HARVARD
School of Engineering
and Applied Sciences

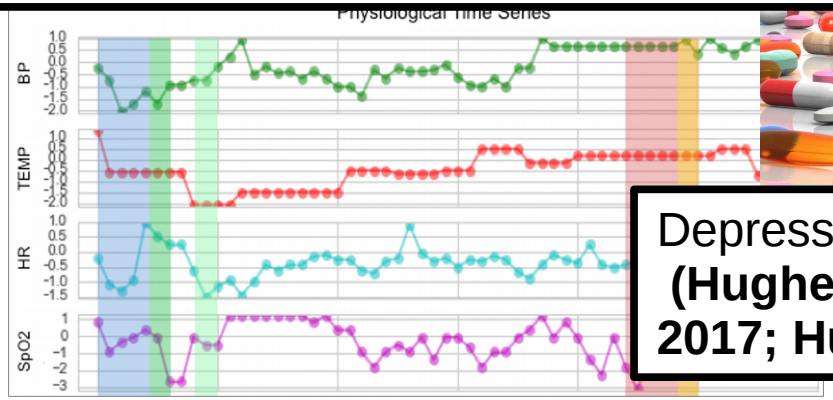
Reinforcement Learning for Healthcare

Finale Doshi-Velez
Harvard University

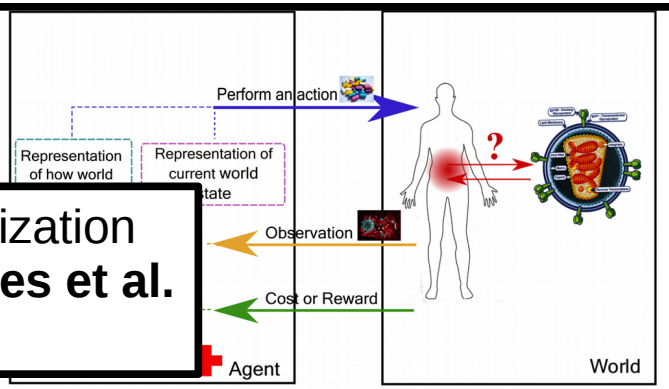
Collaborators: Sonali Parbhoo, Maurizio Zazzi, Volker Roth, Xuefeng Peng, David Wihl, Yi Ding, Omer Gottesman, Liwei Lehman, Matthieu Komorowski, Aldo Faisal, David Sontag, Fredrik Johansson, Leo Celi, Aniruddh Raghu, Yao Liu, Emma Brunskill, and the CS282 2017 Course

Our Lab: ML Towards Effective, Interpretable Health Interventions

Predicting and Optimizing Interventions in ICU (**Wu et al. 2015; Ghassemi et al. 2017; Peng 2018; Raghu 2018; Gottesman 2018**)



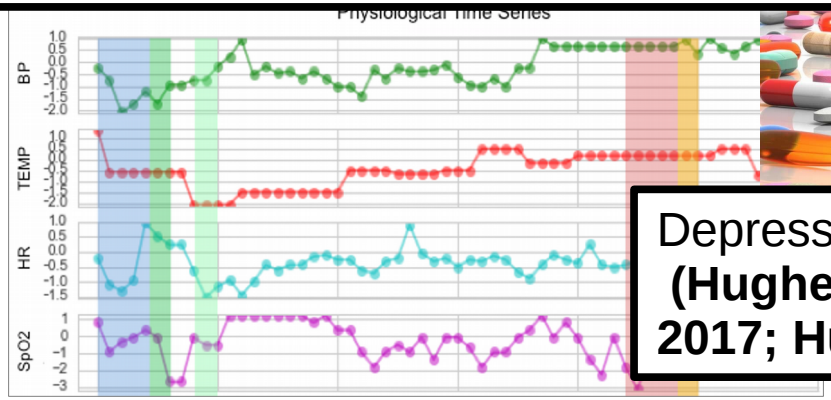
HIV Management Optimization **(Parbhoo et al., 2017, Parbhoo et al. 2018)**



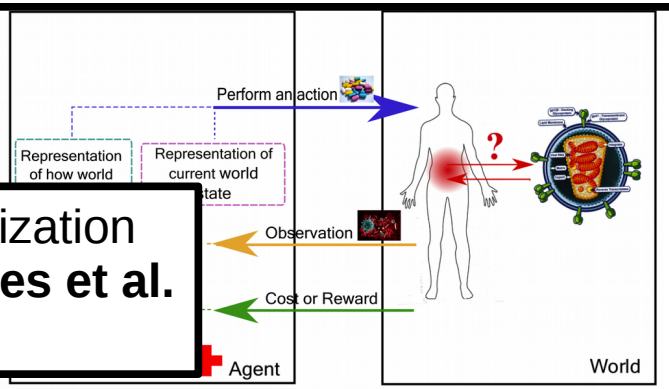
Depression Treatment Optimization **(Hughes et al., 2016; Hughes et al. 2017; Hughes et al. 2018)**

Our Lab: ML Towards Effective, Interpretable Health Interventions

Predicting and Optimizing Interventions in ICU (**Wu et al. 2015; Ghassemi et al. 2017; Peng 2018; Raghu 2018; Gottesman 2018**)



HIV Management Optimization **(Parbhoo et al., 2017, Parbhoo et al. 2018)**

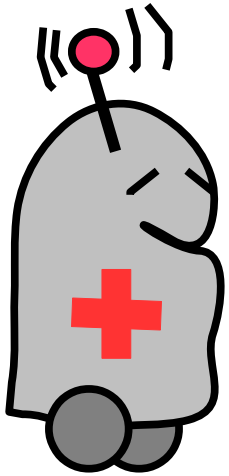


Depression Treatment Optimization **(Hughes et al., 2016; Hughes et al. 2017; Hughes et al. 2018)**

Today: How can reinforcement learning help solve problems in healthcare?

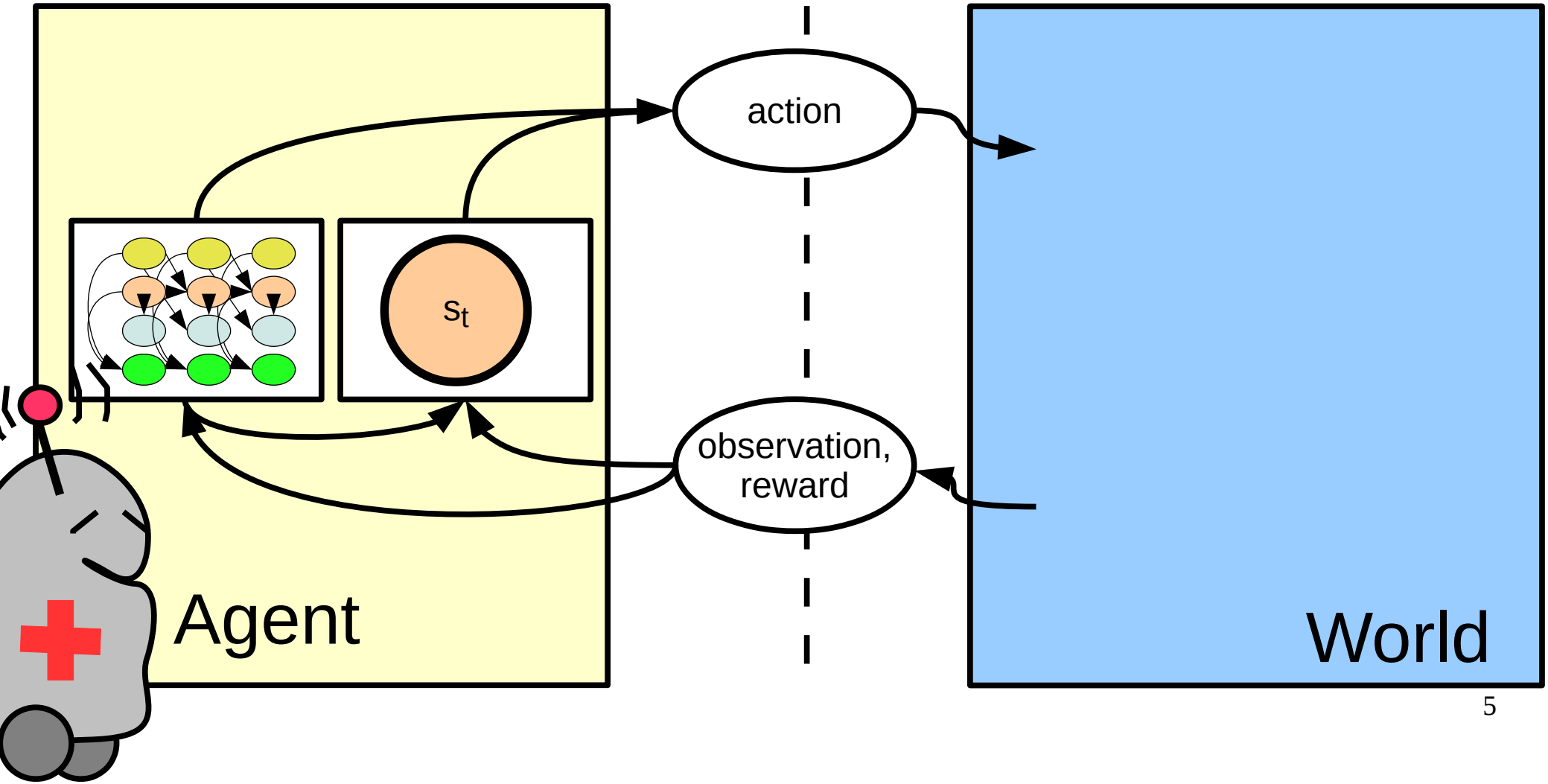
How reinforcement learning can help solve problems in healthcare

- Some health problems are effectively “one-off” (Does this image have a tumor? Will this depression med work?)
- Others involve reasoning about a series of decisions
  - Choosing meds for HIV (prevent drug resistance)
  - Choosing ICU interventions (short vs. longterm effects)



