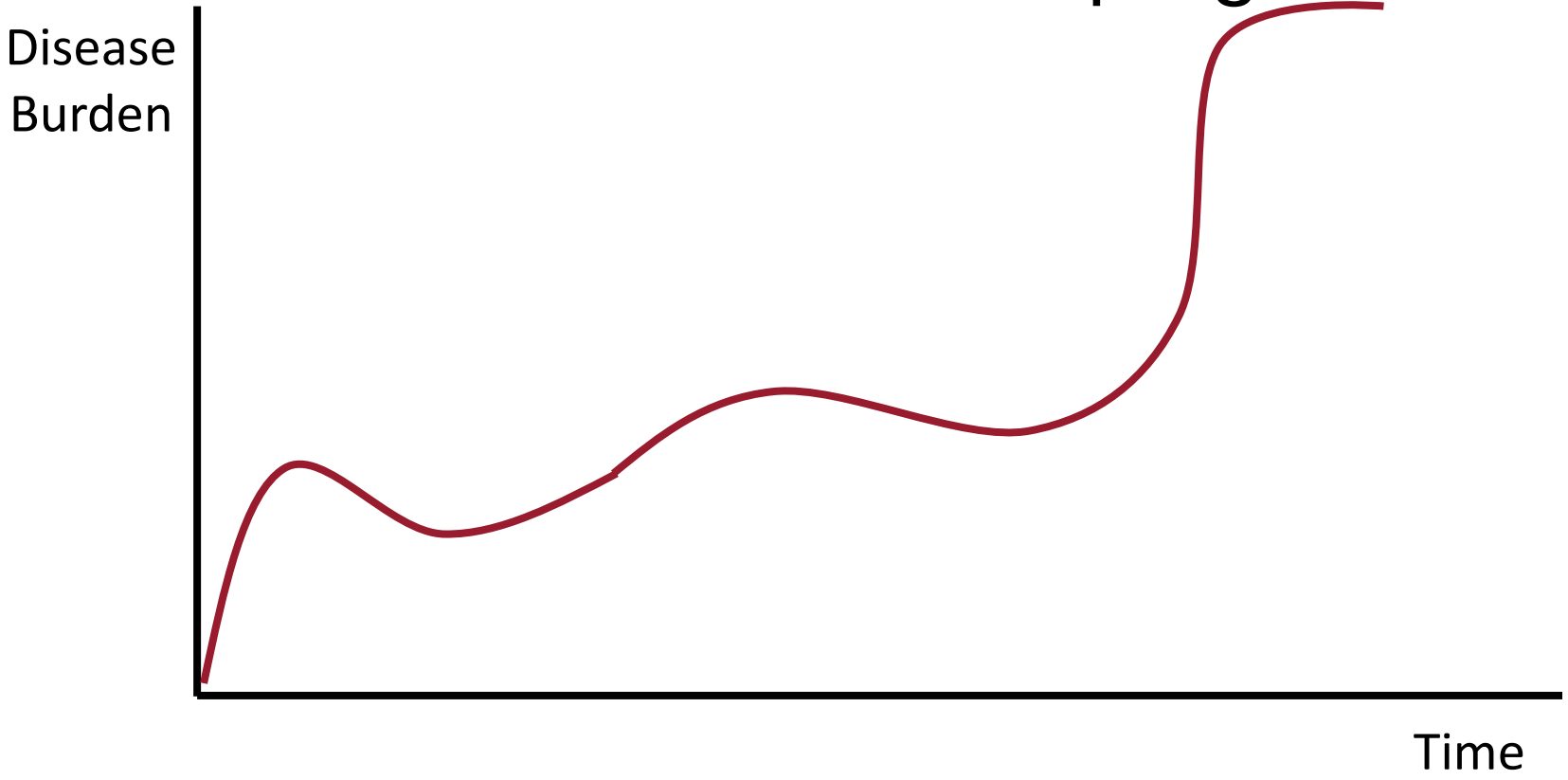
Disease Progression Modeling

Irene Y. Chen
PhD Student, Electrical Engineering and Computer Science

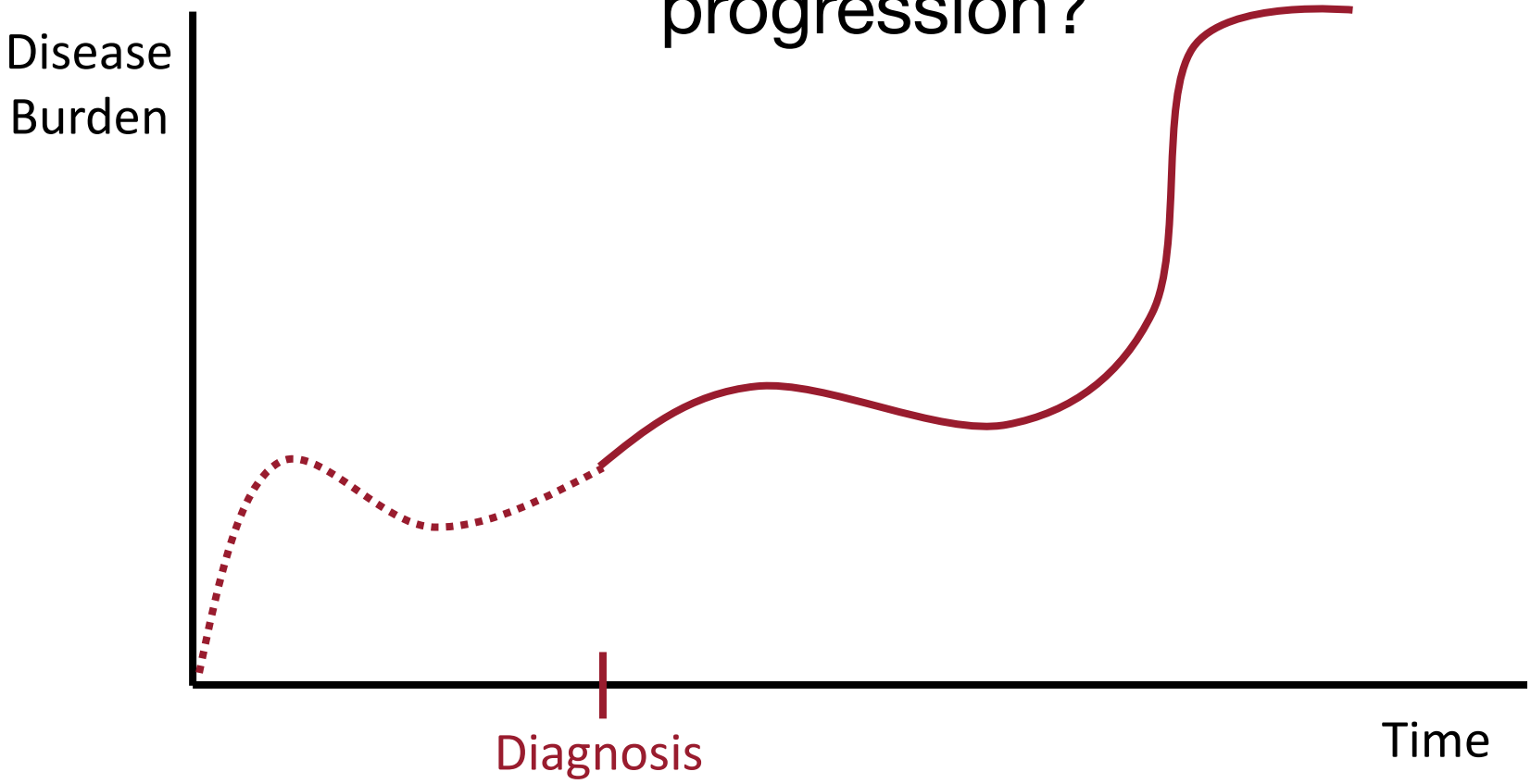
MIT Clinical ML
www.clinicalml.org

How can we learn how a disease will progress?

