Machine Learning for Healthcare
Disease progression modeling

RAHUL G. KRISHNAN

SENIOR RESEARCHER AT MICROSOFT RESEARCH NEW ENGLAND

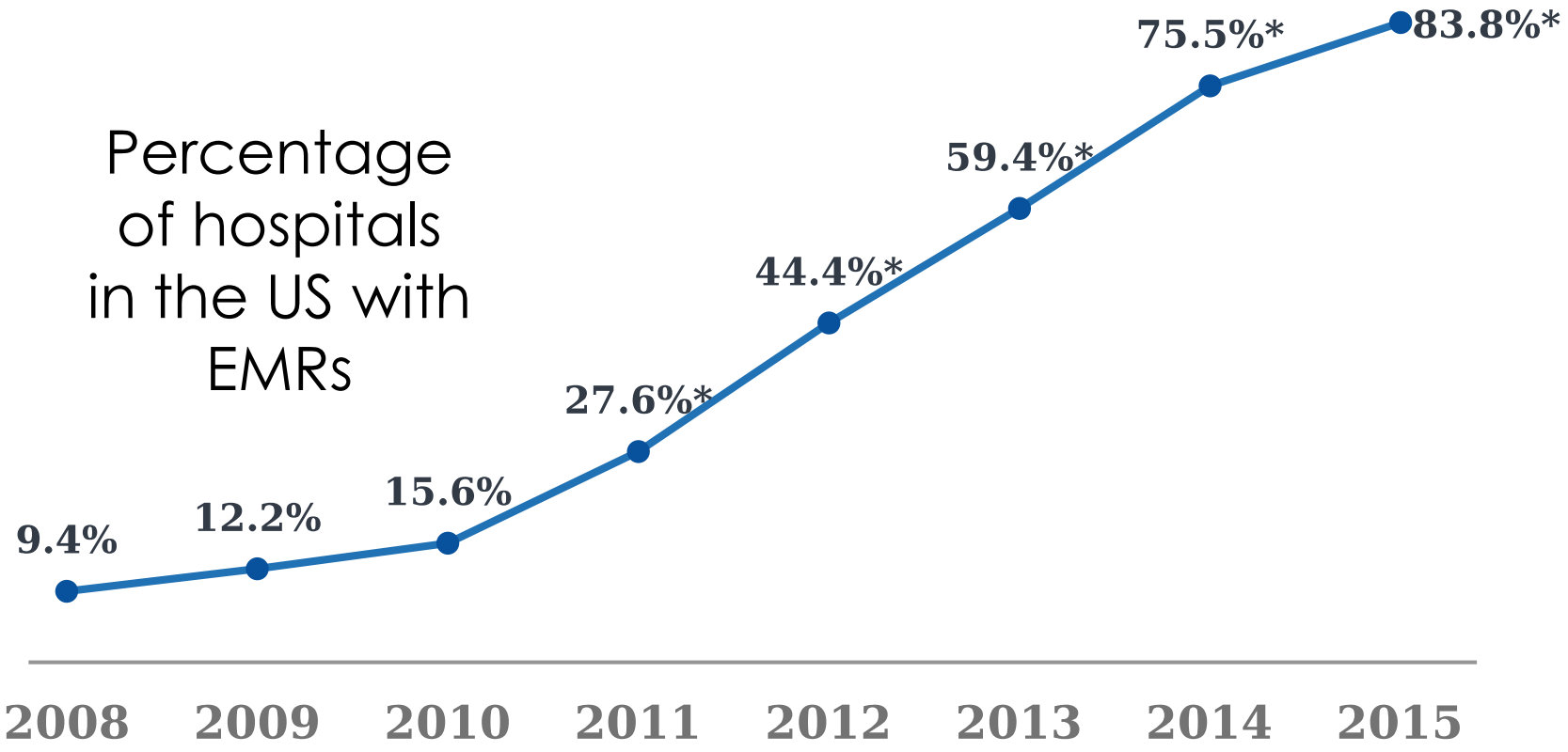
FALL 2021: ASSISTANT PROFESSOR IN CS & MEDICINE

UNIVERSITY OF TORONTO & VECTOR INSTITUTE

Overview

- Introduction
  - ML in healthcare - from data to insights
- Disease progression
  - Motivation
  - Problem statement
  - Deep Markov Models
  - Neural Intervention Effect Functions
  - Neural Pharmacodynamic State Space Models
  - Results
  - Future work
- Conclusion
  - Opportunities in research

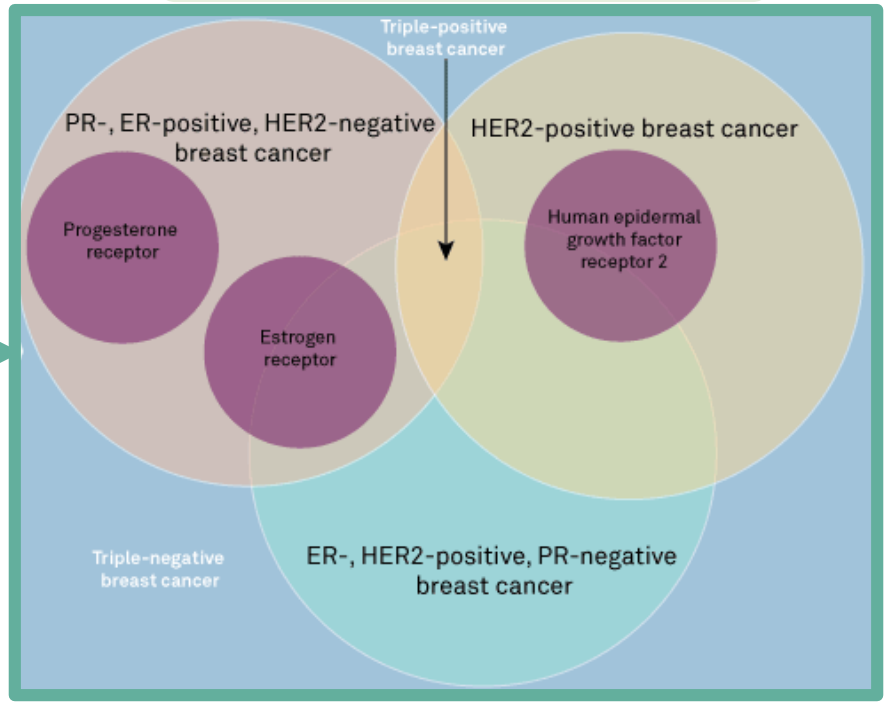
Digitization of electronic healthcare data



[Henry et al., ONC Data Brief, May 2016]