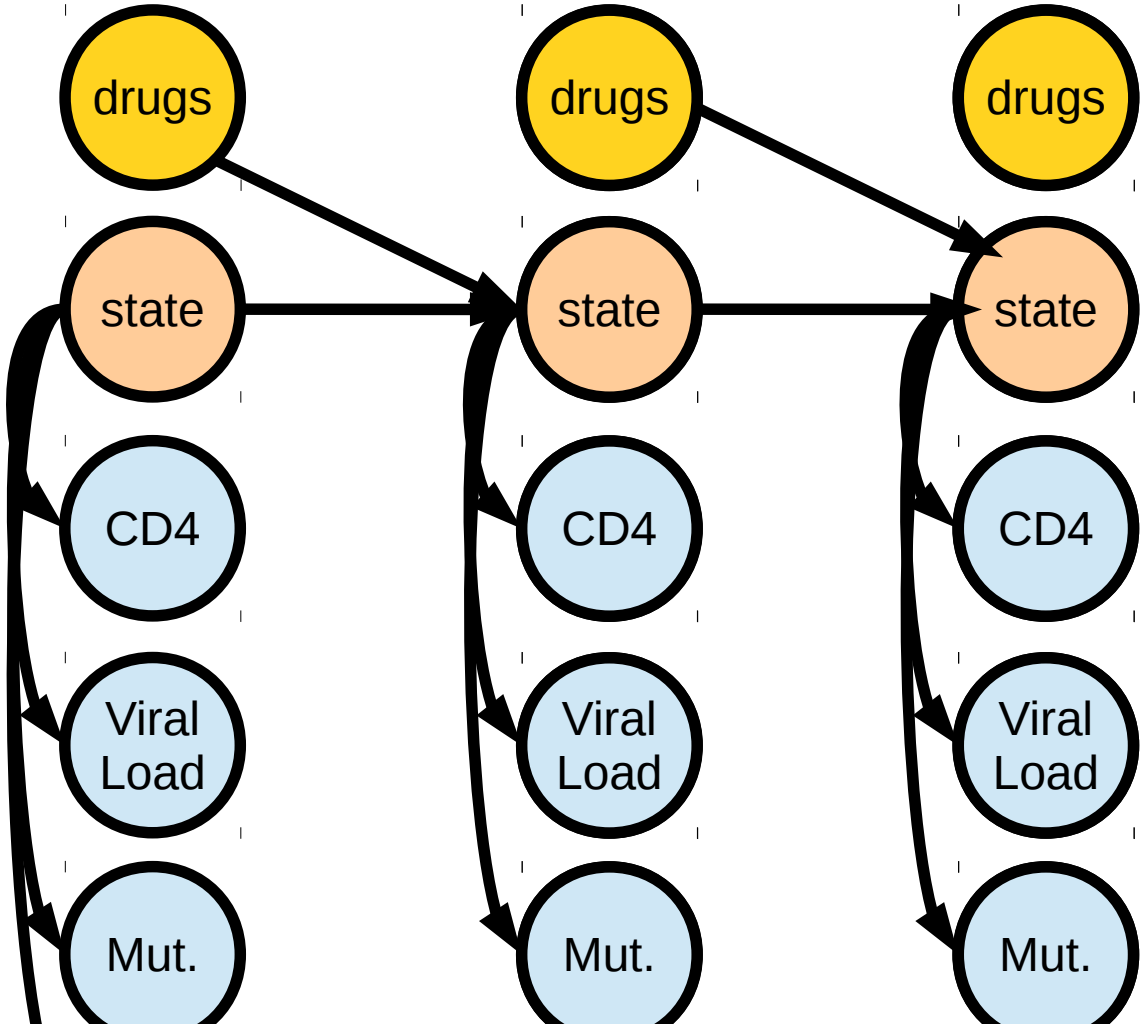
RL can help us think about sequential decisions!

Reinforcement Learning: Formalizing the Problem



Model-based RL for HIV Management

Solves the long-term problem (e.g. Ernst 2005; Parbhoo 2014; Marivate 2015), often in simulation/simplified settings.



Rewards:

If $V_t > 40$:

$r_t = -.7 \log V_t + .6 \log T_t - .2|M_t|$

Else:

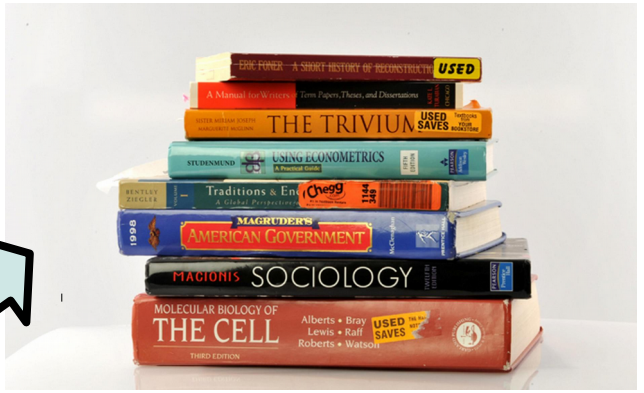
$r_t = 5 + .6 \log T_t - .2|M_t|$

Does it work??

Several works on model-based optimization of HIV treatment (e.g. Ernst 2005; Parbhoo 2014; Marivate 2015), none state of the art...

Drawback of model-based approaches

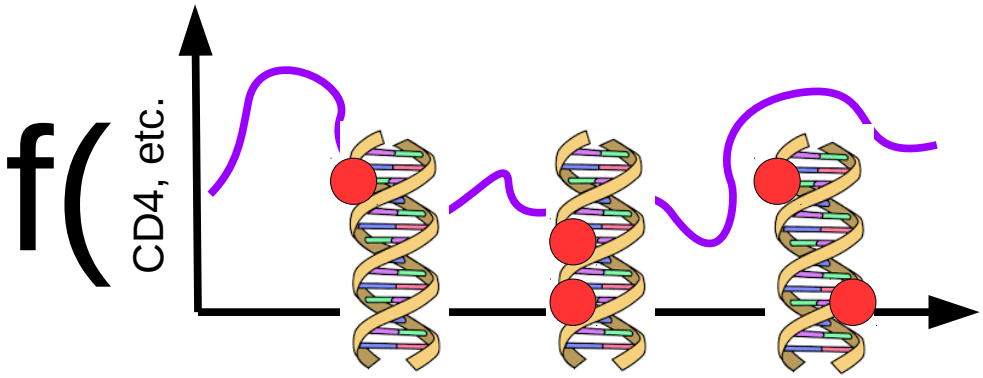
?

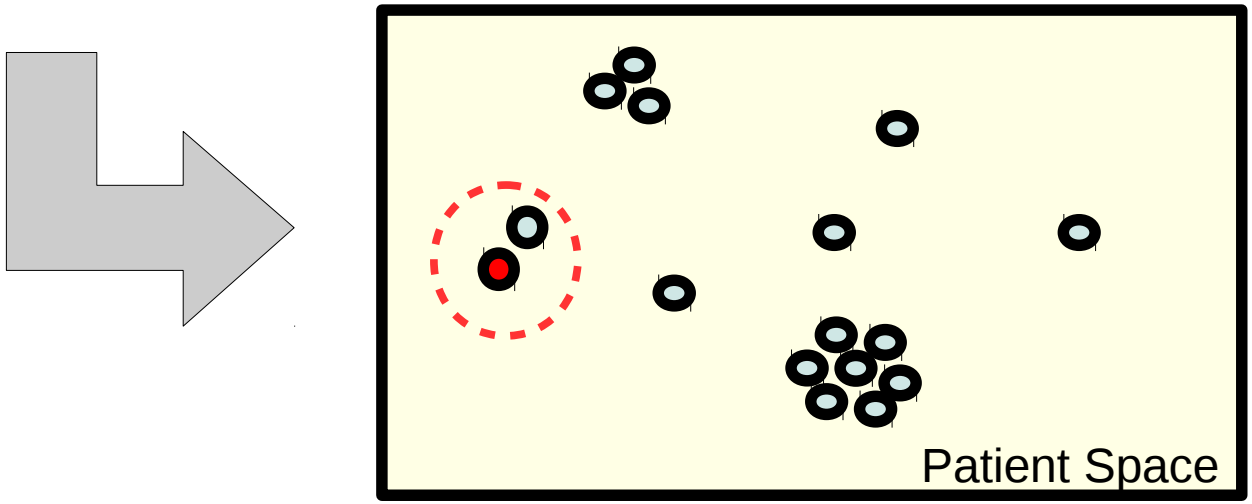


http://78.media.tumblr.com/76b588fd4c4cc02615de4349d179ffc4/tumblr_nyk7v2dech1qcl9ujo1_400.gif
https://agents.allstate.com/images/state/KS.jpg

Alternative: Kernel-based Predictions

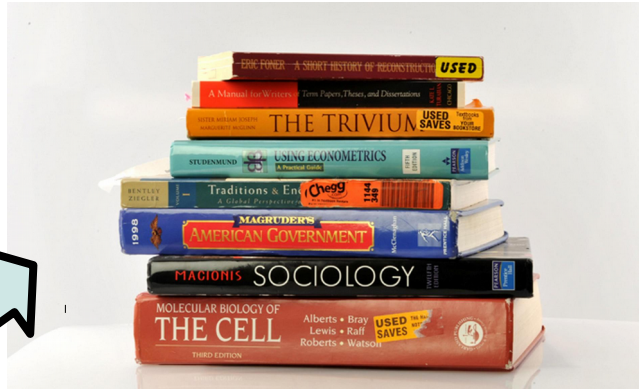
Use the full patient history to predict immediate outcomes (e.g. Bogojeska 2012), but often ignore long term effects.





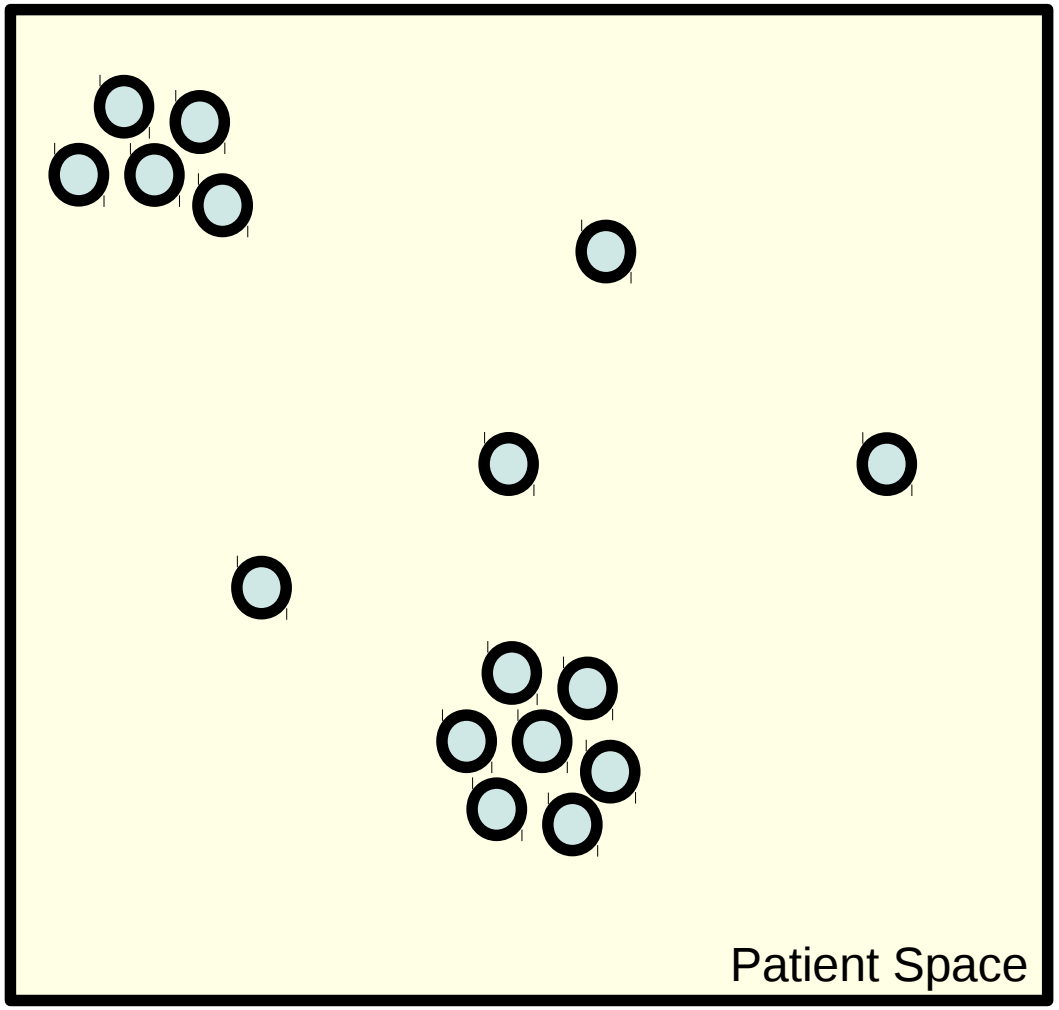
Drawback of neighbor-based approaches

?