When will a disease progress?



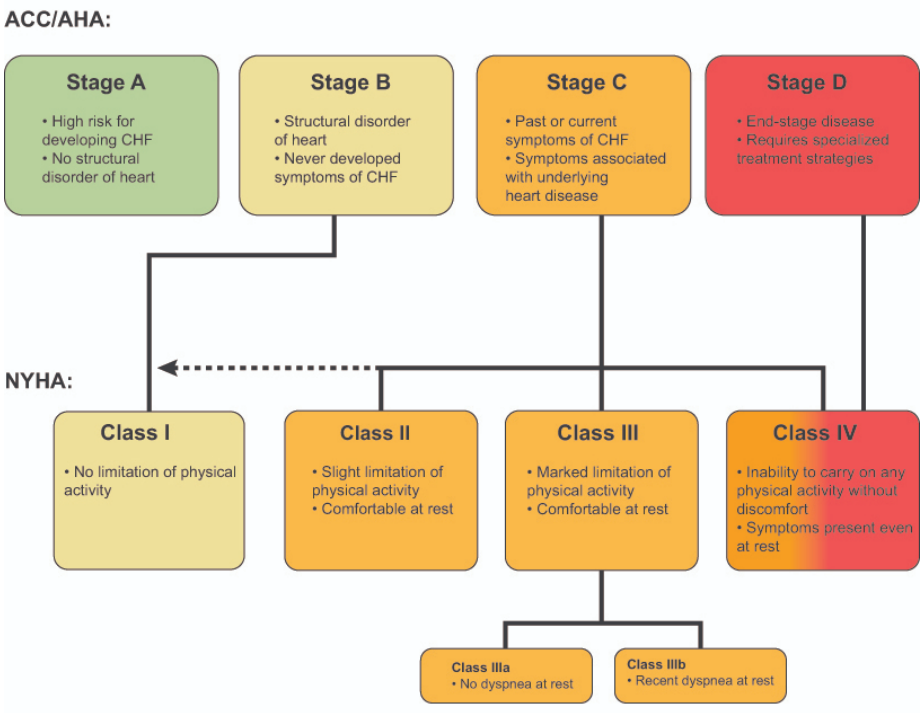
Where is the patient in the disease progression?



What is disease burden?

- Symptoms
  - Cognitive ability
  - Site specific pain
  - Function loss
- Biomarkers
  - Vital signs, e.g. blood pressure
  - Organ health, e.g. brain volume
  - Antibody levels in blood
- **Question**: does disease burden only increase?

What exists clinically now?



GrepMed, Heart Failure Staging

What exists clinically now?

	Early PD		Mid-stage PD	Advanced PD	
Stage of Parkinson's Disease	**1**	**2**	**3**	**4**	**5**
Severity of Symptoms	**MILD** Symptoms of PD are mild and only seen on one side of the body (unilateral involvement)	**MILD** Symptoms of PD on both sides of the body (bilateral involvement) or at the midline	**MODERATE** Symptoms of PD are characterized by loss of balance and slowness of movement	**SEVERE** Symptoms of PD are severely disabling	**SEVERE** Symptoms of PD are severe and are characterized by an inability to rise
	SYMPTOMS Tremor of one hand Rigidity Clumsy Leg One side of the face may be affected, impacting the expression	**SYMPTOMS** Loss of facial expression on both sides Decreased blinking Speech abnormalities Rigidity of the muscles in the trunk	**SYMPTOMS** Balance is compromised Inability to make the rapid, automatic and involuntary adjustments All other symptoms of PD are present	**SYMPTOMS** Patients may be able to walk and stand unassisted, but they are noticeably incapacitated Patient is unable to live an independent life and needs assistance	**SYMPTOMS** Patients fall when standing or turning May freeze or stumble when walking Hallucinations or delusions.

Parkinsonsdisease.net

What data could we have access to?

- Longitudinal vs cross-sectional
  - UK Biobank (cross-sectional)
  - Electronic health records (cross-sectional OR longitudinal)
  - Insurance claims (longitudinal)
  - Disease registries (longitudinal)
- Multimodal
  - Clinical biomarkers, medical imaging, clinical notes, etc

Today's talk

- What is disease progression?
- **Three approaches to disease progression**
- What could go wrong?
- Pop quiz

Approach 1: Supervised learning

- **Goal**: Predict disease status for 6, 12, 18, and 24 months separately.
- **Challenge**:
  - Separate prediction tasks
  - Assumes constant measurement
  - Labels are very noisy
  - Fewer time points as time progresses

Approach 1: Supervised learning

- **Goal**: Predict disease status for 6, 12, 18, and 24 months separately.

Number of patients M months after baseline
(Alzheimer's Disease Neuroimaging Initiative)

M06	M12	M24	M36	M48
648	642	569	389	87

M06 = 6 months after baseline

Zhou et al, "A multi-task learning formulation …", KDD 2012.

Approach 2: Multi-task learning

- **Goal**: Predict disease status for 6, 12, 18, and 24 months jointly.
- **Idea**: Treat problem as a *multi-task* learning problem where learning for 12 months would impact learning for 18 months.
  - Use common biomarkers across all time
  - Allow for specific biomarkers at specific times
  - Incorporate temporal smoothing

Convex fused sparse group lasso

- Simultaneously learn multiple outputs by solving

$$\min_W \|XW - Y\|_F^2 + \theta_1 \|W\|_F^2 + \theta_2 \sum_{i=1}^{t-1} \left\| w^i - w^{i+1} \right\|_2^2$$

squared loss between outcomes Y, learned weights W, and biomarkers X

regularize weights

temporal smoothing between times

Zhou et al, "A multi-task learning formulation …", KDD 2012.

Convex fused sparse group lasso

- Simultaneously learn multiple outputs by solving

$$\min_W \|S \odot (XW - Y)\|_F^2 + \theta_1 \|W\|_F^2 + \theta_2 \|WH\|_F^2 + \delta \|W\|_{2,1}$$