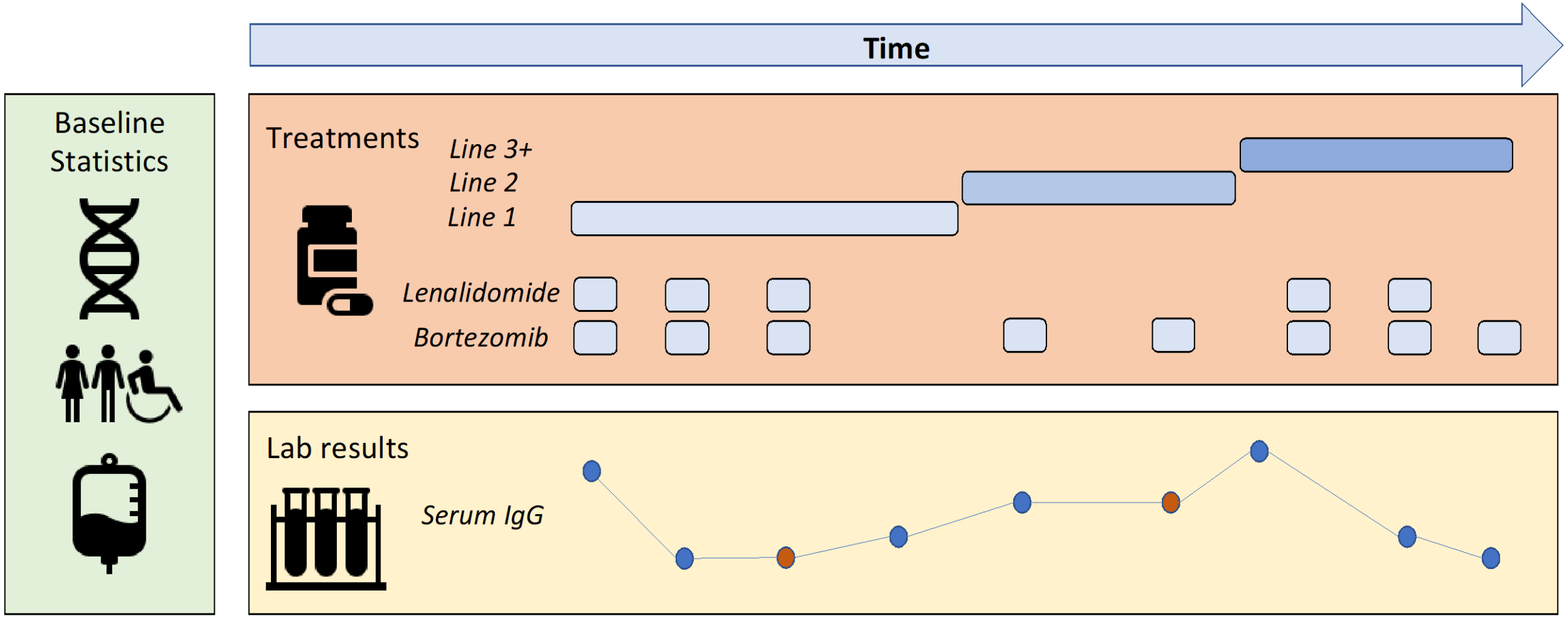
Clinical data

Understand disease biology



Build clinical tools

Disease registries

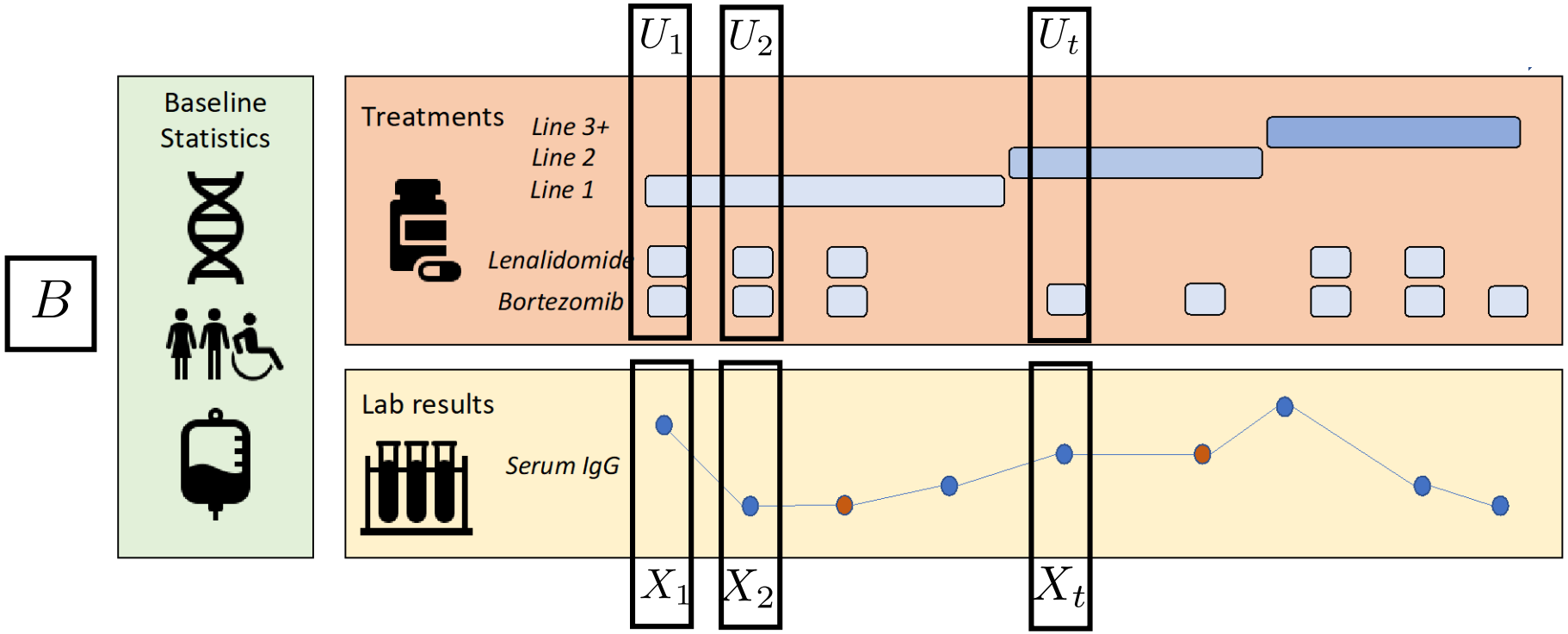


Disease progression

Neural Pharmacodynamic State Space Models,
Hussain & Krishnan, Sontag, ICML 2021

Modeling disease progression

- What can we learn about diseases using data of patients who suffer from it?
- **Goal:** Unsupervised learning of clinical biomarkers: $\text{maximize} \sum_{i=1}^{N} \log p(\mathbf{X}^i | \mathbf{U}^i, B^i)$