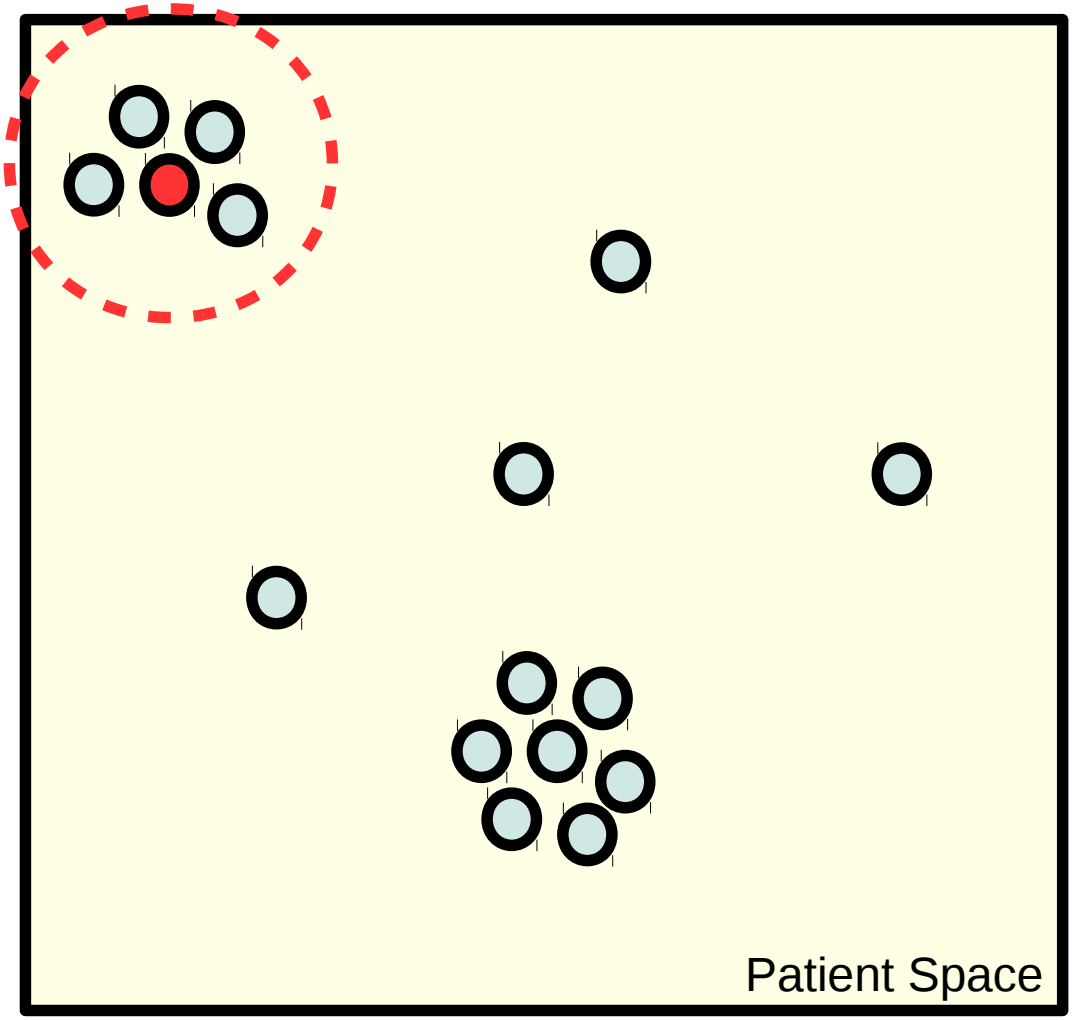
https://i.pinimg.com/736x/c3/82/34/c3823490400661372003eee39beb0356--dorothy-gale-wizard-of-oz.jpg
http://78.media.tumblr.com/76b588fd4c4cc02615de4349d179ffc4/tumblr_nyk7v2dech1qcl9ujo1_400.gif
https://agents.allstate.com/images/state/KS.jpg

Our insight: Models and kernels have complementary strengths!

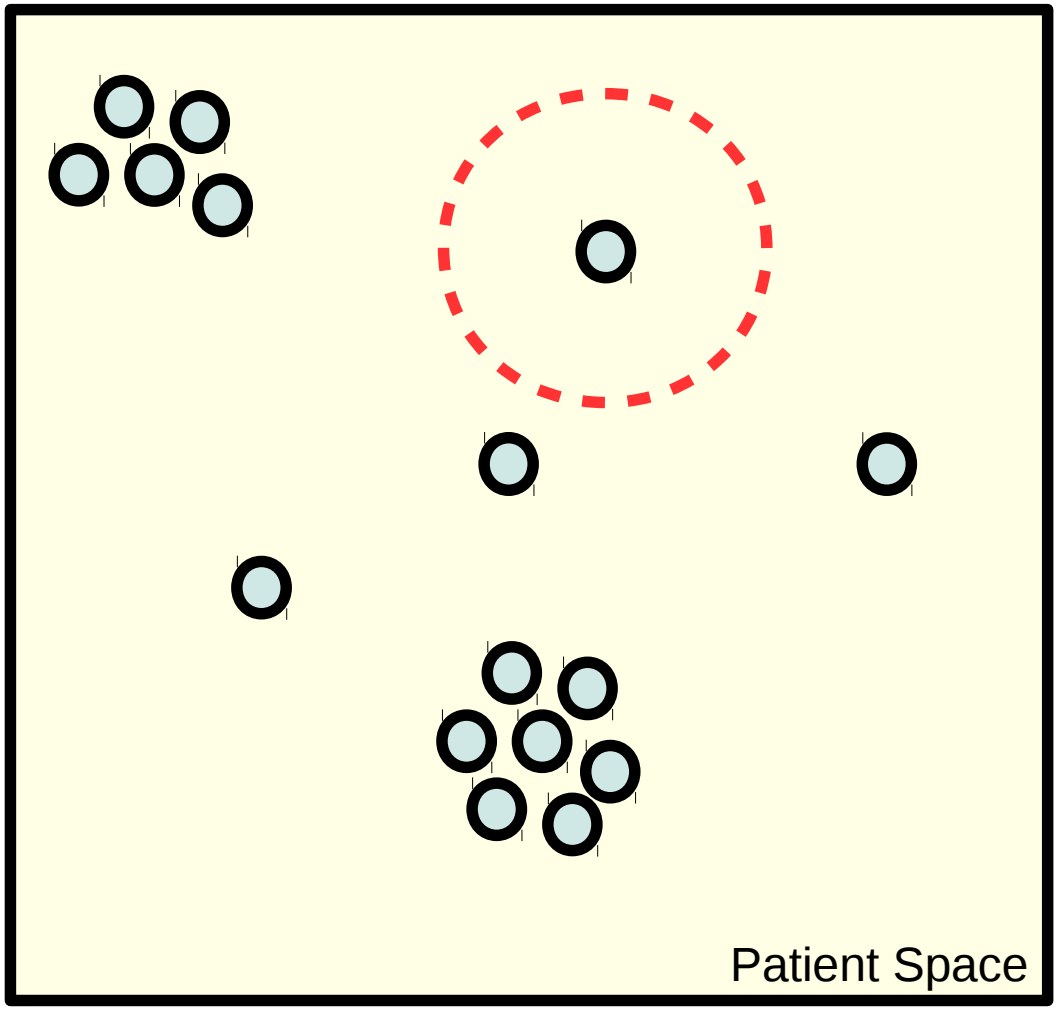


Our insight: Models and kernels have complementary strengths!

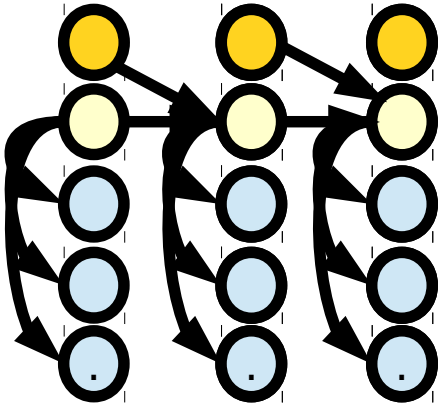
Patients in clusters may be best modeled by their neighbors



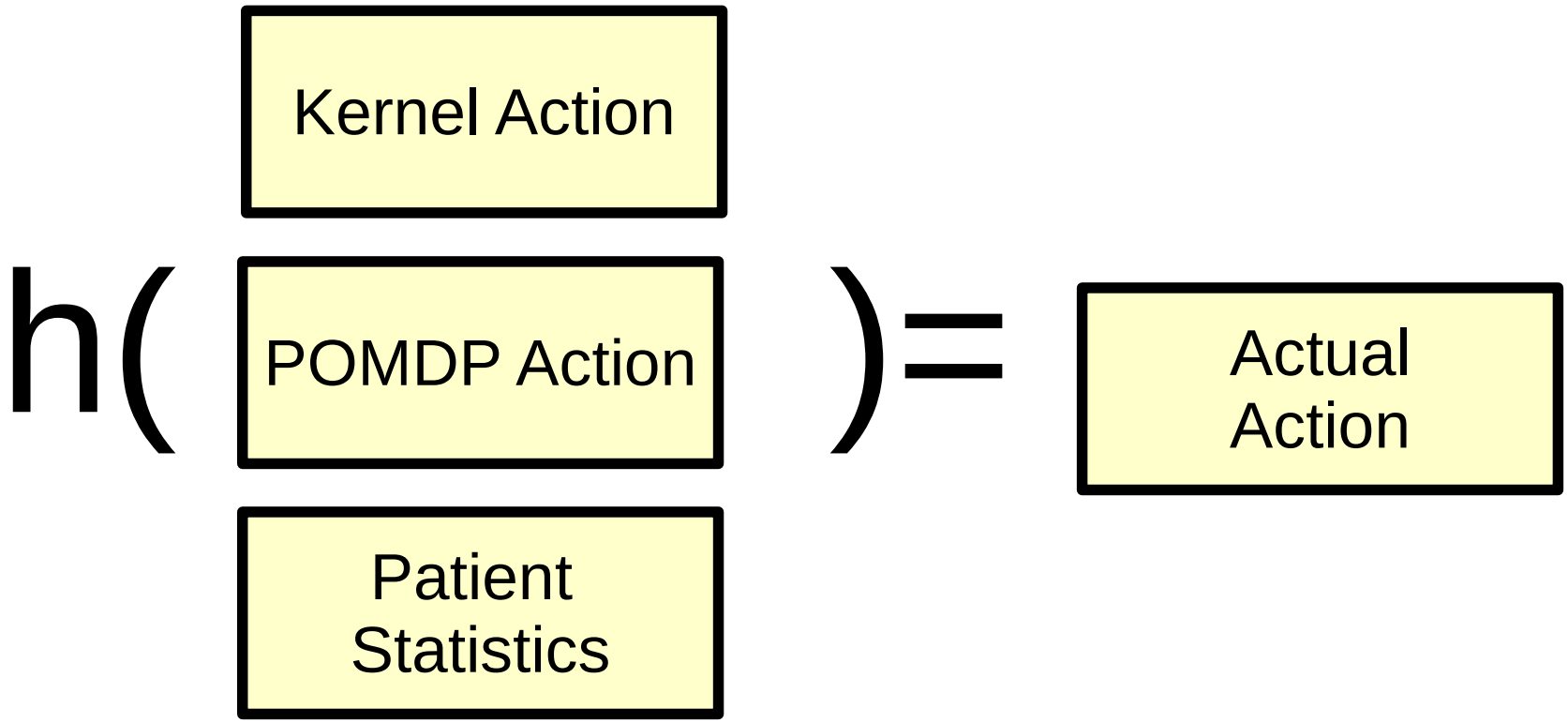
Our insight: Models and kernels have complementary strengths!