Allow for missing values with mask S

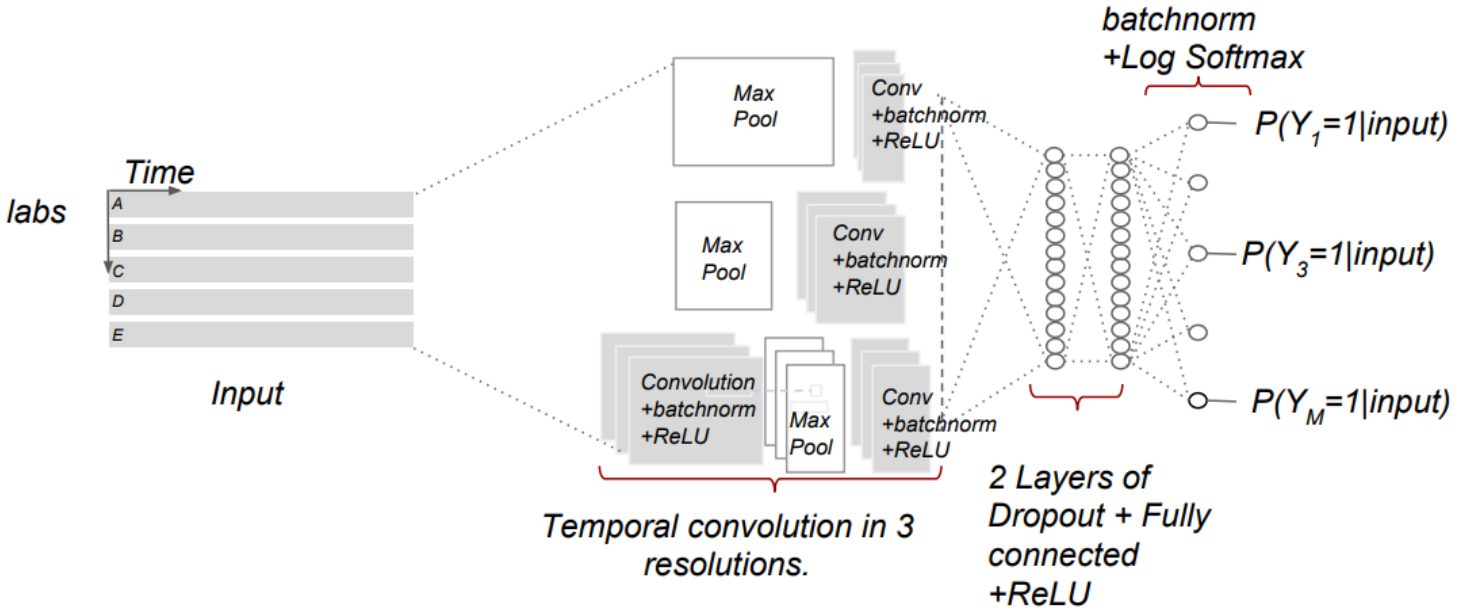
regularize weights

temporal smoothing between times

Additional regularization for reducing dim of data

Zhou et al, "A multi-task learning formulation …", KDD 2012.

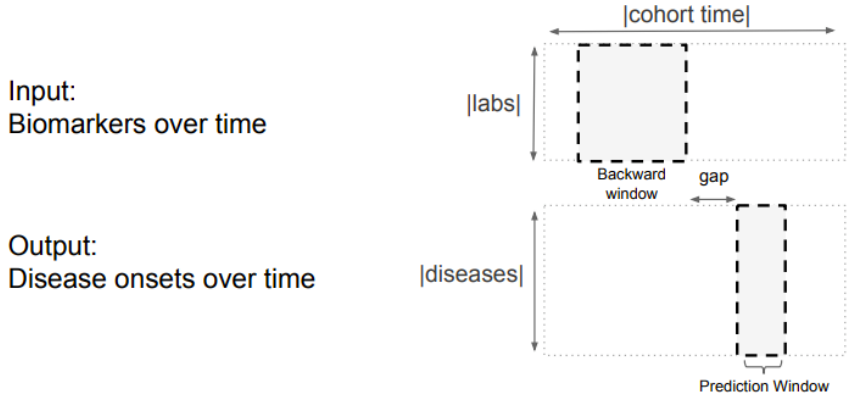
Multi-task prediction for disease and time



Razavian et al, “Multi-task Prediction of Disease Onsets from Longitudinal Lab Tests,” 2016.

Multi-task prediction for disease and time

- **Data**: Longitudinal lab test values from insurance claims
- **Goal:** Early diagnosis across diseases for a fixed future time window



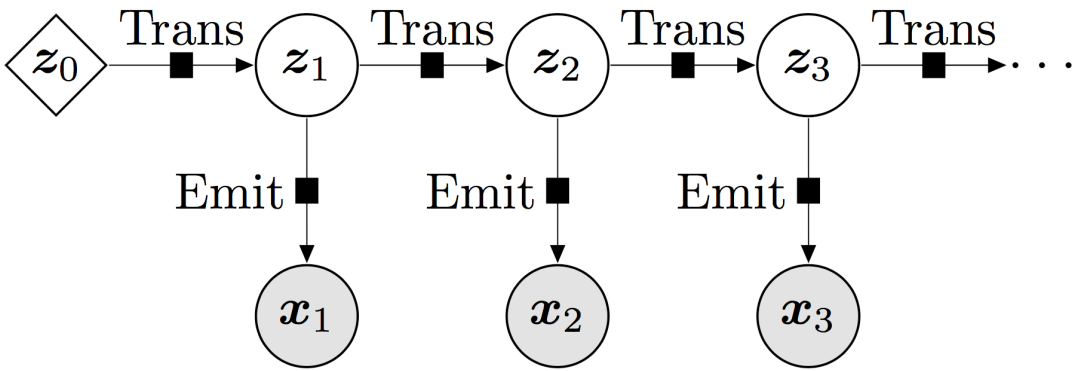
Razavian et al, “Multi-task Prediction of Disease Onsets from Longitudinal Lab Tests,” 2016.

Approach 3: Unsupervised learning

- **Goal**: What if we learned continuously without specifying outcome of interest?
- **Idea**: How can we use *unsupervised learning* to find patterns in data for more robust learning
  - Hidden Markov Models
  - Recurrent Neural Networks
  - Single cell biology

Approach 3a: Hidden Markov Models

- **Goal**: We can model our data with a HMM



- $p(x_{123}, z_{123}) = p(x_1|z_1)\ p(x_2|z_2)\ p(x_3|z_3)\ p(z_3|z_2)\ p(z_2|z_1)\ p(z_1|z_0)$

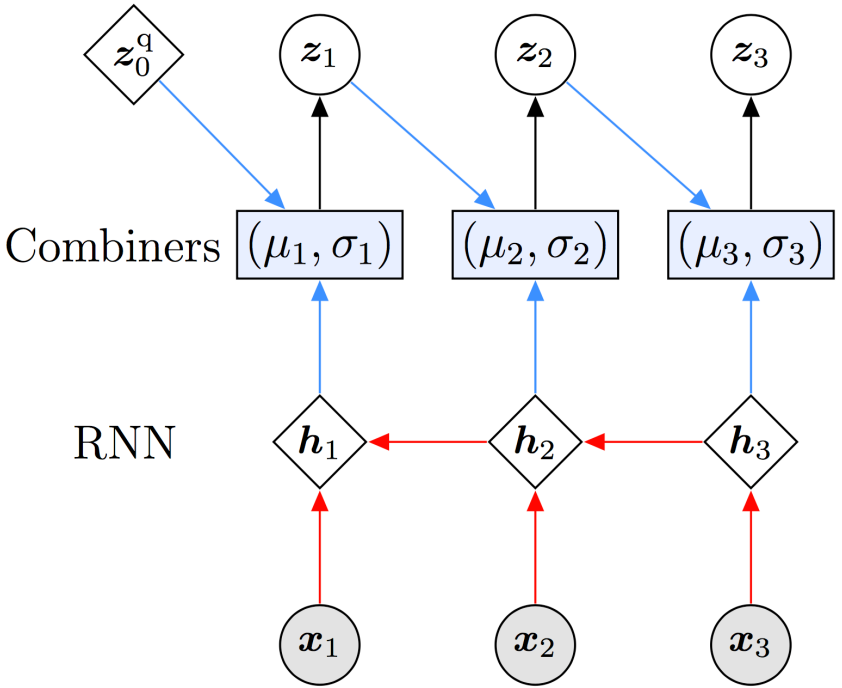
Krishnan et al, "Structured Inference Networks for Nonlinear State Space Models," 2016.