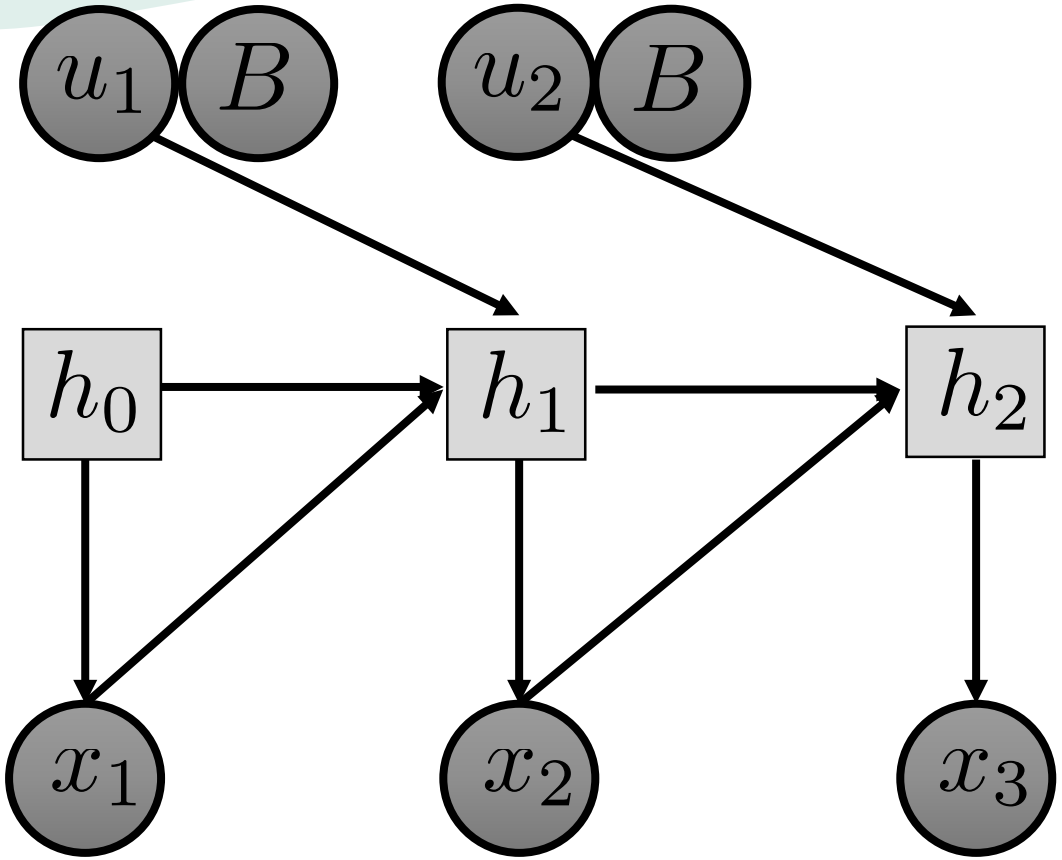


MULTIPLE MYELOMA
Research Foundation

Technical challenges in healthcare data

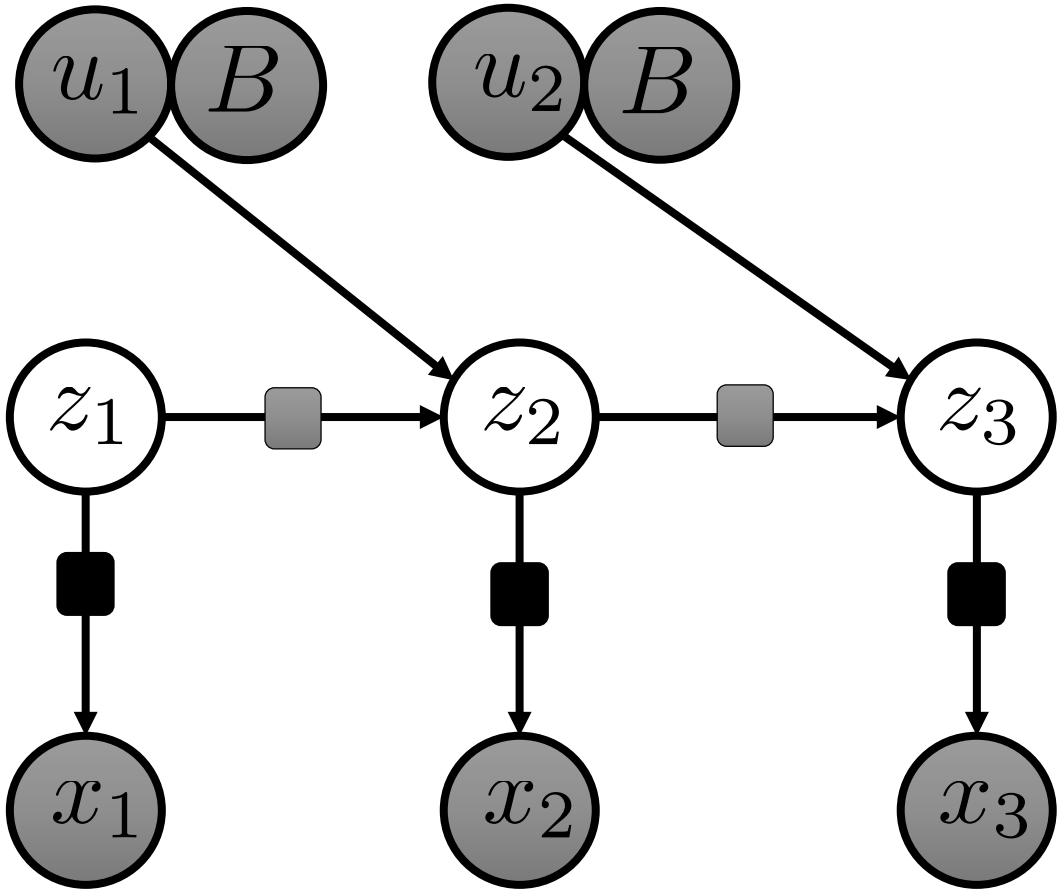
- High-dimensional longitudinal data X has nonlinear variation
- Missingness in X
- Left and right censorship
- Complex variation in X due to treatment protocols U
- **Rare diseases:** Small number of samples to learn from

Statistical models of sequential data



Recurrent Neural Network (RNN)

$$p(\mathbf{X}|\mathbf{U},B) = \prod_{t=1}^{T} p(X_t|X_{<t}, U_{<t}, B))$$



(Gaussian) state space models (SSM)

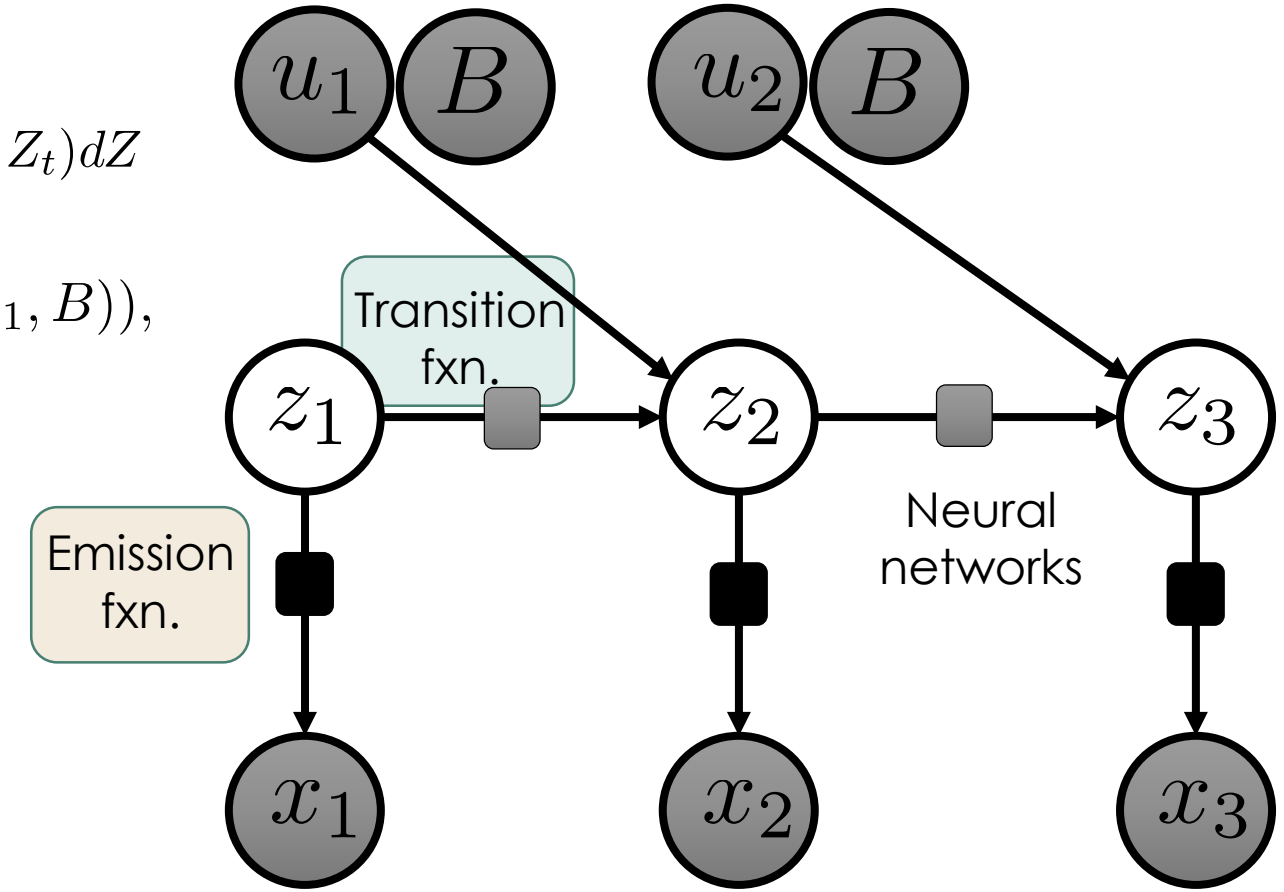
$$p(\mathbf{X}|\mathbf{U},B) = \int_Z \prod_{t=1}^{T} p_\theta(Z_t|Z_{t-1}, U_{t-1}, B) p_\theta(X_t|Z_t) dZ$$

Deep Markov Models [DMM]

$$p(\mathbf{X}|\mathbf{U},B) = \int_Z \prod_{t=1}^{T} p_\theta(Z_t|Z_{t-1},U_{t-1},B)p_\theta(X_t|Z_t)dZ$$

$$Z_t|\cdot \sim \mathcal{N}(\mu_\theta(Z_{t-1},U_{t-1},B), \Sigma_\theta^t(Z_{t-1},U_{t-1},B)),$$
$$X_t|\cdot \sim \mathcal{N}(\kappa_\theta(Z_t), \Sigma_\theta^e(Z_t))$$