Patients without neighbors may be better modeled with a model



Key Idea: Combine!



Does it work??

- 32,960 patients from EU Resist Database; hold out 3,000 for testing.
- Observations: CD4s, viral loads, mutations
- Actions: 312 common drug combinations (from 20 drugs)

Approach	DR Reward
Random Policy	-7.31 ± 3.72
Neighbor Policy	9.35 ± 2.61
Model-Based Policy	3.37 ± 2.15
Policy-Mixture Policy	11.52 ± 1.31

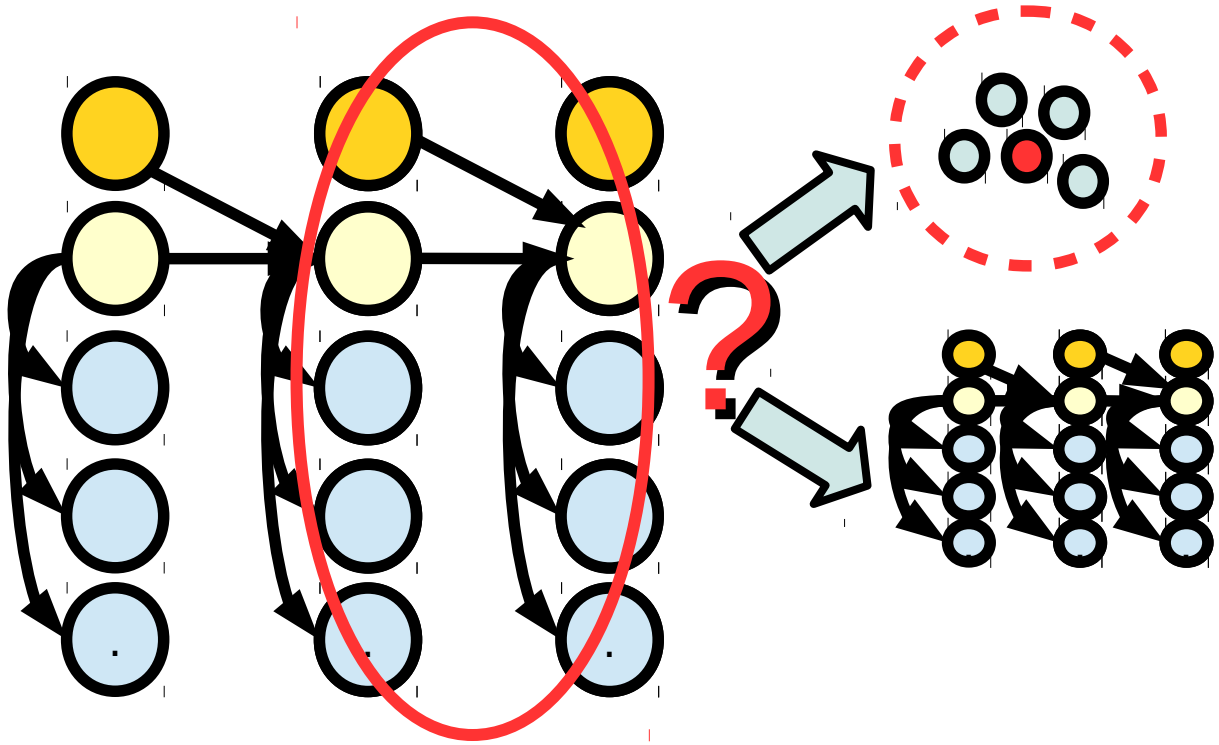
*Mixture chooses POMDP about 30% of the time.

Parbhoo et al., AMIA 2017

Does it work??

- 32,960 pa
  Resist Da
  out 3,000
- Observat
  viral load
- Actions: 3
  drug com
  20 drugs)

Extension: Putting the mixing in the model.



ward

± 3.72

2.61

2.15

± 1.31

0% of the time.

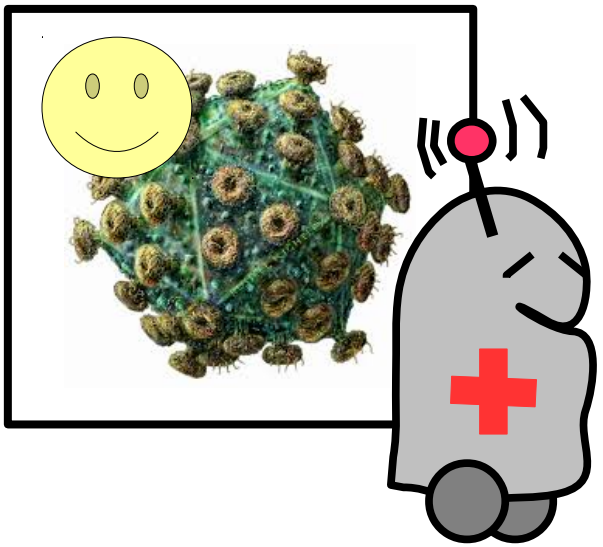
Parbhoo et al., PLoS Medicine 2018

Does it work??

- 32,960 patients from EU Resist Database; hold out 3,000 for testing.
- Observations: CD4s, viral loads, mutations
- Actions: 312 common drug combinations (from 20 drugs)

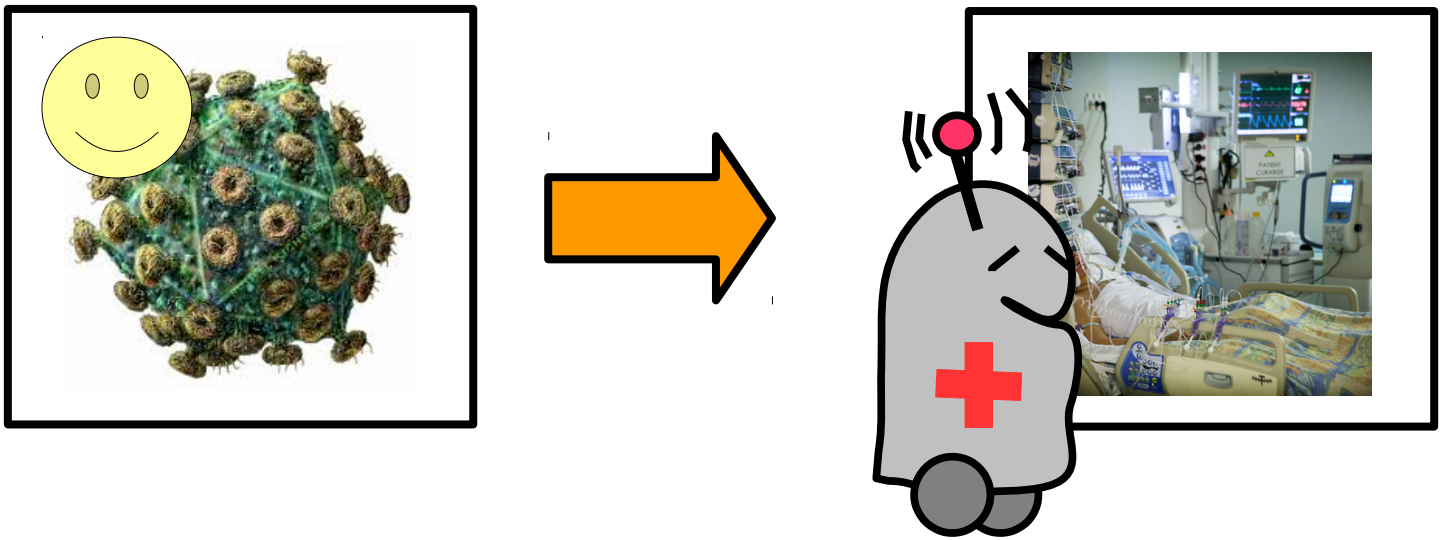
Approach	DR Reward
Random Policy	-7.31 ± 3.72
Neighbor Policy	9.35 ± 2.61
Model-Based Policy	3.37 ± 2.15
Policy-Mixture Policy	11.52 ± 1.31
Model-Mixture Policy	12.47 ± 1.38

Where next?



http://www.rkm.com.au/VIRUS/HIV/HIV-images/HIV-virus.jpg
https://img.clipartxtras.com/c3cfa613babc26f66f50cec50717941f_adventure-clipart-black-and-white-pencil-and-in-color-adventure-internet-explorer-clipart-black-and-white_509-600.jpeg

What about the ICU?



http://www.rkm.com.au/VIRUS/HIV/HIV-images/HIV-virus.jpg
https://img.clipartxtras.com/c3cfa613babc26f66f50cec50717941f_adventure-clipart-black-and-white-pencil-and-in-color-adventure-internet-explorer-clipart-black-and-white_509-600.jpeg

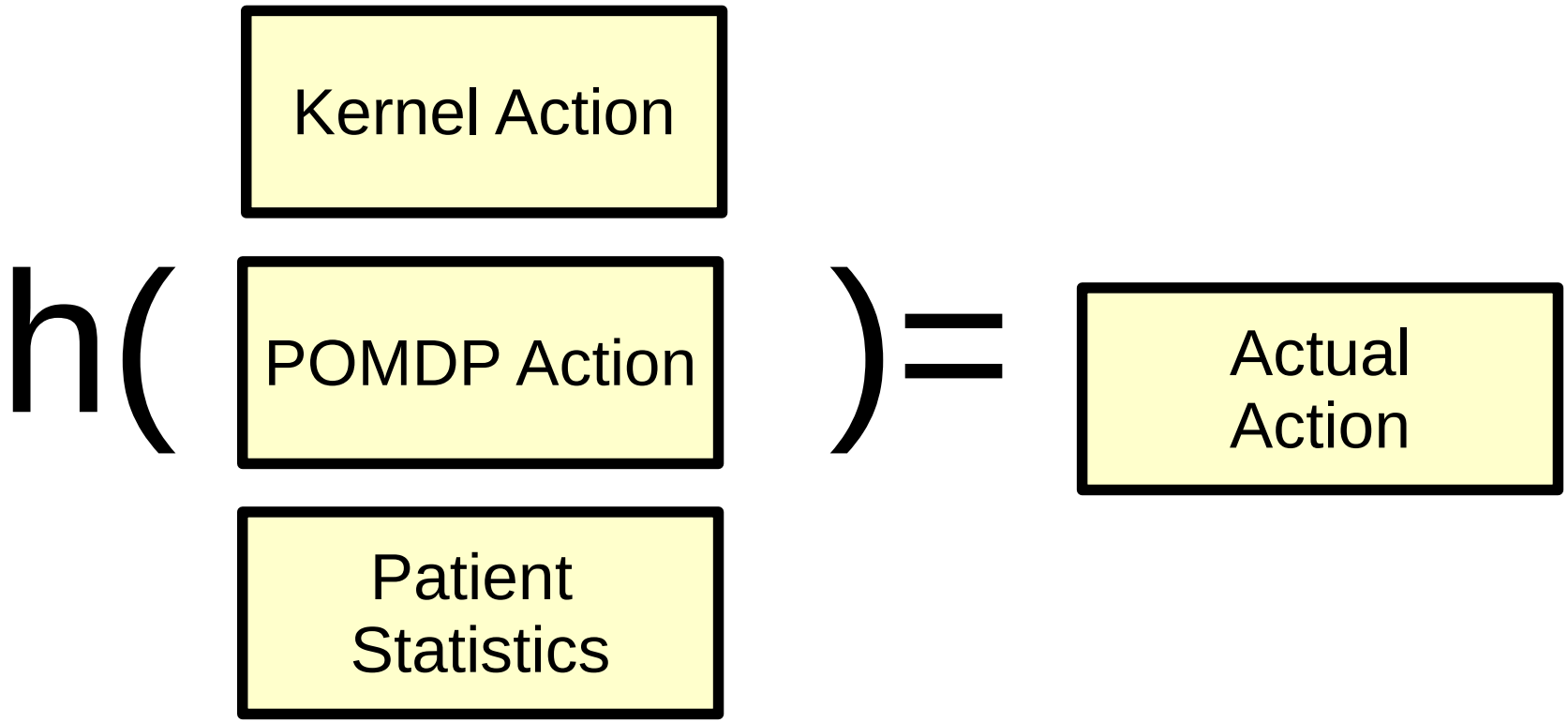
Managing Sepsis in the ICU

- Cohort of 15,415 patients with sepsis from the MIMIC dataset (same as Raghu et al. 2017); contains vitals and some lab tests.
- Actions: focus on vasopressors and fluids, used to manage circulation.
- Goal: reduce 30-day mortality; rewards based on probability of 30-day mortality:

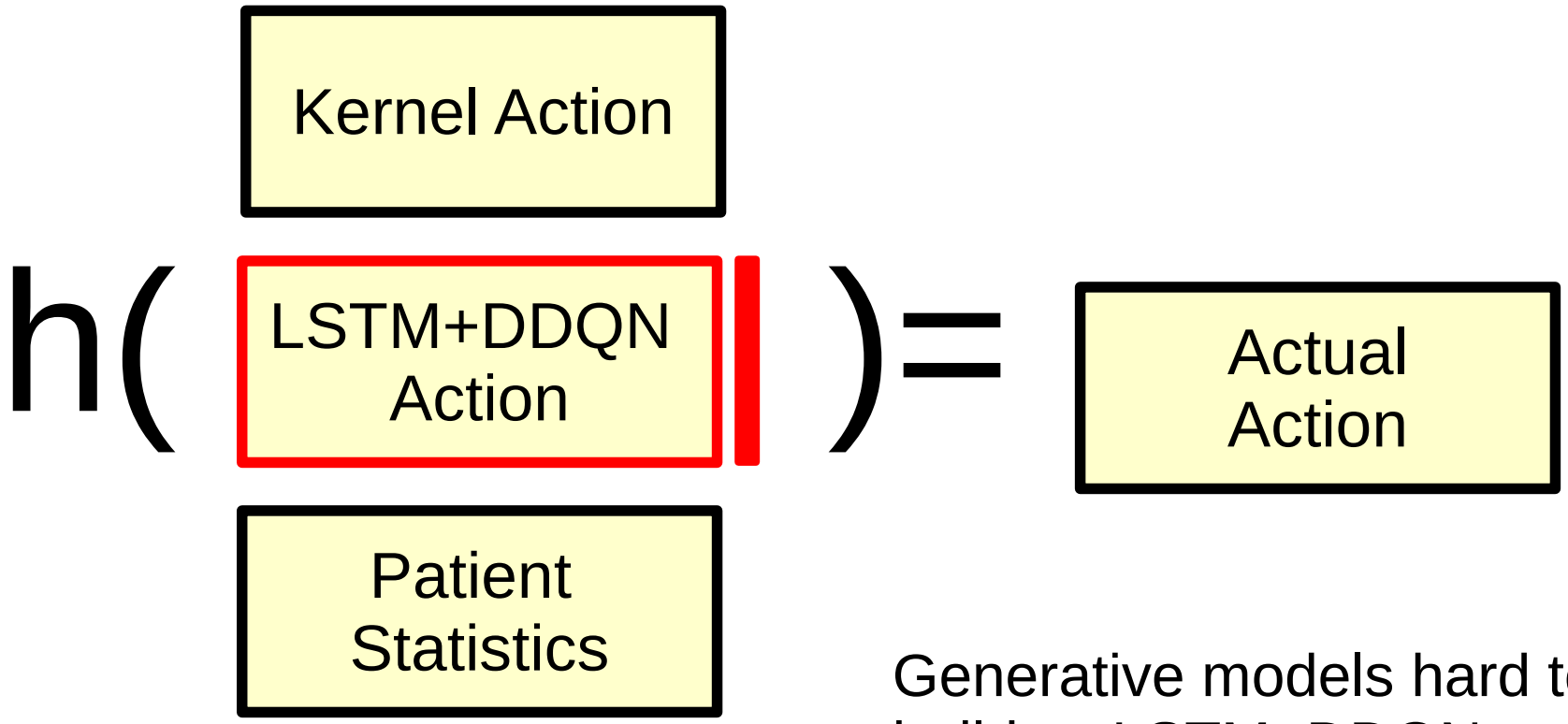
$$r(o, a, o') = -\log \frac{f(o')}{1 - f(o')} f(o') + \log \frac{f(o)}{1 - f(o)}$$

Peng et al., AMIA 2018

Can we apply the same idea?



Can we apply the same idea?

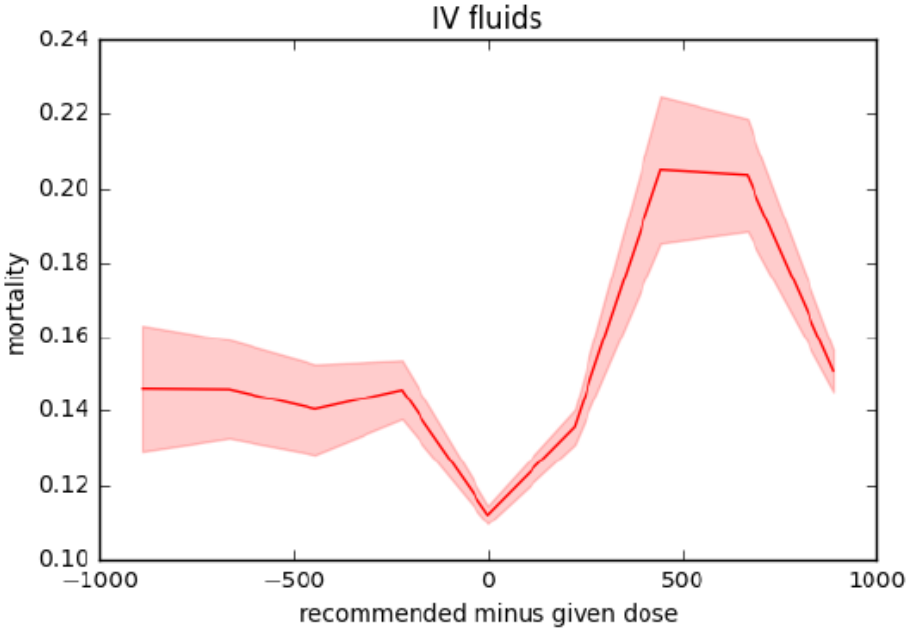


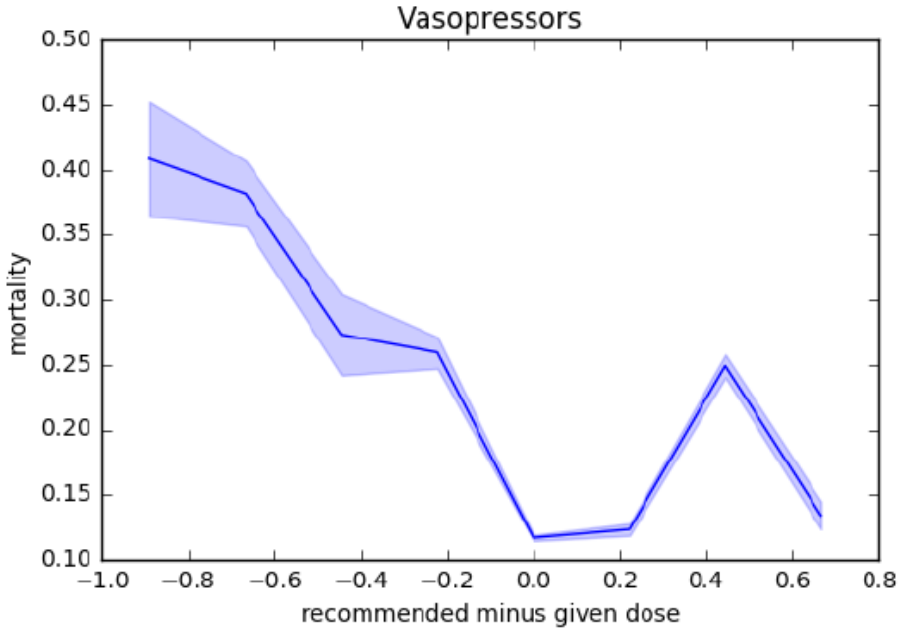
Generative models hard to build → LSTM+DDQN

LSTM+DDQN suggests never-taken actions → hard cap.

Again, Mixtures do best!

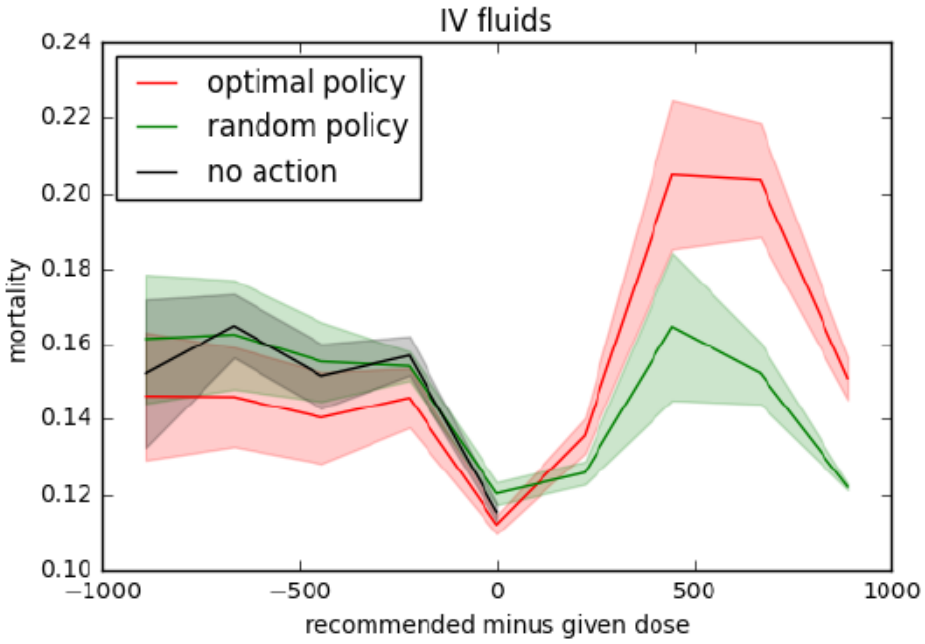
	Physician	Kernel	DQN	MoE_{V_d,Q_d}	MoE_{V_b,Q_b}
non-recurrent encoded	3.76	3.73	4.06	3.93	4.31
recurrent encoded	3.76	4.46	4.23	5.03	5.72