Approach 3a: Hidden Markov Models

- **Idea**: We use variational inference to learn single parametric function $f(\boldsymbol{x})$ for variational distribution

$$q(z|f(x))$$

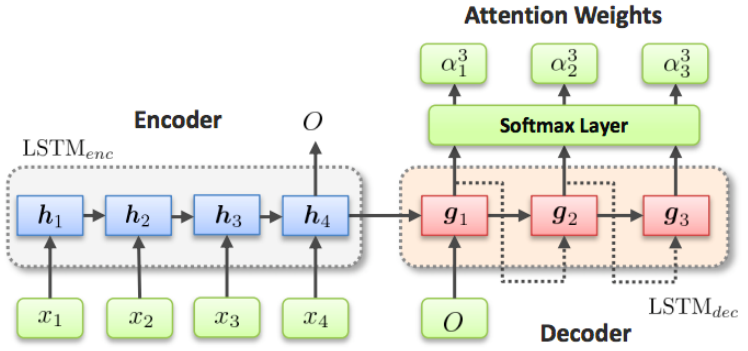
- You can run a RNN backwards and use the hidden states

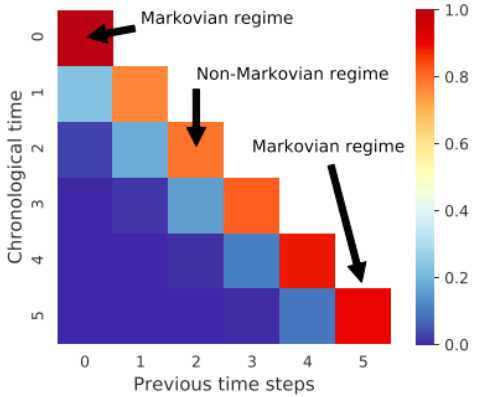


Krishnan et al, "Structured Inference Networks for Nonlinear State Space Models," 2016.

Approach 3b: Recurrent NNs

- **Goal**: learn “memoryful” dynamics with attentive state space
- **Idea**: progression from a long time ago could impact future disease state

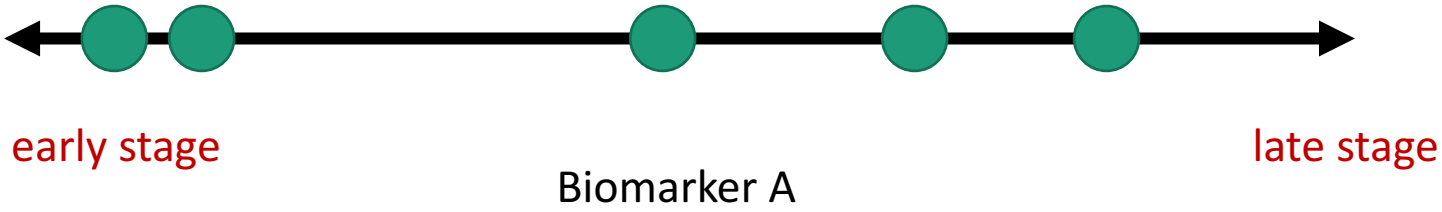




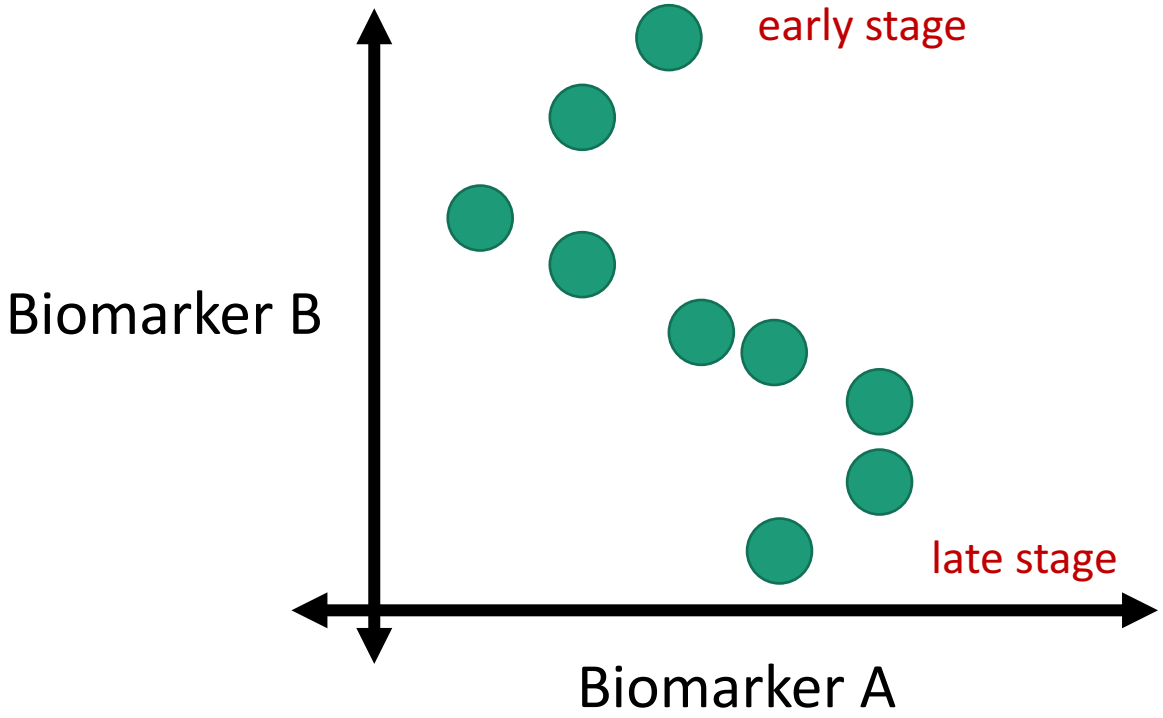
Alaa and van der Schaar, “Attentive State Space Modeling of Disease Progression,” 2019.