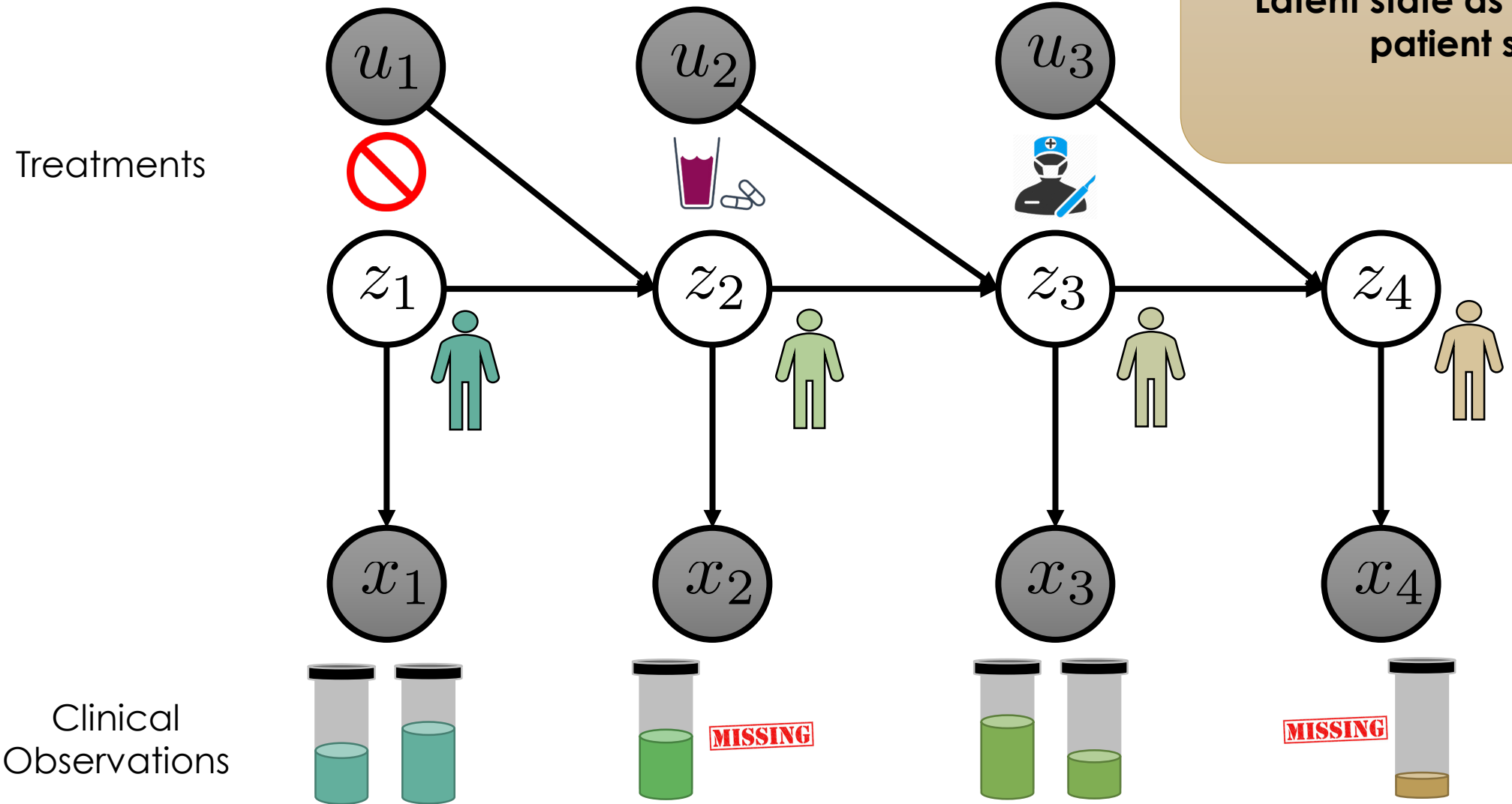
- Parameter estimation via maximum likelihood
- Use variational inference with an inference network for approximating the intractable posterior distribution



Structured inference networks for nonlinear state space models, Krishnan, Shalit & Sontag, AAAI 2017

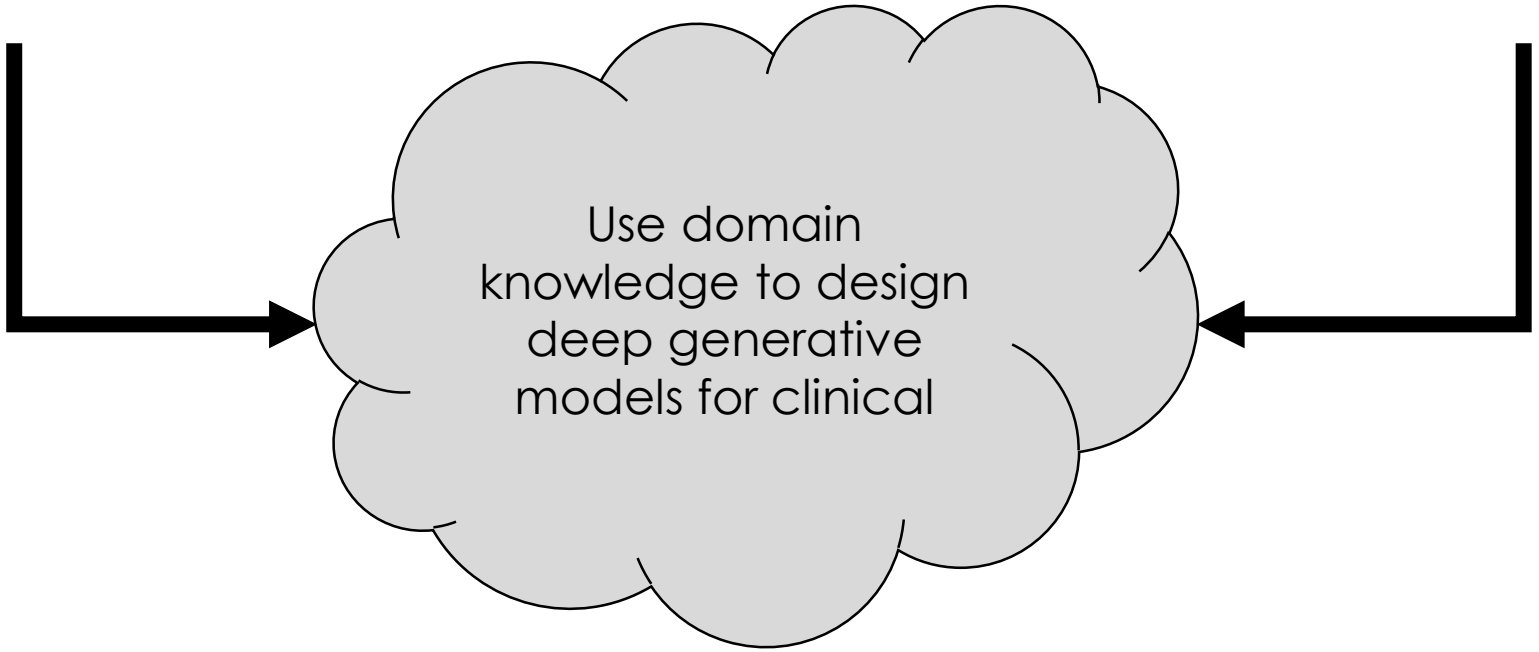
How does the DMM work?

Latent state as a proxy for patient state



A middle ground for models of sequential data

- RNNs/DMMs
  - Powerful black-box models for sequences
  - Susceptible to overfitting when data is scarce
- SSMs
  - Latent variable model with history of use in disease progression
  - Linearity can be a limitation

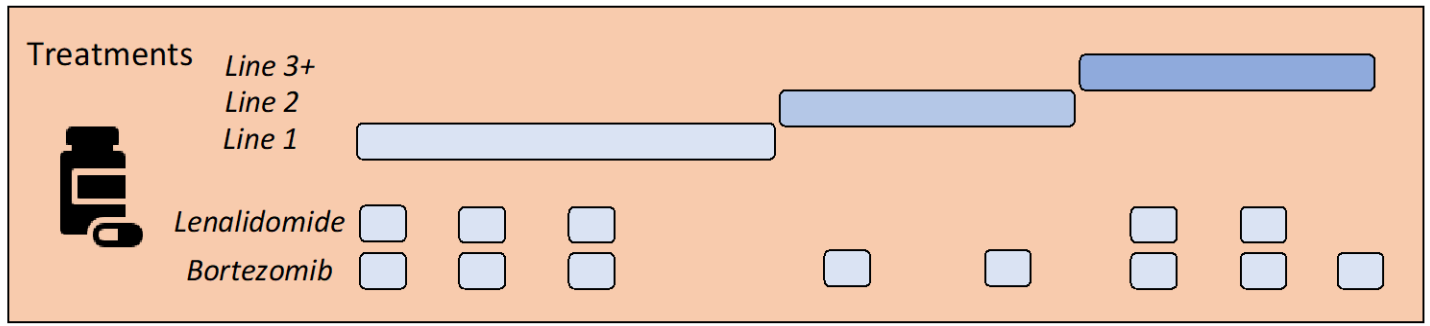


Domain knowledge for disease progression

- What is the right domain knowledge to use for cancer progression?
  - **Lines of therapy**
  - **Mechanism of drug-effect**
- How do we use this knowledge?
  - Design a new neural architecture for the transition function

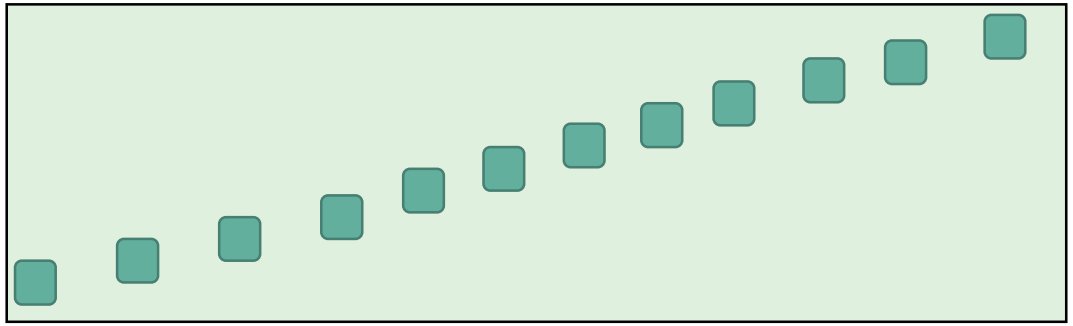
$$Z_t|\cdot \sim \mathcal{N}(\mu_\theta(Z_{t-1}, U_{t-1}, B), \Sigma^t_\theta(Z_{t-1}, U_{t-1}, B)),$$
$$X_t|\cdot \sim \mathcal{N}(\kappa_\theta(Z_t), \Sigma^e_\theta(Z_t))$$

