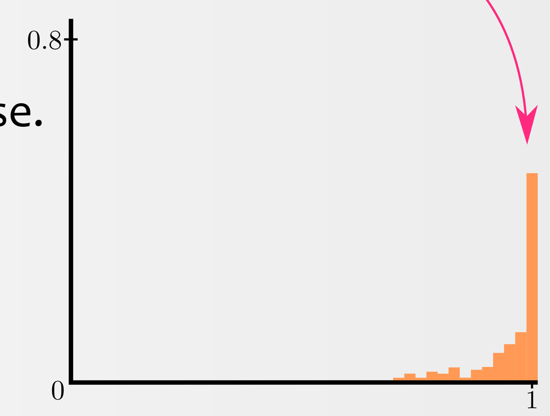
$$\frac{1}{P_{\rm bm}(0) + Q_{\rm bm}(0)} \begin{bmatrix} P_{\rm bm}(\lfloor t \rfloor) + Q_{\rm bm}(\lfloor t \rfloor) \\ P_{\rm bm}(\lfloor t-1 \rfloor) + Q_{\rm bm}(\lfloor t-1 \rfloor) \\ \vdots \\ P_{\rm bm}(\lfloor t-9 \rfloor) + Q_{\rm bm}(\lfloor t-9 \rfloor) \end{bmatrix}$$

The reinforcement learner produced scores clustered around the maximum.




Bin width 0.025

While the optimal controller scores were much more diffuse.

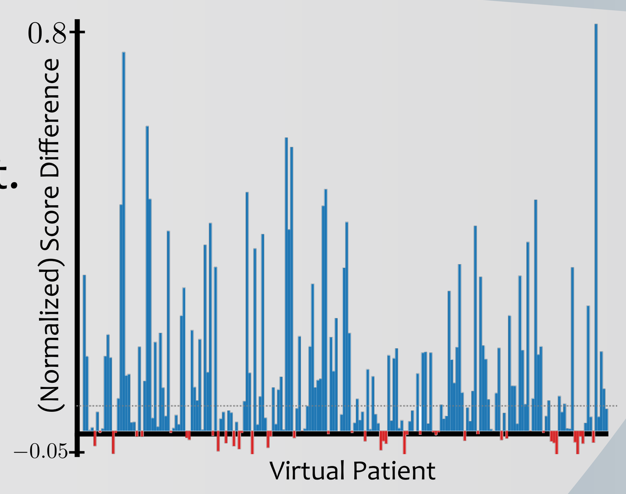



As we perturbed further from the mean, the OC scores became more diffuse still while the RL scores remained clustered between 0.925 and 1.000

Each bar represents a virtual patient.

The height is the difference between the RL score, obtained from applying the schedule learned from $Q^*(s,t;\ \cdot\ )$

and the OC score, obtained from applying the nominal control.




The RL derived schedules remain nearly optimal as perturbation strength increases

## Perturbed Virtual Patients

We prepared 3 sets of 200 virtual patients at 15 %, 20 %, and 25% perturbation strength level by perturbing the 5 nominal patient [1] parameter values by latin hypercube sampling a scaling factor up to the strength.

We treated these values as unknowns during the training and testing processeses.

Only non-dimensionalizing the objective functional by the theoretical maximum for display purposes.

## Other Personalizations

We also considered OC based personalizations

These were obtained via various interpolation stratgies

In all situations we solve the OC problem for 50 training patients and interpolate based on bone-marrow state and time

As in the other case, the RL remains robust as perturbation strength increases.

The difference between the RL and the OC scores can be as high as 0.35 or as low as -0.05. (Again where the swings favour the RL)

## (Some) Works Cited

1. John Carl Panetta. A mathematical model of breast and ovarian cancer treated with paclitaxel. Mathematical biosciences, 146(2):89–113, 1997

2. John Carl Panetta and K Renee Fister. Optimal control applied to cell-cycle-specific cancer chemotherapy. SIAM Journal on Applied Mathematics, 60(3):1059–1072, 2000.

3. Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI conference on artificial intelligence, volume 30, 2016.

4. Brydon Eastman, Michelle Przedborski, and Mohammad Kohandel. Reinforcement learning derived chemoterapeutic schedules for robust patient-specific therapy. bioRxiv, 2021.