Approach 3c: Single Cell Biology

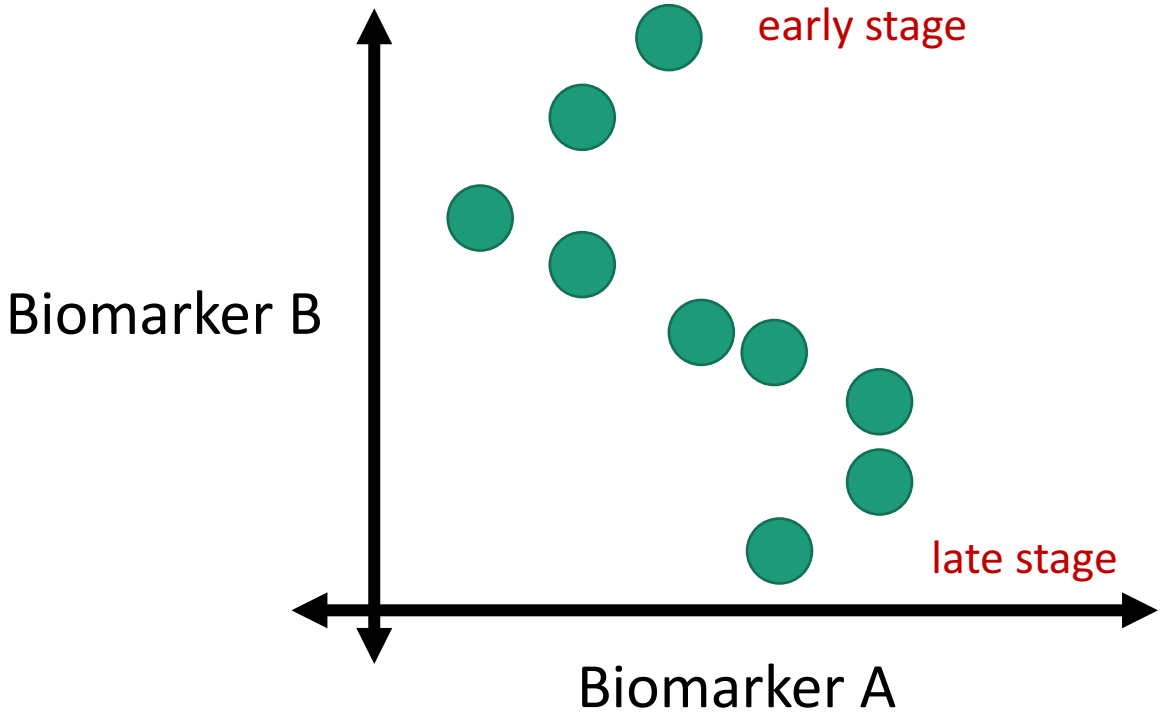
- **Goal**: How can we learn from cross-sectional data?
- **Idea**: If we observe enough data across all stages, we can learn alignment.
- For 1-D case for a meaningful biomarker, we can place values across a line.



Approach 3c: Single Cell Biology

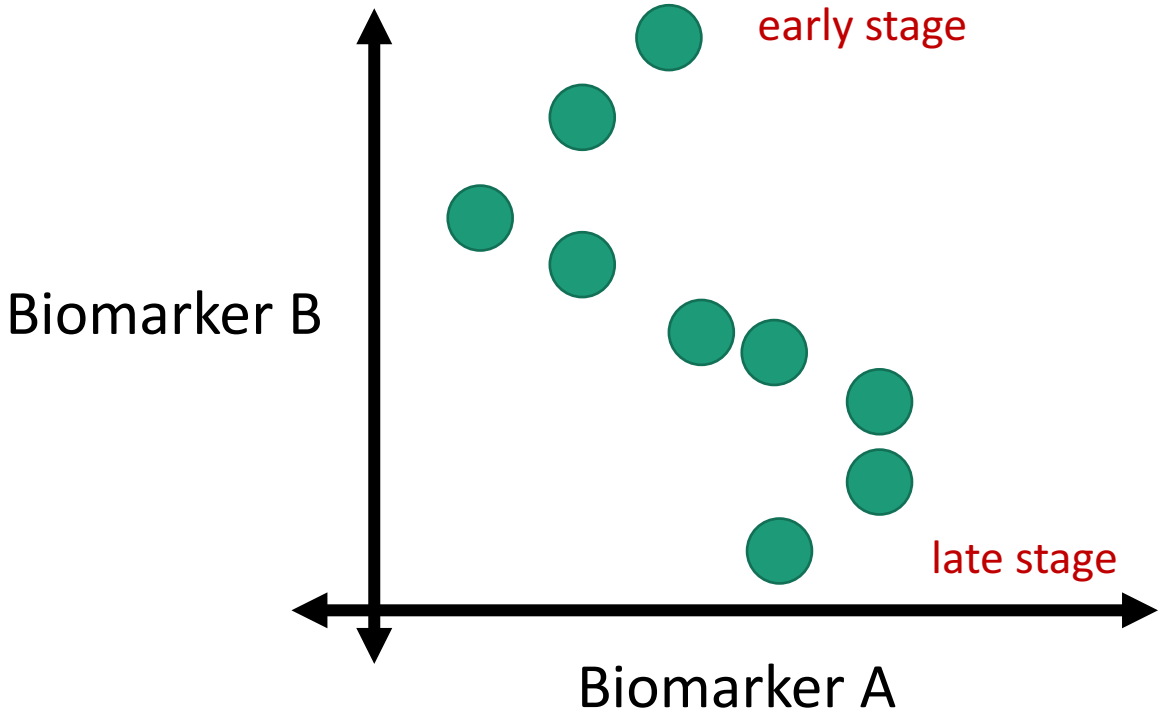


Approach 3c: Single Cell Biology



- With enough data, we can recognize structure
- Sequential data from same patient may help

Approach 3c: Single Cell Biology



- What assumptions are we making?
- What if we have many dimensions?
- What if we have many dimensions?