From lines of therapy to local and global clocks



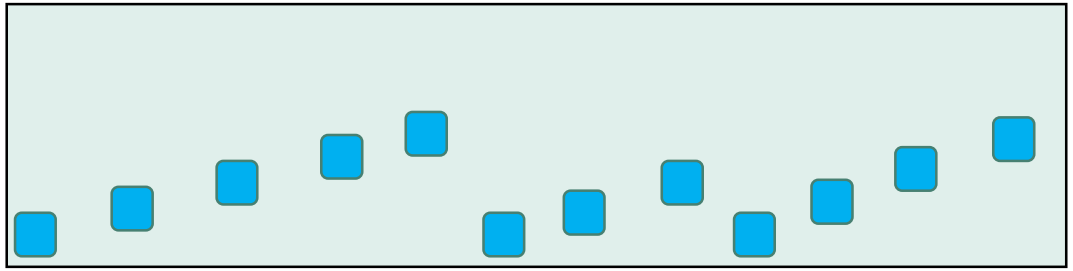
Capture time from start of therapy

Global clock

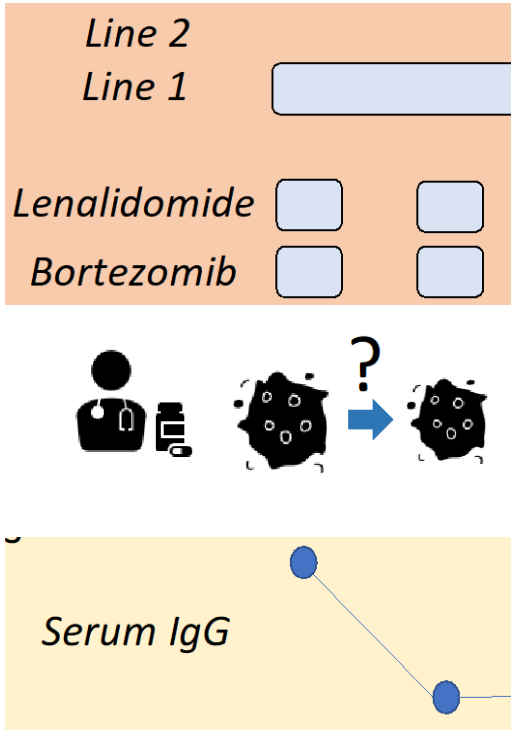


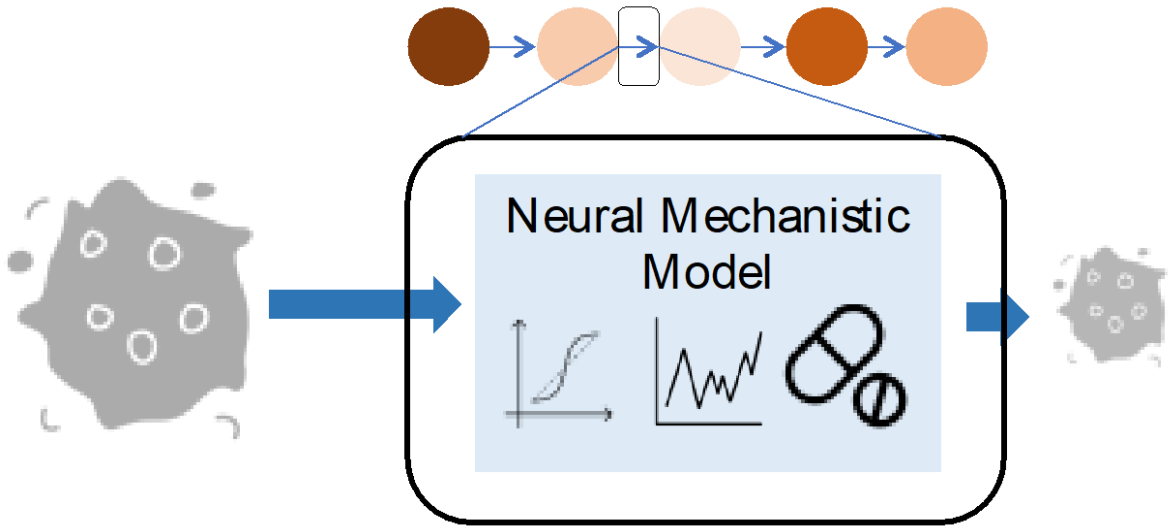
Capture time relative to progression event

Local clock



Approximating the mechanistic effect of drugs





Pharmacokinetics and Pharmacodynamics

- Pharmacokinetics
  - How drugs move within the body.
- Pharmacodynamics
  - Study of how the body responds to the drugs being prescribed
- Traditional PK-PD models are designed to model dynamics of a single biomarker due to a single drug
- **Our work:** proposes new neural architectures to model the effect of multiple simultaneous interventions on latent representations

Neural intervention effect functions

- Modeling baseline conditional variation

$$g_1(Z_{t-1},U_{t-1},B) = Z_{t-1} \cdot \tanh(b_{\text{lin}} + W_{\text{lin}}[U_{t-1},B])$$

- Modeling slow gradual relapse after treatment
  - Log-cell kill $g_2(Z_{t-1},U_{t-1},B) = Z_{t-1} \cdot (1 - \rho \log(Z_{t-1}^2) - \beta \exp(-\delta \cdot \text{lc}_{t-1}))$,
  - Captures rapid variation in representations due to treatment

$$g_3(Z_{t-1},U_{t-1},B) = \begin{cases} b_0 + \alpha_{1,t-1}/[1+\exp(-\alpha_{2,t-1}(\text{lc}_{t-1} - \frac{\gamma_l}{2}))], \\ \text{if } 0 \le \text{lc}_{t-1} < \gamma_l \\ b_l + \alpha_{0,t-1}/[1+\exp(\alpha_{3,t-1}(\text{lc}_{t-1} - \frac{3\gamma_l}{2}))], \\ \text{if } \text{lc}_{t-1} \ge \gamma_l \end{cases}$$

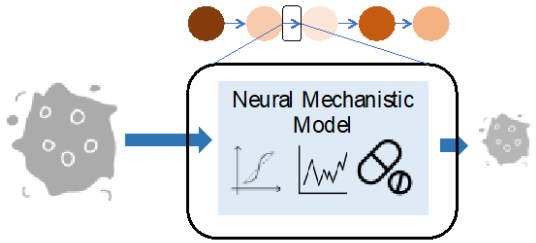
Inspired by:

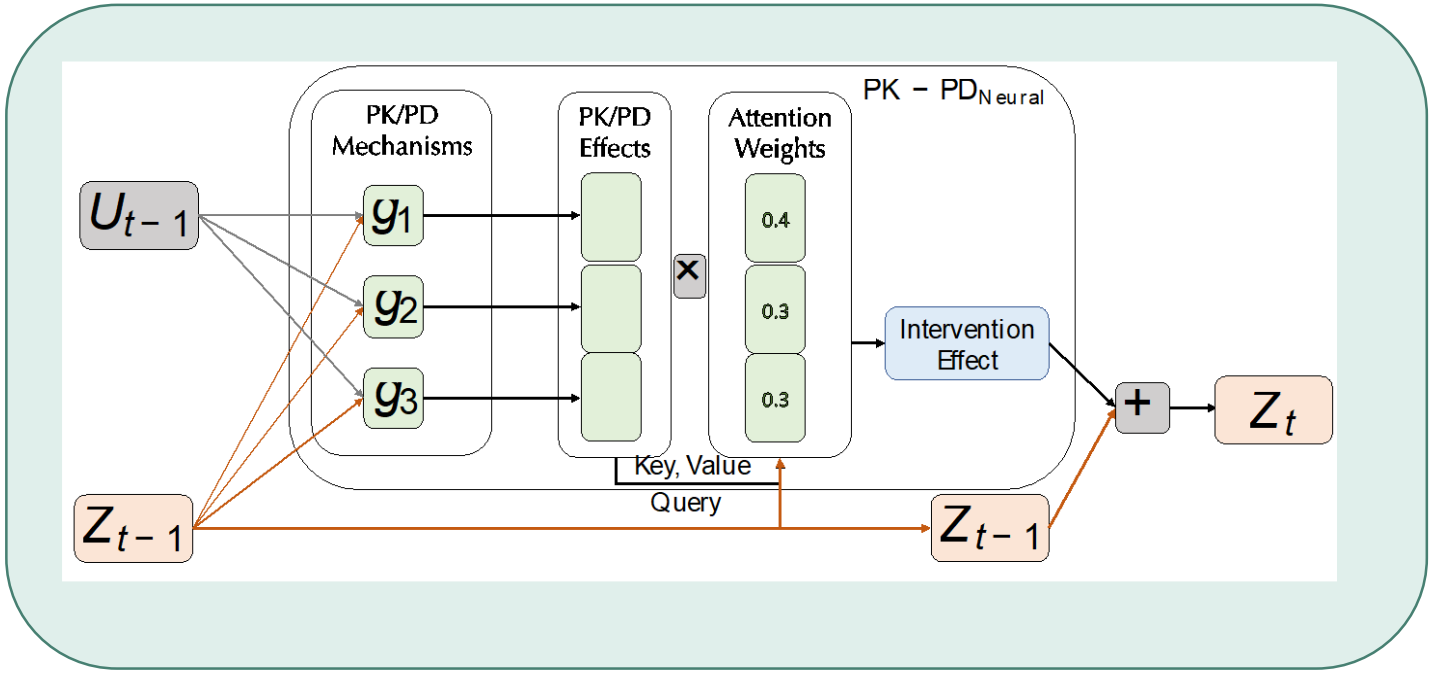
(West & Newton, 2017)