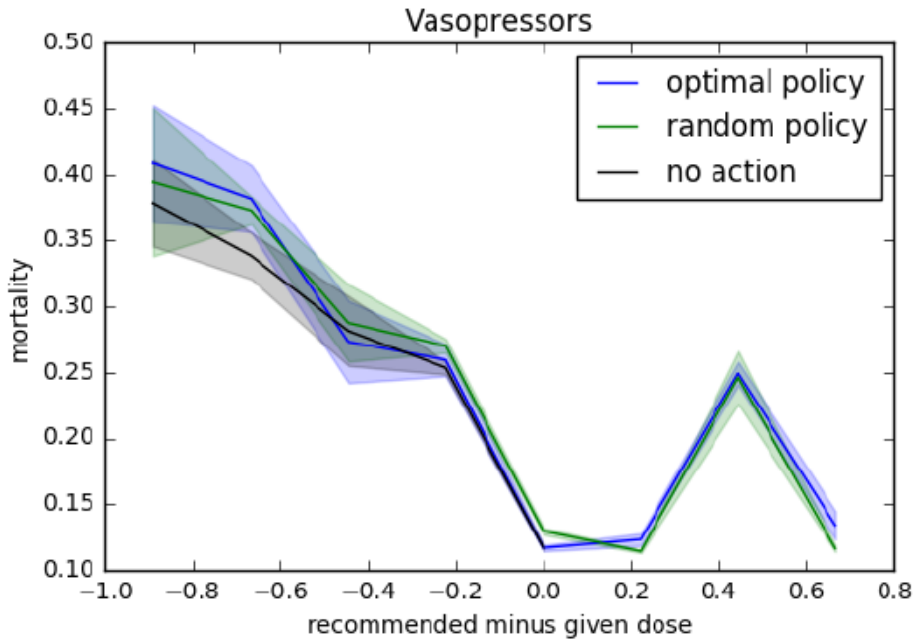




Again, Mixtures do best!

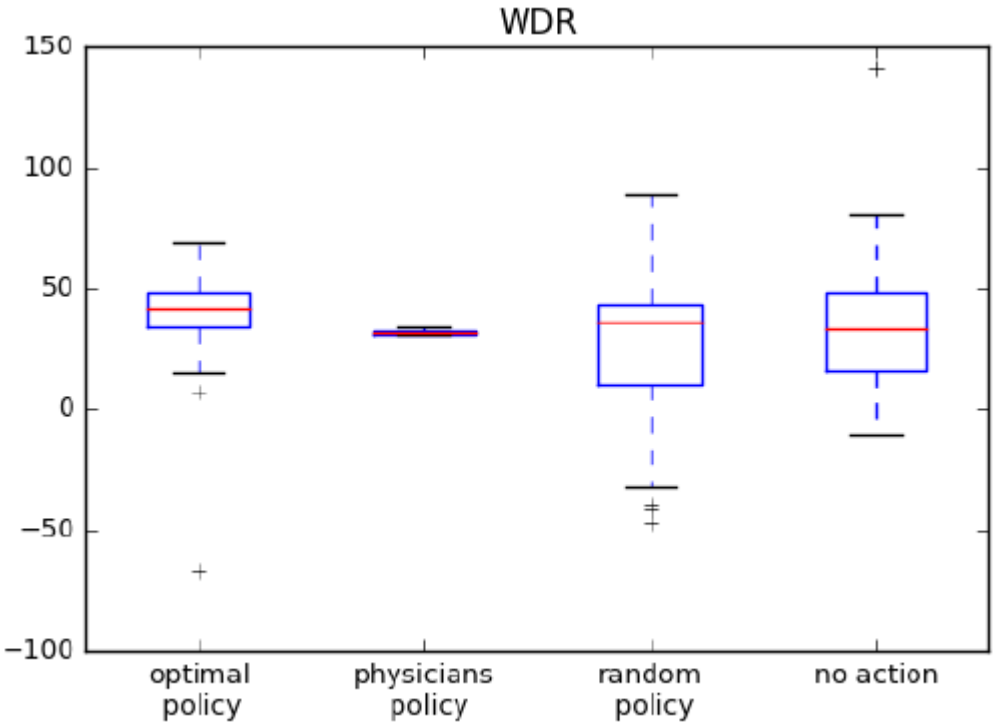
	Physician	Kernel	DQN	MoE_{V_d,Q_d}	MoE_{V_b,Q_b}
non-recurrent encoded	3.76	3.73	4.06	3.93	4.31
recurrent encoded	3.76	4.46	4.23	5.03	5.72





Evaluation Challenges: OPE

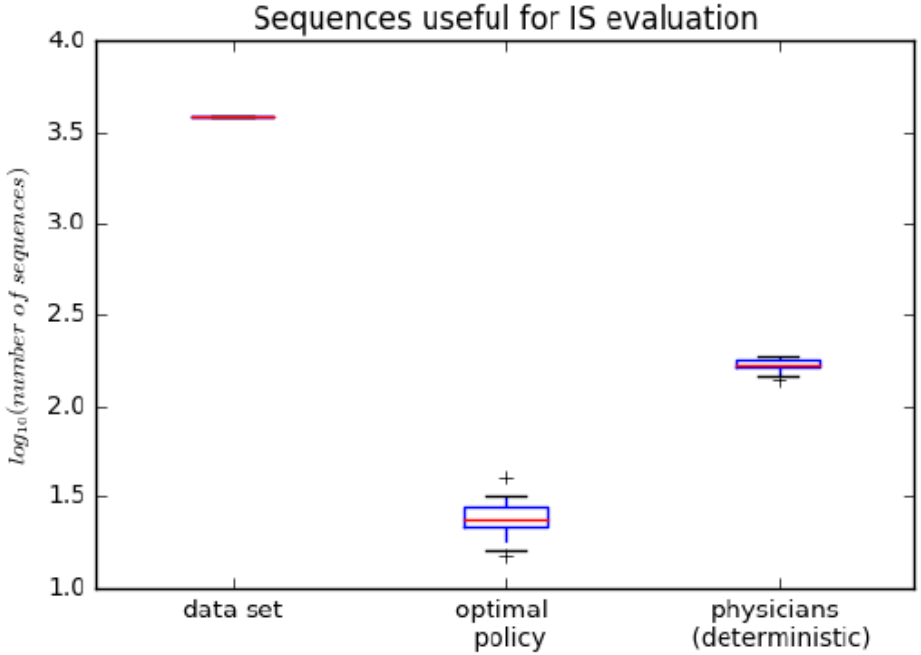
Statistical methods have high variance

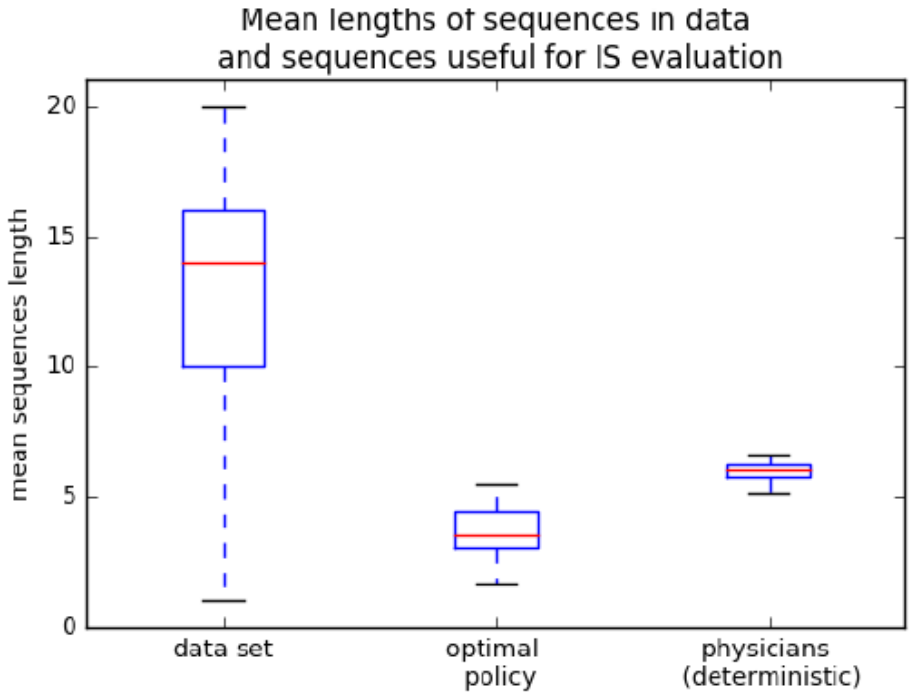


Gottesman et al. 2018, Raghu et al. 2018

Evaluation Challenges: OPE

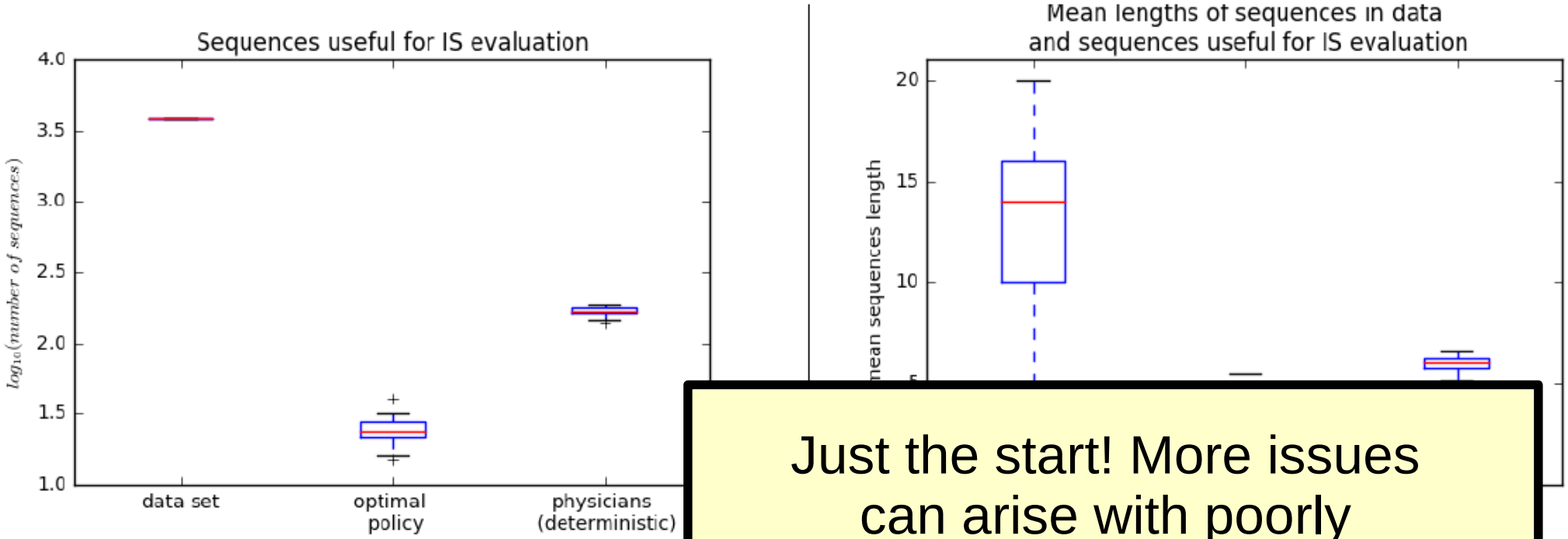
They can select non-representative cohorts





Evaluation Challenges: OPE

They can select non-representative cohorts



Just the start! More issues can arise with poorly estimated behavior policies, poor representation choices

So can we trust those results?