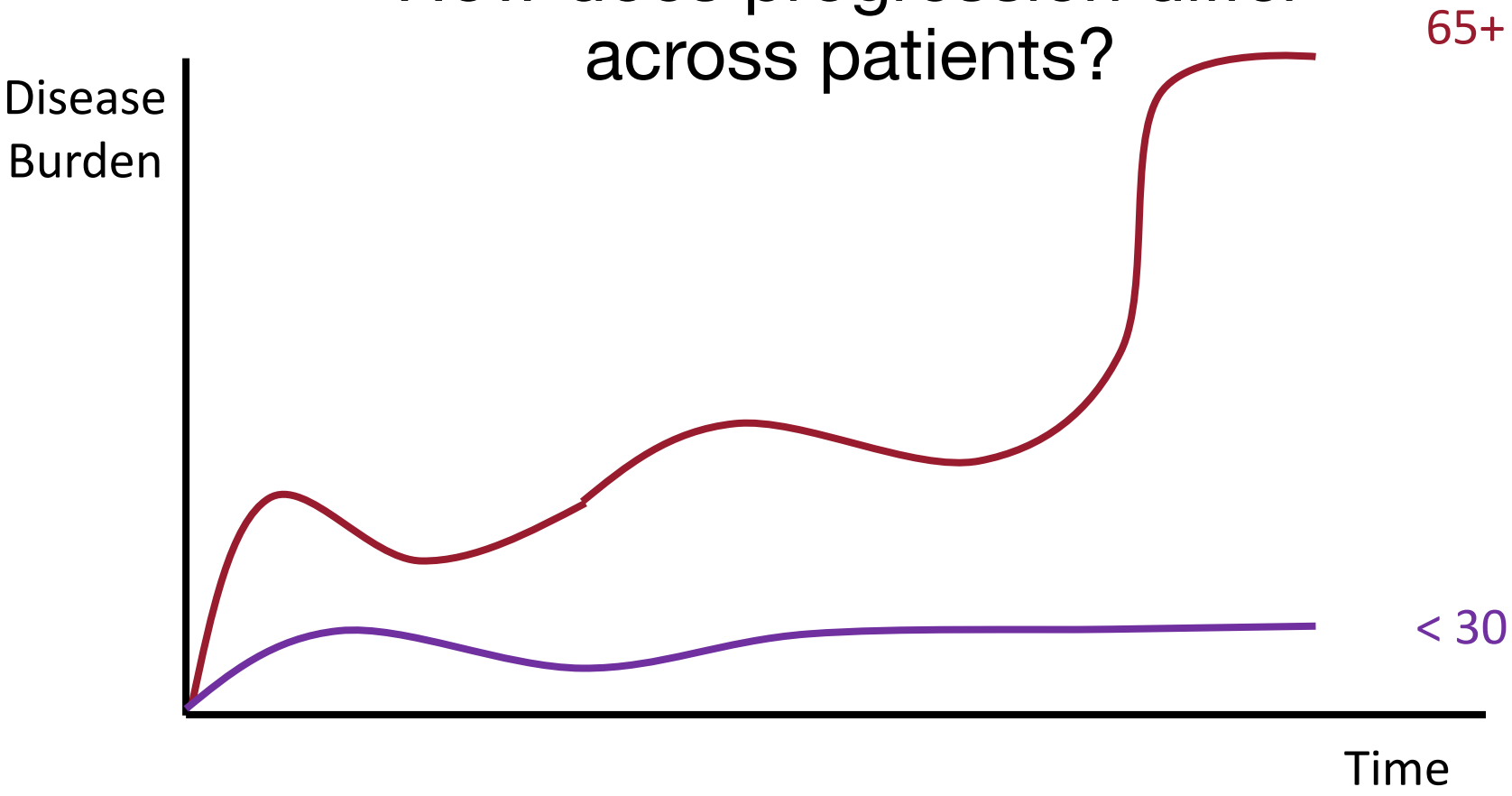
Today’s talk

- What is disease progression?
- Three approaches to disease progression
- **What could go wrong?**
- Quiz time

What are potential complications?

- Subtypes
- Treatment policies from clinicians
- Misaligned / censored data
- Non-stationarity

How does progression differ across patients?



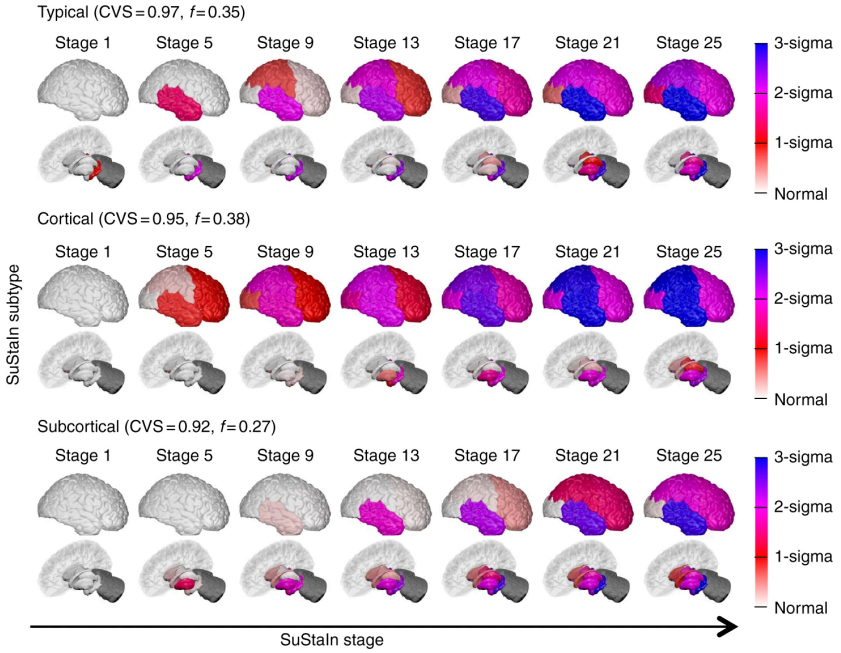
What if there are different subtypes?

- **Asthma**: 1) transient early wheezers, 2) persistent wheezers, 3) late onset wheezers
- **Autism:** 1) seizures, 2) gastrointestinal, 3) psychiatric, and 3) unknown.
- **Heart failure**: 1) reduced ejection fraction and 2) preserved ejection fraction (three types as well)
- *Challenge*: how do we separate subtype and progression?

[1] Deliu et al; *Pulmonary Therapy* 2016
[2] Doshi-Velez et al; *Pediatrics*, 2014.
[3] Shah et al; *Circulation*, 2016.

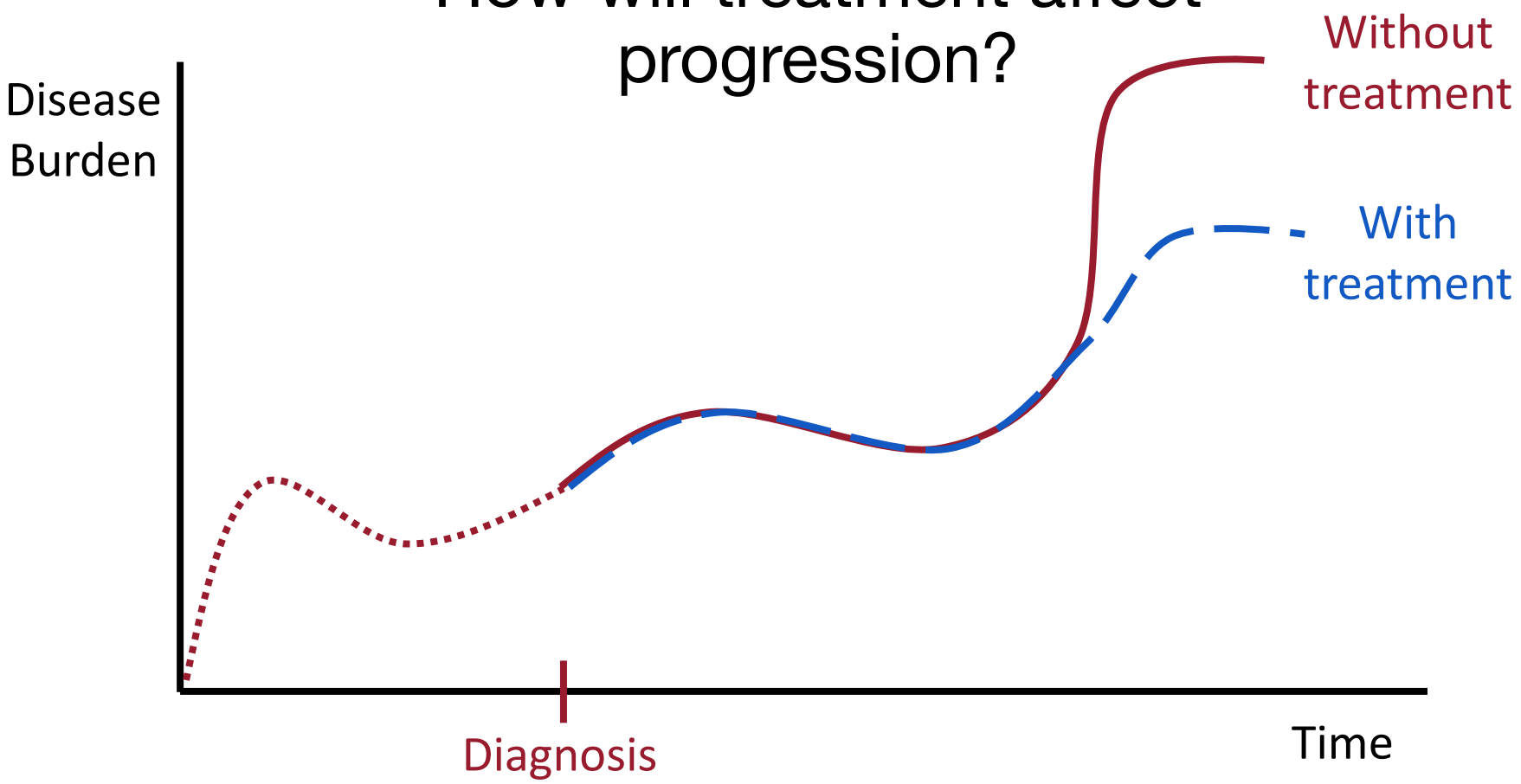
Can we learn stage and subtype jointly?