Inspired by:

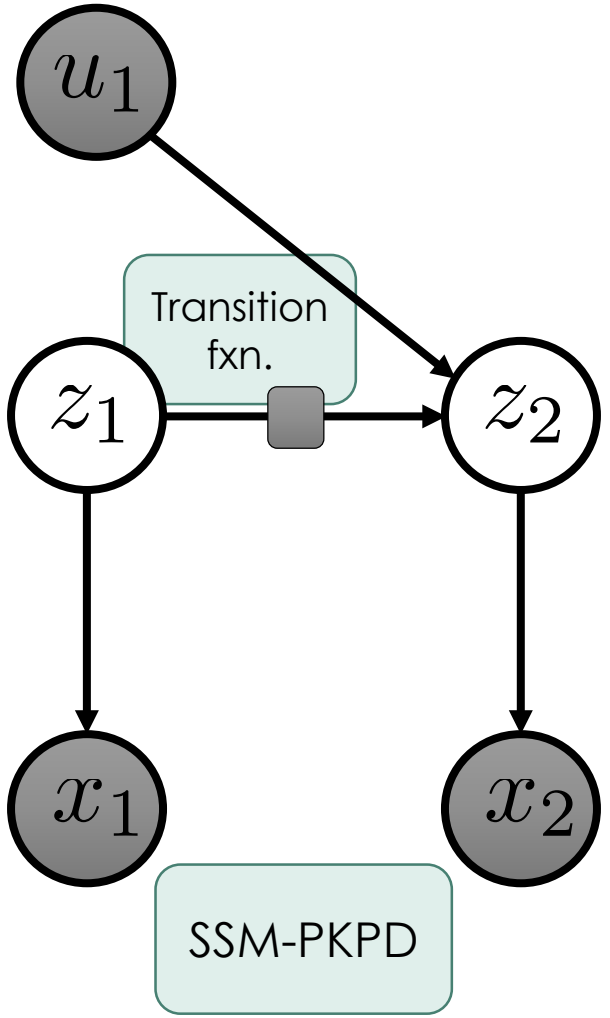
Xu *et al.* (2016)

Neural architecture for the transition function





$$Z_t|\cdot \sim \mathcal{N}(\mu_\theta(Z_{t-1}, U_{t-1}, B), \Sigma^t_\theta(Z_{t-1}, U_{t-1}, B)),$$
$$X_t|\cdot \sim \mathcal{N}(\kappa_\theta(Z_t), \Sigma^e_\theta(Z_t))$$



Cohort characteristics

- 1143 patients aligned by the start of treatment for multiple myeloma
- Treated according to current standard of care
- Worked with an oncologist to select:
  - **X**: 16 clinical biomarkers over time
  - **U**: 9 indicators of treatments (such as drugs and line of therapy)
  - **B**: 16 baseline features
    - PCA projections of RNA SNP data
    - Demographics

Baselines

$\textbf{SSM}_{\text{Linear}}$ parametrizes $\mu_\theta(Z_{t-1}, U_{t-1}, B)$ with a linear function. (Perotte *et al.*, 2015)

$\textbf{SSM}_{\text{NL}}$: nonlinear SSM to capture variation in the clinical biomarkers Krishnan *et al.* (2017)

$\textbf{SSM}_{\text{MOE}}$: We use an SSM whose transition function is parameterized via a Mixture-of-Experts (Jacobs *et al.*, 1991; Jordan & Jacobs, 1994)

$\textbf{SSM}_{\text{Attn.Hist.}}$: We implement a variant of the SSM in Alaa & van der Schaar (2019)

Generalization against baselines

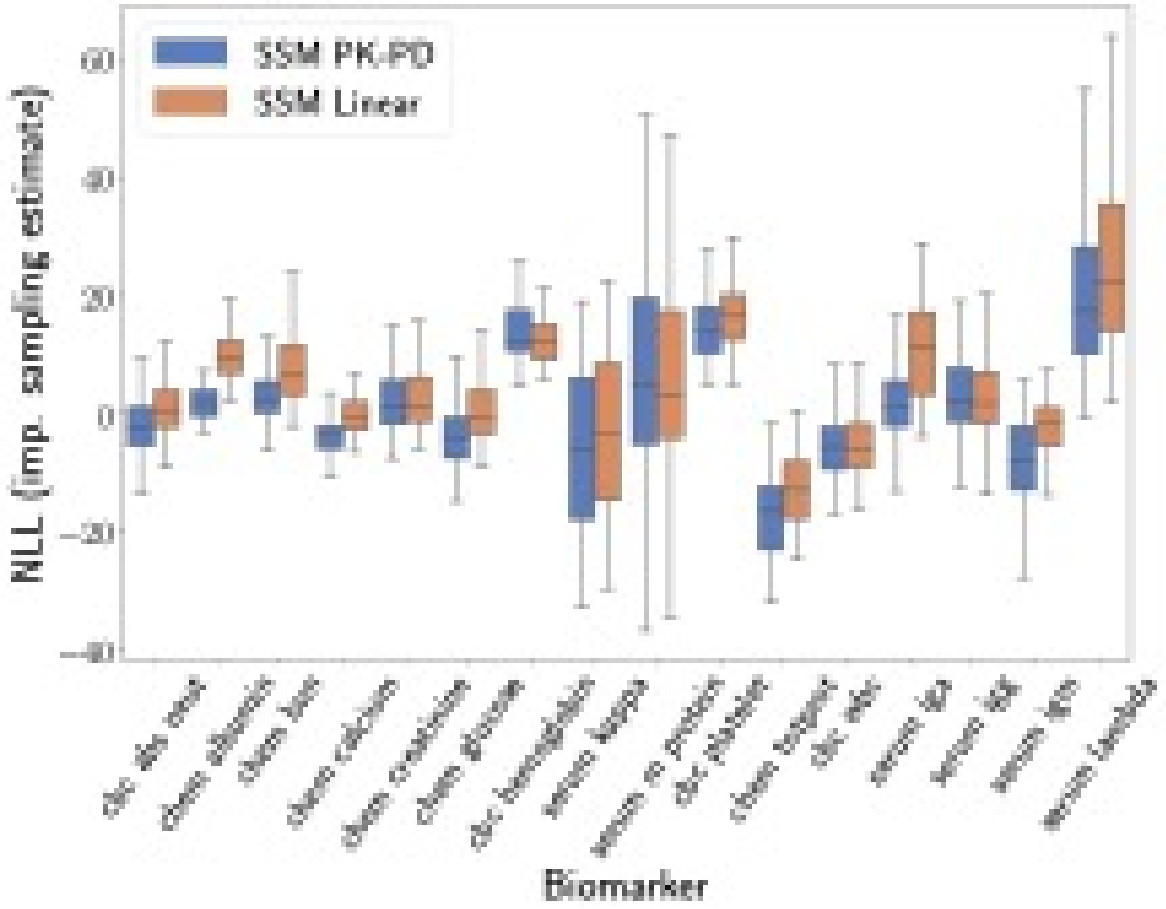
Semi-synthetic data

Training Set Size	SSM Linear	SSM NL	SSM MOE	SSM Attn. Hist.	SSM PK-PD	SSM PK-PD (w/o TExp)
100	58.57 +/- .06	69.04 +/- .11	60.98 +/- .04	76.94 +/- .02	**55.34 +/- .03**	**58.39 +/- .05**
1000	53.84 +/- .02	44.75 +/- .02	51.57 +/- .03	73.80 +/- .03	**39.84 +/- .02**	**38.93 +/- .01**