Positive Evidence: Replication

Our HIV results hold across two distinct cohorts.

EU Resist

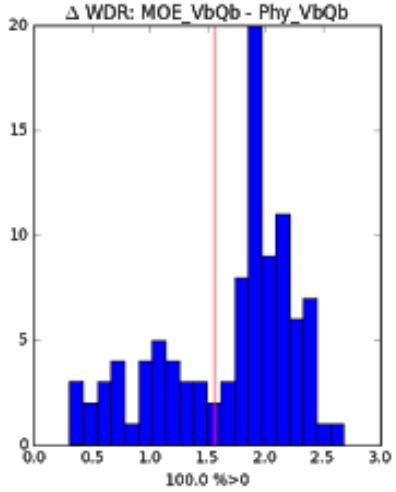
	Doubly Robust	Importance Sampling	Weighted Importance
Random Policy	–2.31 ± 1.42	–3.48 ± 1.36	–2.80 ± 1.27
Short-term Kernel	2.17 ±1.4	2.18 ±1.20	2.16 ±1.71
Long-term Kernel	9.47 ± 1.70	5.72 ± 1.81	6.97 ± 1.29
POMDP	6.04 ± 2.18	4.15 ± 2.28	6.67 ± 1.74
Mixture-of-experts	**11.83 ± 1.26**	**12.50 ± 1.19**	**11.07 ± 1.21**

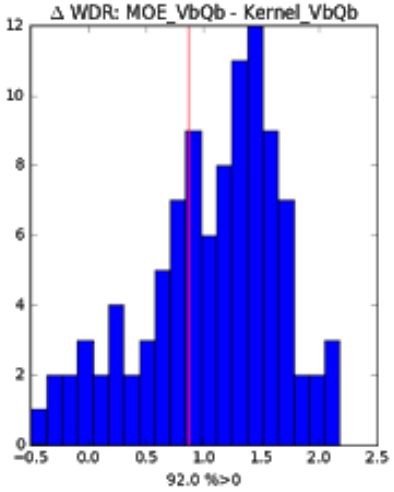
Swiss HIV Cohort

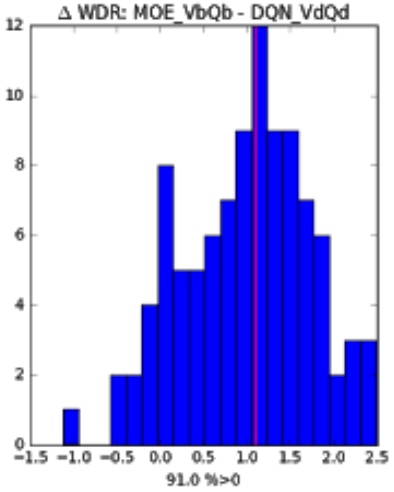
	Doubly Robust	Importance Sampling	Weighted Importance
Random Policy	–6.33 ± 3.47	–5.57 ± 2.17	–6.18 ± 3.24
Short-term Kernel	1.64 ±1.86	2.03 ±1.81	2.17 ±1.74
Long-term Kernel	9.67 ± 1.49	7.38 ± 1.72	7.64 ± 1.92
POMDP	5.46 ± 2.05	6.72 ± 2.88	7.76 ± 2.10
Mixture-of-experts	**10.73 ± 1.02**	**13.59 ± 1.57**	**11.83 ± 1.31**

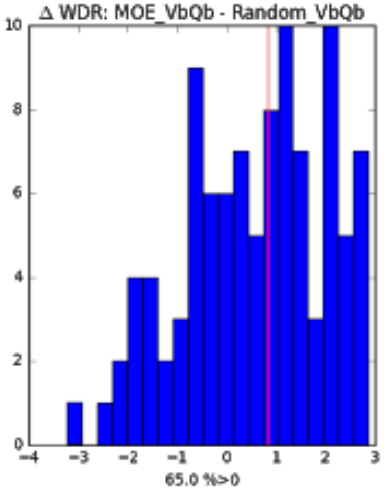
Positive Evidence: Sensitivity Analysis

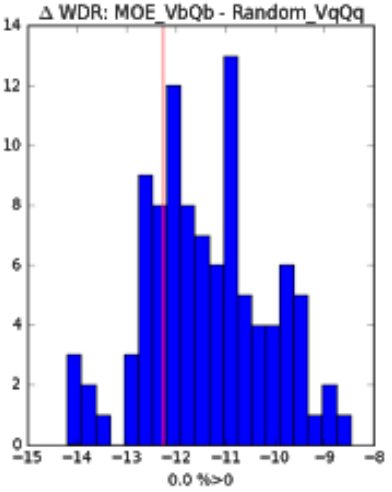
Sepsis: results hold with different control variates

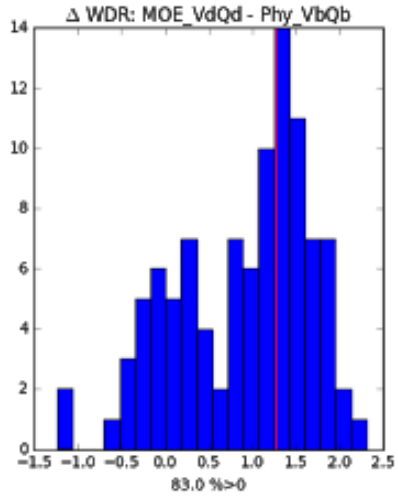


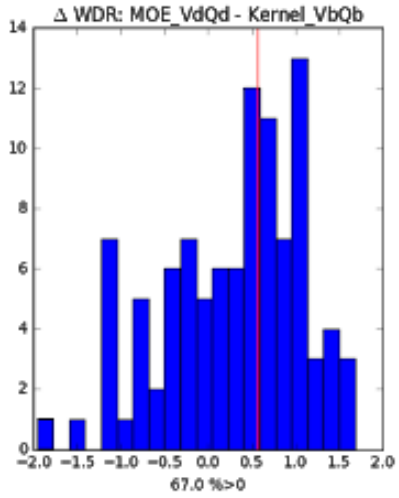


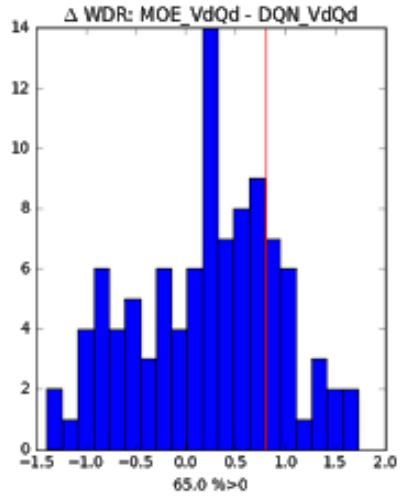


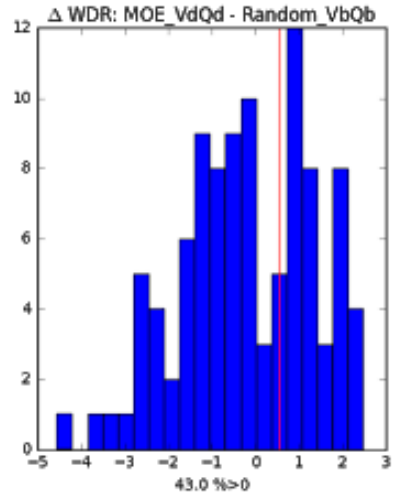


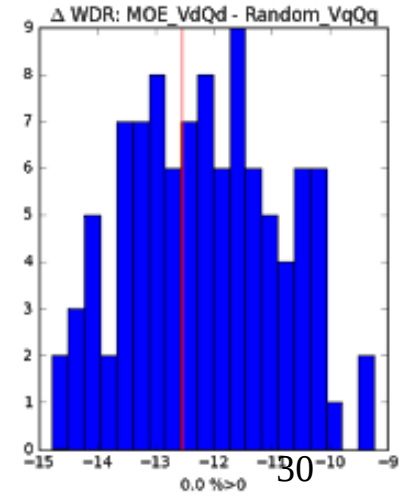












Positive Evidence: Expert Confirmation

- HIV: Checking against standard of care:

	NNRTIs	NRTIs	PIs	Fusion/Entry Inhibitors
First-line therapy	12 157	3 054	774	128
Second-line therapy	4 068	8 764	6 082	1 042

- As well as three expert clinicians:

	Clinician 1	Clinician 2	Clinician 3
Agree	18	15	13
Partially Agree	10	11	13
Disagree	2	4	4

Responsible RL in Healthcare

- Representation learning: are the essential variables at least used by clinicians being retained?
- Statistical evaluation: How large is the bias, variance in off-policy evaluation?
- Interpretation: Do clinicians find it valid?

Finally, remember that the ML will operate within the health system!

Responsible RL in Healthcare

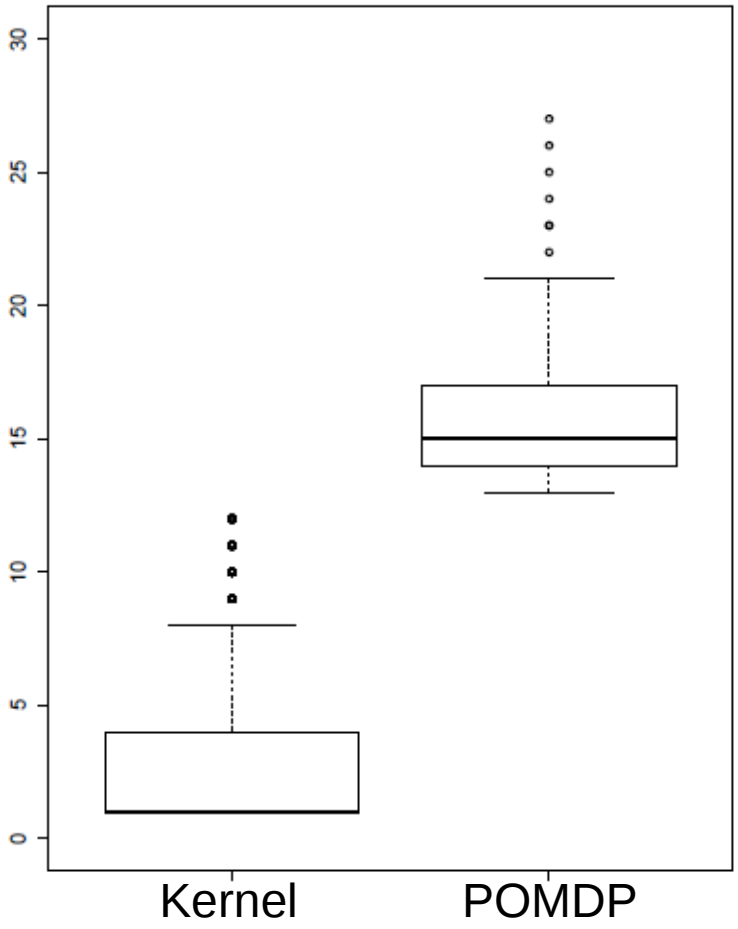
- Representation learning: are the essential variables at least used by clinicians being retained?
- Statistical evaluation: How large is the bias, variance in off-policy evaluation?
- Interpretation: Do clinicians find it valid?

Finally, remember that the ML will operate within the health system!