- Learn stages and subtypes of Alzheimer's disease
- Assume piecewise linear functions for separate subtypes
- Infer latent parameters through MCMC



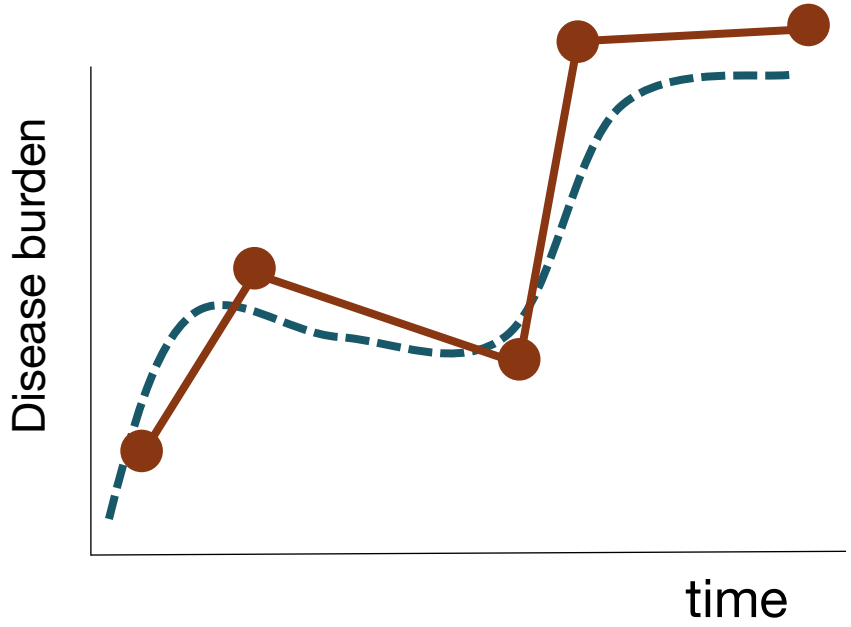
Young et al, "Uncovering the heterogeneity …", Nature Communications 2018.

How will treatment affect progression?



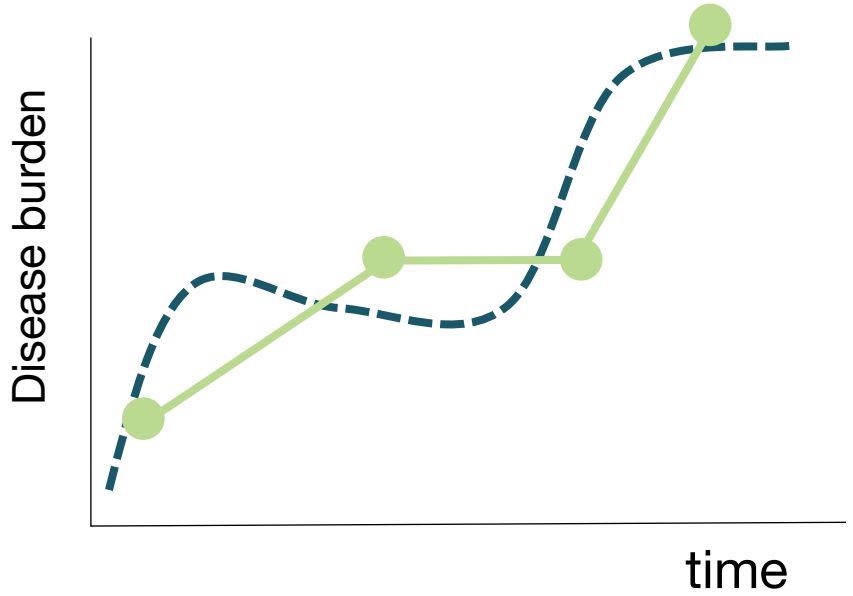
Mis-aligned / censored data

- **Problem:** Patients may enter the system at different times
  - Access to care
  - Switch hospitals so records begin in the middle of progression
- **Problem:** Patients may leave the system
  - If we align by death: not enough data
  - Patients can also leave system without defined labels about outcome



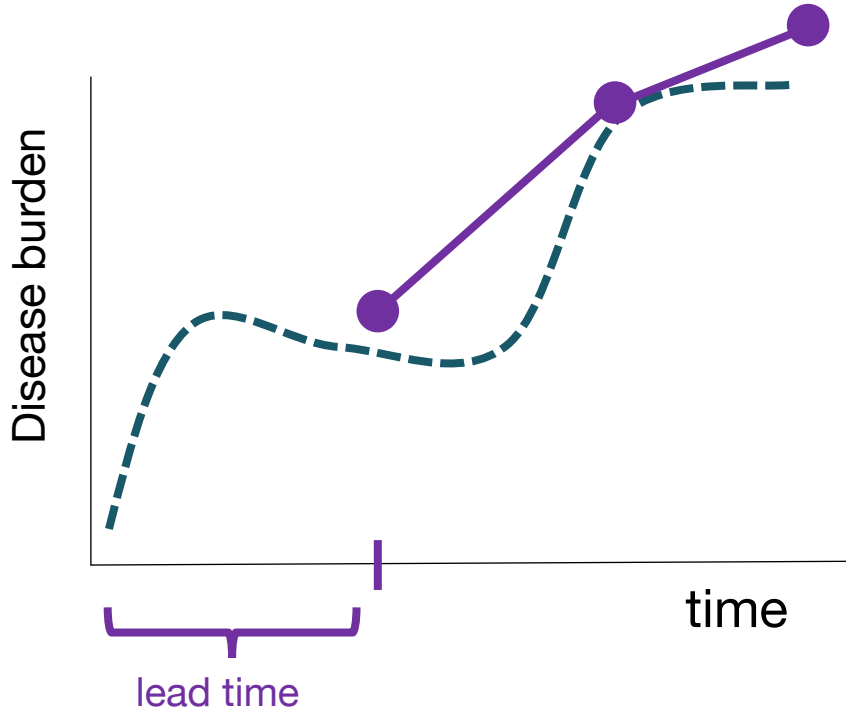
Mis-aligned / censored data

- **Problem:** Patients may enter the system at different times
  - Access to care
  - Switch hospitals so records begin in the middle of progression
- **Problem:** Patients may leave the system
  - If we align by death: not enough data
  - Patients can also leave system without defined labels about outcome