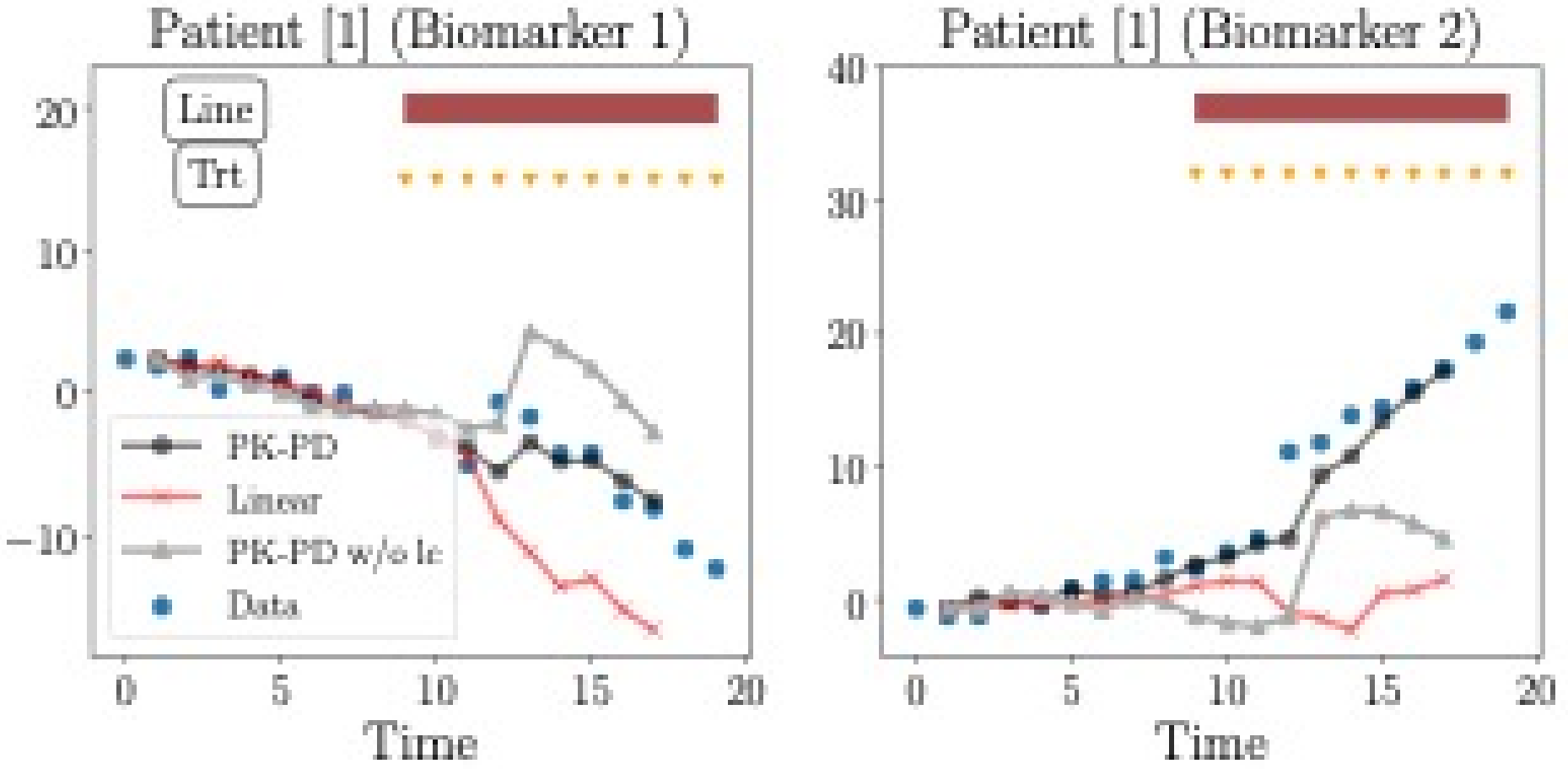
Multiple Myeloma data

Evaluation Metric	SSM PK-PD vs. SSM Linear	SSM PK-PD vs. SSM NL	SSM PK-PD vs. SSM MOE	SSM PK-PD vs. SSM Attn. Hist.
Pairwise Comparison (↑)	0.796 (0.400)	0.760 (0.426)	0.714 (0.450)	0.934 (0.247)
Counts (↑)	PK-PD: 158, LIN: 6	130, 12	94, 8	272, 0
NELBO (↓)	PK-PD: 61.54, LIN: 74.22	61.54, 79.10	61.54, 73.44	61.54, 105.04
# of Model Parameters	PK-PD: 23K, LIN: 7K	23K, 51K	23K, 77K	23K, 17K

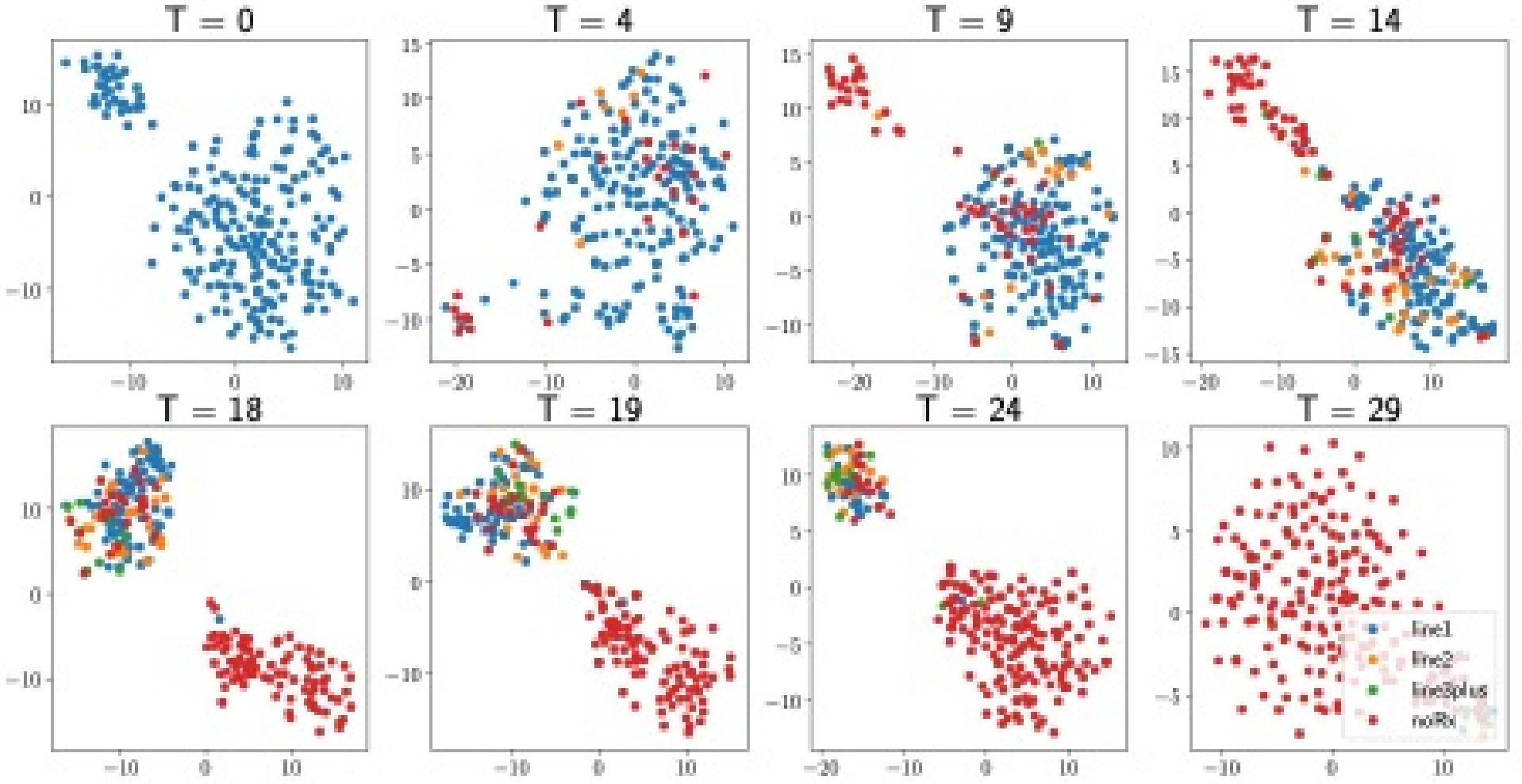
Where do the gains come from?