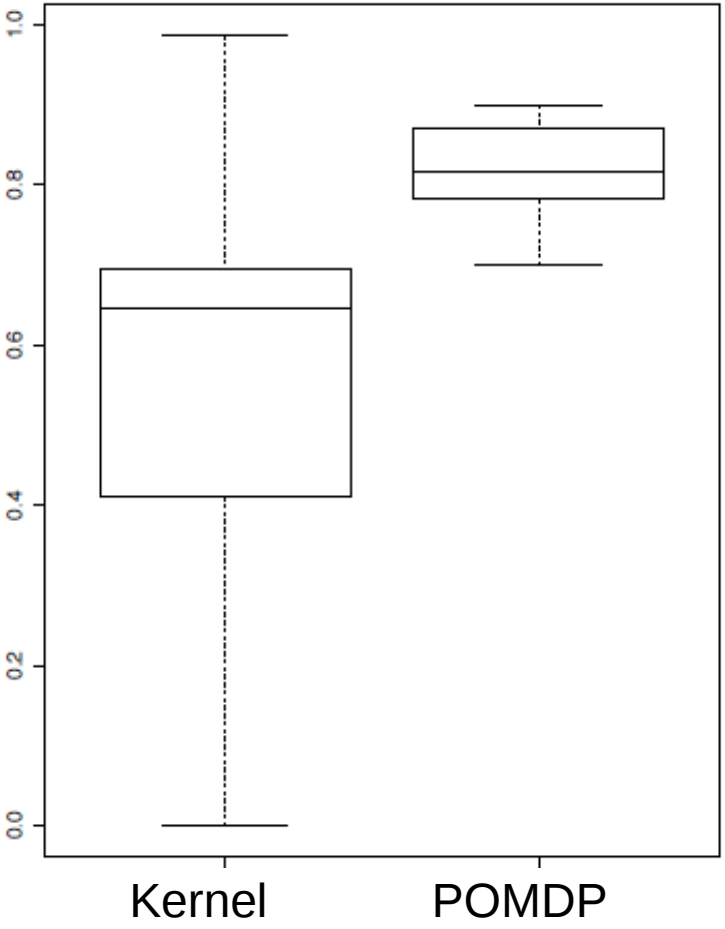
Plugs: MlforHC.org (ML for Health → ML + Clinical researchers)
JAMA Network Open (ML for Health → to clinicians)

And: Our hypothesis was correct! Model used when neighbors are far

History Length

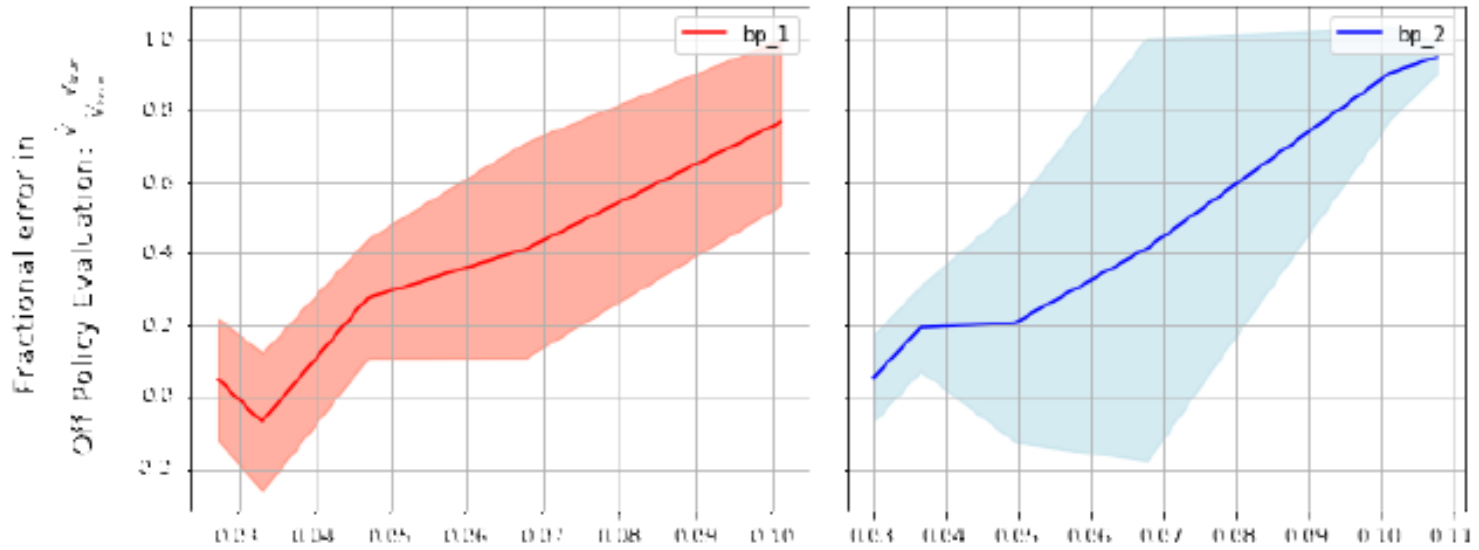


2nd Quantile Distance



Evaluation Challenges: OPE

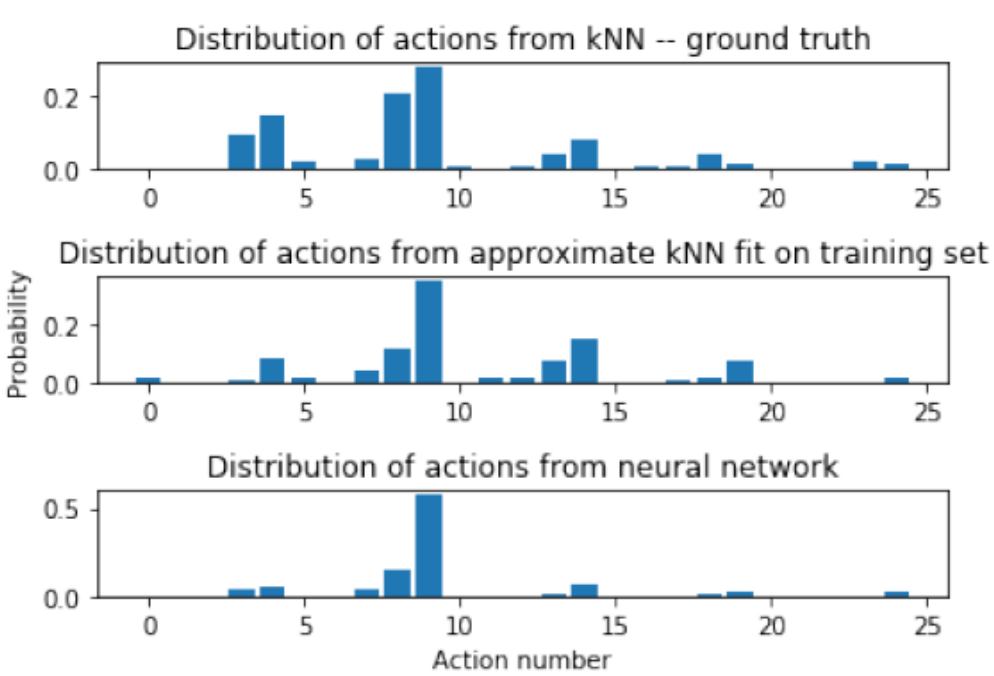
Toy problem: value estimate error grows with behavior policy estimate error



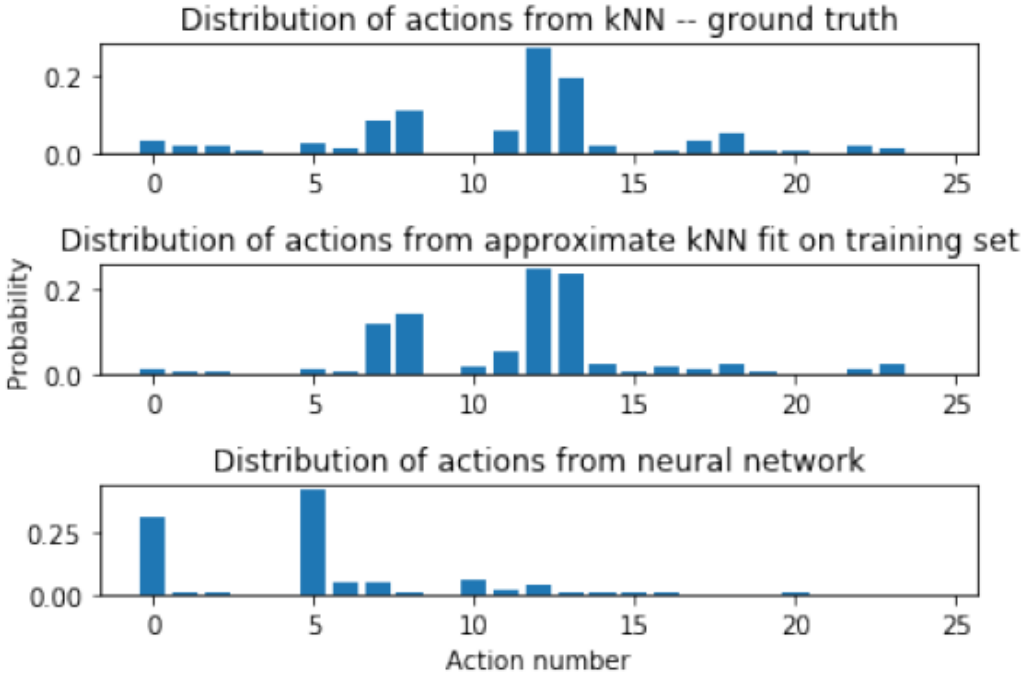
$$\mathrm{WDR}(D) := \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t w_i^t r_t^{H_i} - \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t (w_t^i \hat{Q}^{\pi_e}(S_t^{H_i}, A_t^{H_i}) - w_{t-1}^i \hat{V}^{\pi_e}(S_t^{H_i}))$$

Evaluation Challenges: OPE

Sepsis: Neural networks definitely not calibrated...



(a) Overconfident predictions



(b) Incorrect predictions

$$\mathrm{WDR}(D) := \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t w_i^t r_t^{H_i} - \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t\left(w_t^i \hat{Q}^{\pi_e}(S_t^{H_i}, A_t^{H_i}) - w_{t-1}^i \hat{V}^{\pi_e}(S_t^{H_i})\right)$$

Evaluation Challenges: OPE

kNN is more calibrated

Severity	LR	RF	NN	Approx kNN
0 - 4	0.249	0.214	0.213	**0.129**
5 - 9	0.269	0.254	0.246	**0.152**
10 - 13	0.309	0.309	0.399	**0.210**
14 - 23	0.356	0.337	0.426	**0.199**

Calibration helps

Behaviour Policy Model	MDP Approximate Model	MSE
Approximate kNN	Fitted Q Iteration	**3.05**
Approximate kNN	Kernel-based RL	6.54
Approximate kNN	Discrete SARSA	6.53
Neural network	Fitted Q Iteration	3.53
Neural network	Kernel-based RL	10.2

$$\mathrm{WDR}(D) := \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t w_i^t r_t^{H_i} - \sum_{i=1}^{I}\sum_{t=0}^{T}\gamma^t\left(w_t^i \hat{Q}^{\pi_e}(S_t^{H_i}, A_t^{H_i}) - w_{t-1}^i \hat{V}^{\pi_e}(S_t^{H_i})\right)$$

And even the best OPE can’t counteract bad modeling choices...

https://i0.wp.com/lipstickandlifesaving.com/wp-content/uploads/2017/03/img_9064.jpg