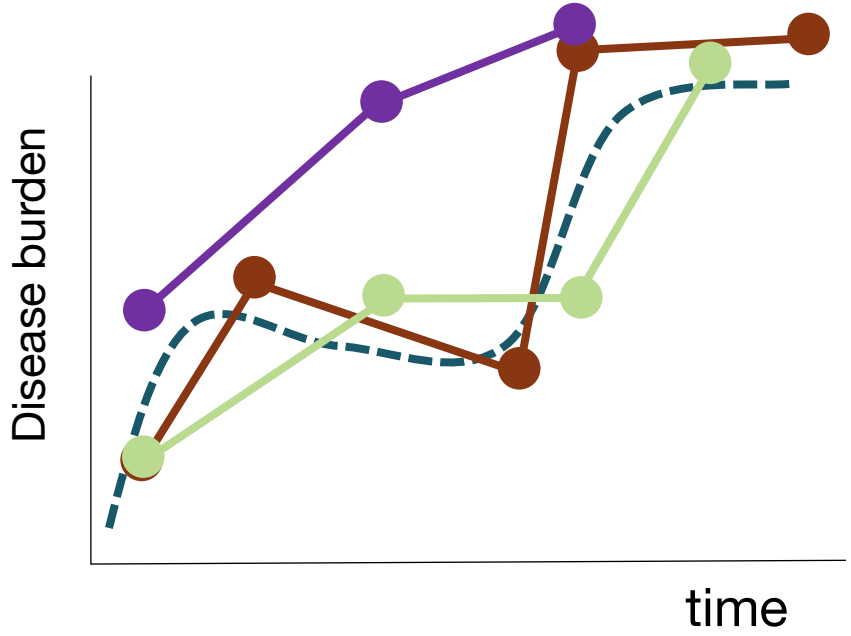
Mis-aligned / censored data

- **Problem:** Patients may enter the system at different times
  - Access to care
  - Switch hospitals so records begin in the middle of progression
- **Problem:** Patients may leave the system
  - If we align by death: not enough data
  - Patients can also leave system without defined labels about outcome



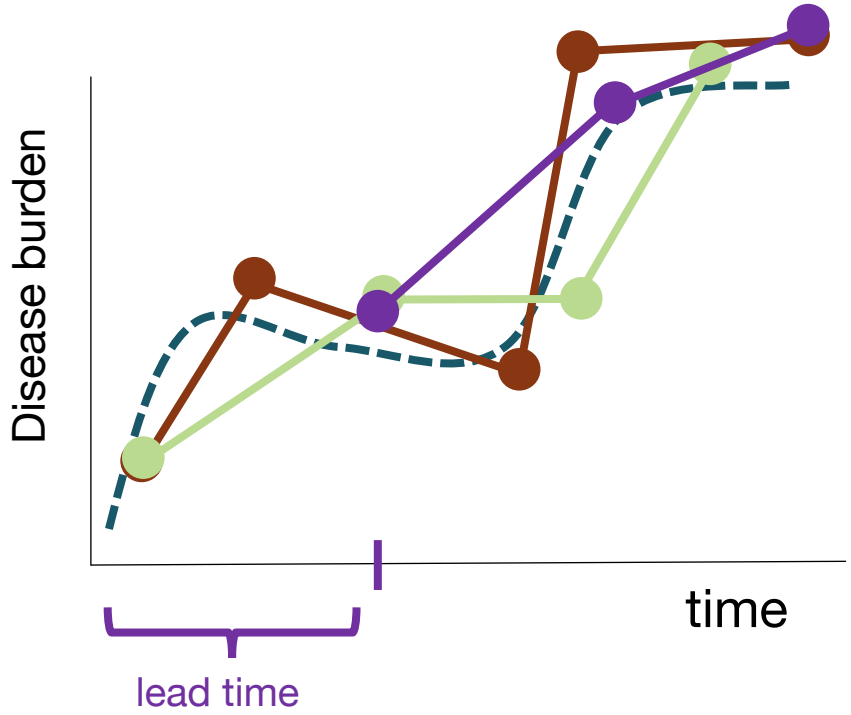
Mis-aligned / censored data

- **Problem:** Patients may enter the system at different times
  - Access to care
  - Switch hospitals so records begin in the middle of progression
- **Problem:** Patients may leave the system
  - If we align by death: not enough data
  - Patients can also leave system without defined labels about outcome



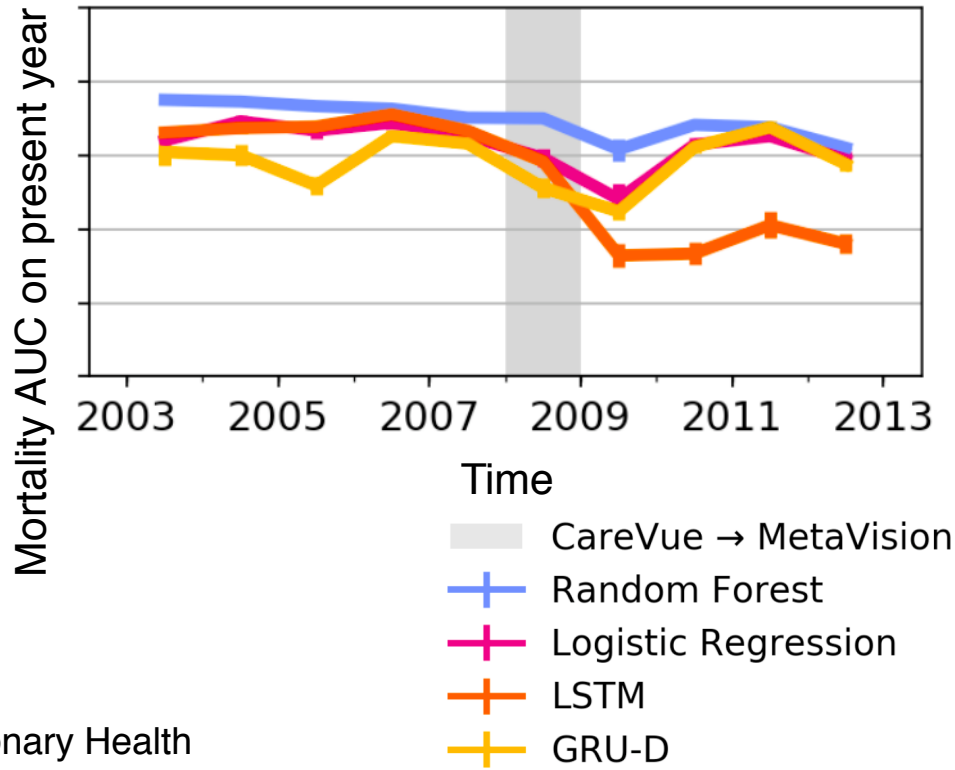
Mis-aligned / censored data

- **Problem:** Patients may enter the system at different times
  - Access to care
  - Switch hospitals so records begin in the middle of progression
- **Problem:** Patients may leave the system
  - If we align by death: not enough data
  - Patients can also leave system without defined labels about outcome



Non-stationarity: The future is the ultimate confounder

- Predict **MIMIC-III** patient mortality based on 181 lab and vitals, aggregated based on clinical domain knowledge
- Train on all **prior years**
- Model performance can **degrade** over time



Nestor et al, “Feature Robustness in Non-stationary Health Records,” 2019.

Questions

FAQs: Where is the deep learning?

FAQs: Where is the deep learning?

1. Any function can be fit with a deep net.
   - Supervised learning
   - Multi-task learning
   - Transition functions of Markov model
2. Learn low-dimensional representation and fit any model on top of that

FAQs: How do we measure success?

FAQs: How do we measure success?

1. Prediction tasks have accuracy metrics
2. Unsupervised learning have log-likelihood
3. Compare against clinical guidelines

FAQs: How is this same/different to RL?

FAQs: How is this same/different to RL?

1. If we assume all patients treated the same, we can ignore treatments entirely
2. In RL, we have rewards each time step (unless POMDP)
3. Disease progression modeled as RL may run into concerns about lack of decision support

Looking forward

- **Disease progression is a nail with many hammers.** Depending on clinical needs, we can model with great simplicity or great complexity.
- **There exit many pieces of the clinical puzzle.** We need to think critically about all components of clinical pipeline – making assumptions when needed for task.