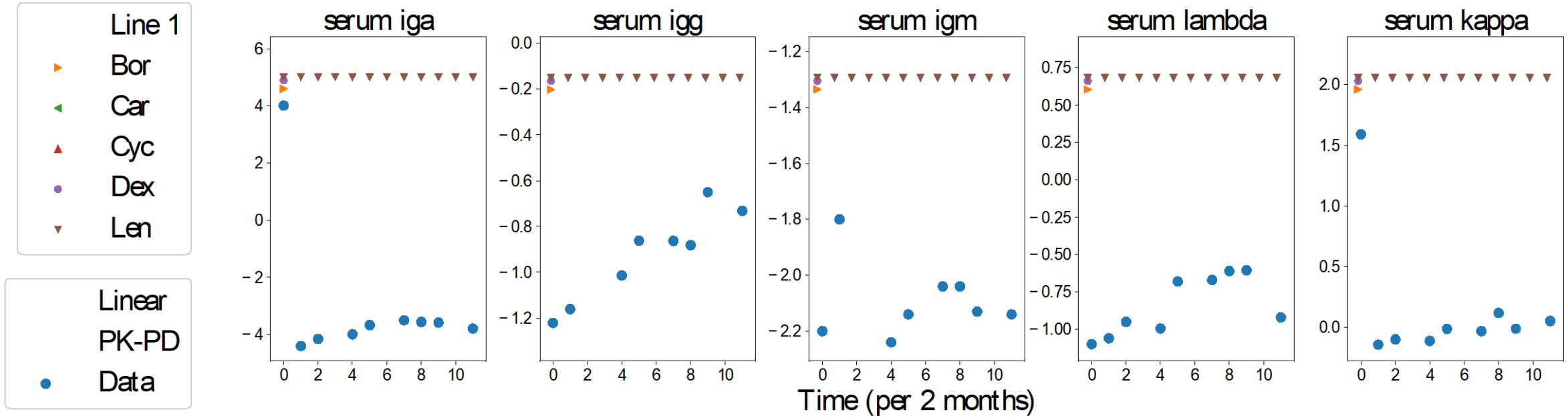
What utility do the clocks have?



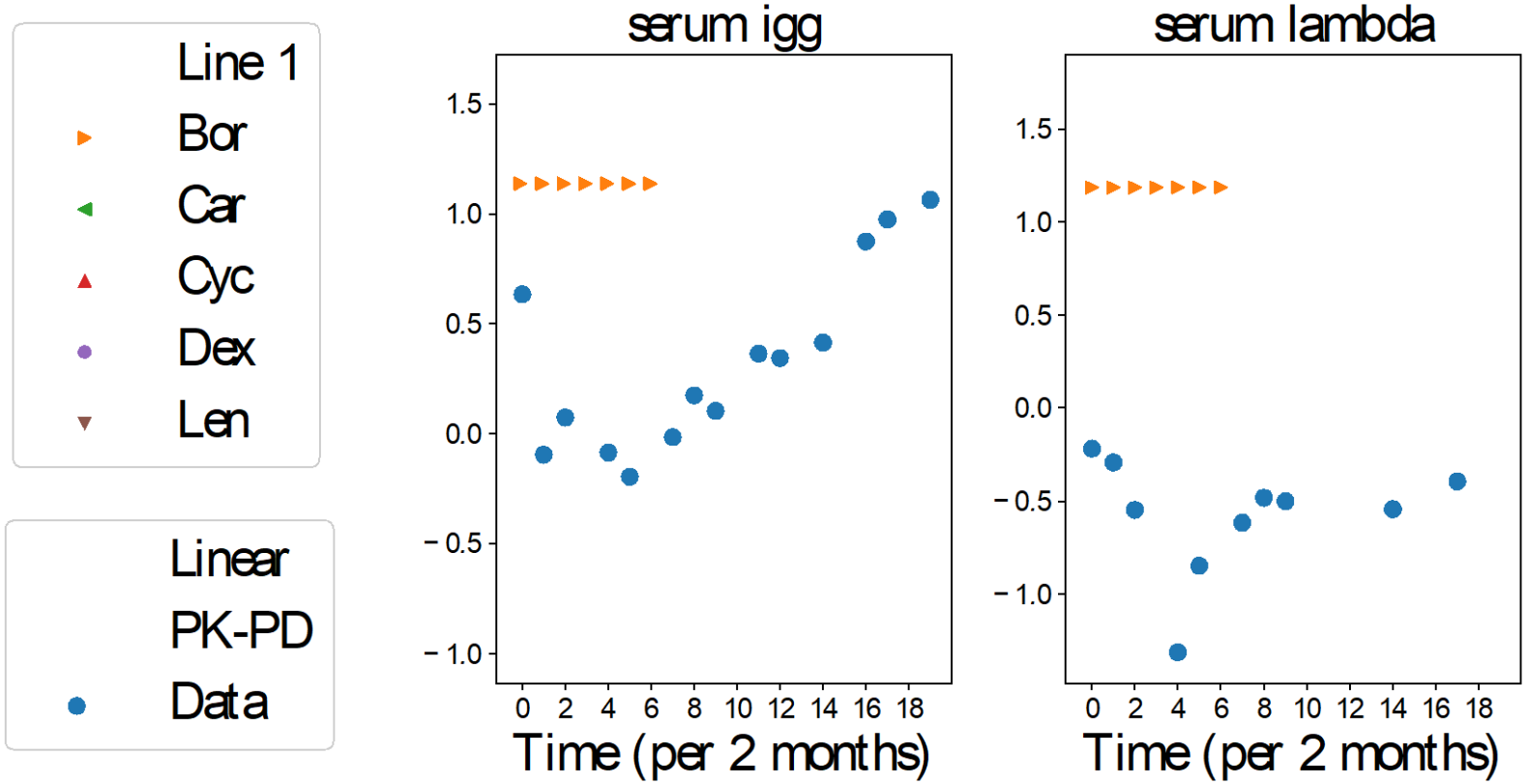
Introspection into the learned latent space



Predicting clinical biomarkers into the future (using baseline data)



Predicting clinical biomarkers into the future (after observing the patient for 15 months)



Conclusion &

Opportunities in research

Conclusion - Idea in a slide

- **Goal**:
  - Conditional density estimation of non-linear time-varying data
- **Challenge:**
  - Data is scarce and missing,
  - Traditional methods overfit or are insufficiently expressive
- **Approach:**
  - **Key idea:** Incorporate domain knowledge in how interventions affect latent representations to improve generalization
    - Pharmacodynamic modeling -> Neural Intervention Effect Functions
    - Treatment protocols -> local and global clocks
  - **PKPD-SSM:** Neural pharmacodynamic state space model
- **Consequence:**
  - Improvements in model's ability to generalize and forecast clinical biomarkers

Conclusion – take aways

- When applying deep generative models to real data, think deeply about and incorporating domain knowledge
- Incorporating structure of the problem into the model can improve generalization (especially when data is scarce)

Future Work

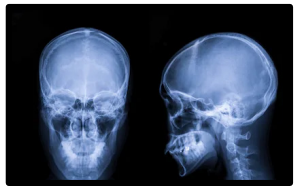
- Validating results in a larger, independent cohort
  - Working with collaborators to study the model on data from Veteran's Affairs (VA)
- Developing clinical decision support tools
  - Understanding the needs of oncologists when treating patients and what forecasting tasks might be of interest
- From predictive to counterfactual models
  - Use as a starting point for model-based reinforcement learning

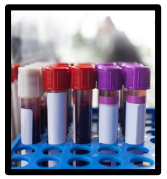
Opportunities in research Multi-modal decision making in oncology

Clinical notes

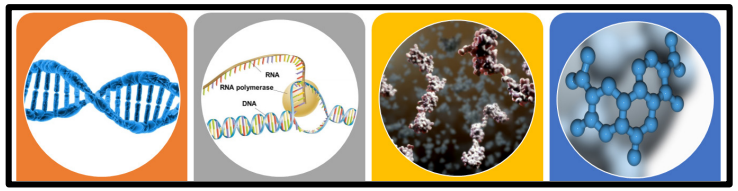
Patient found of floor at commencement of shift. Had climbed out of bed and hit head. Assisted back to bed. Obs stable. Cut above right eye – steri strips in place. Dr attended and sutured x3 to laceration on scalp. Very drowsy, unable to take meds due to drowsiness. Very poor fluid intake. ?may require IV therapy?

Imaging





Lab tests



Genetics

- Forecasting patient data
- Predicting time to progression
- Likelihood of successful treatment
- Disease sub-typing

Questions?