Diagnosing Alzheimer's the Bayesian Way

Arya Pourzanjani, Benjamin Bales,
Michael Harrington, Linda Petzold

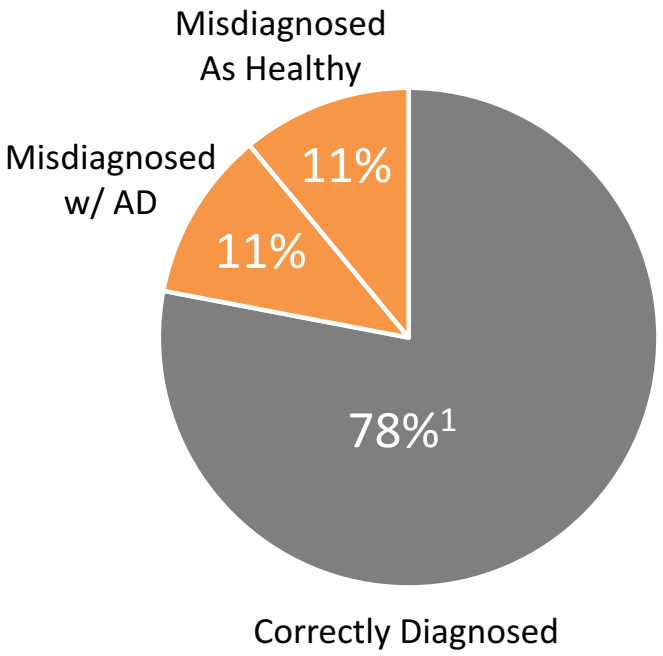

FIRE CO., INC.
SQUA

Alzheimer's Disease is ranked the most debilitating disease[1] because it primarily targets the mind

- memory loss
  erratic behavior, trouble remembering loved ones
- trouble performing daily functions
  like eating or bathing
- worsened moods
  depression and anxiety in patients and caretakers

[1] *Healthcare Business and Tech 2012*

We can help Alzheimer's patients now, but only if we can diagnosis accurately



Misdiagnosed patients with Alzheimer's miss out on memory-boosting therapies

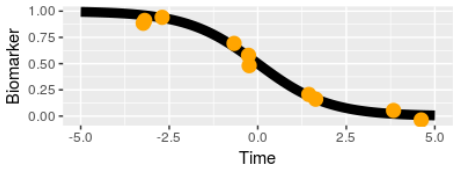
Misdiagnosed healthy patients can be subject to unnecessary treatments costing up to $14k/year[2]

New drugs are being developed, but we need a precise, accurate way to recruit patients and test them

[1]CBS News 2016
[2]Hunter et al. 2013

Diagnosis is currently done manually by physicians and is imprecise and subjective

Physicians take and examine patient biomarkers
blood tests, verbal memory tests, and MRI



Healthy patient biomarkers start at a baseline distribution then decay asymptotically towards a lower baseline distribution as the disease progresses

Physicians currently diagnose in to three classes based on available biomarkers

- imprecise, low granularity
- only looking at single snapshot of patient
- subjective, prone to inconsistency

We need to diagnose the Bayesian way

1. We need a generative model to describe how Alzheimer's and biomarkers progress

2. We need uncertainty estimates of a given patient's disease progression

3. We need multilevel models because every patient is different

To understand complex a process like Alzheimer's progression, we need complex generative models

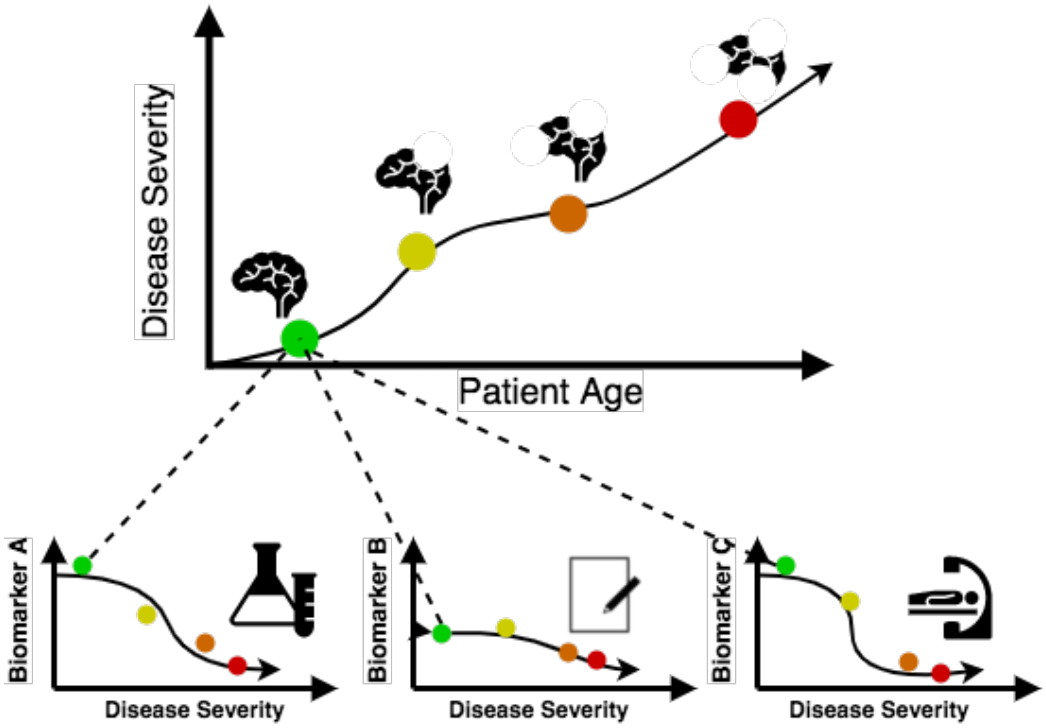
Disease progression curve for *i*th patient

$$\overbrace{d_i(t)}$$

$$\underbrace{y_{ij}(t)} \sim \mathcal{N}(\underbrace{f_j(d_i(t))}, \sigma_j^2)$$

Observation of *j*th biomarker for patient *i* at time *t*

4-parameter logistic function for *j*th biomarker evaluated at disease severity *i*



Jedynak et al. 2012

Generative modeling advances our understanding of the disease more so than simple classification

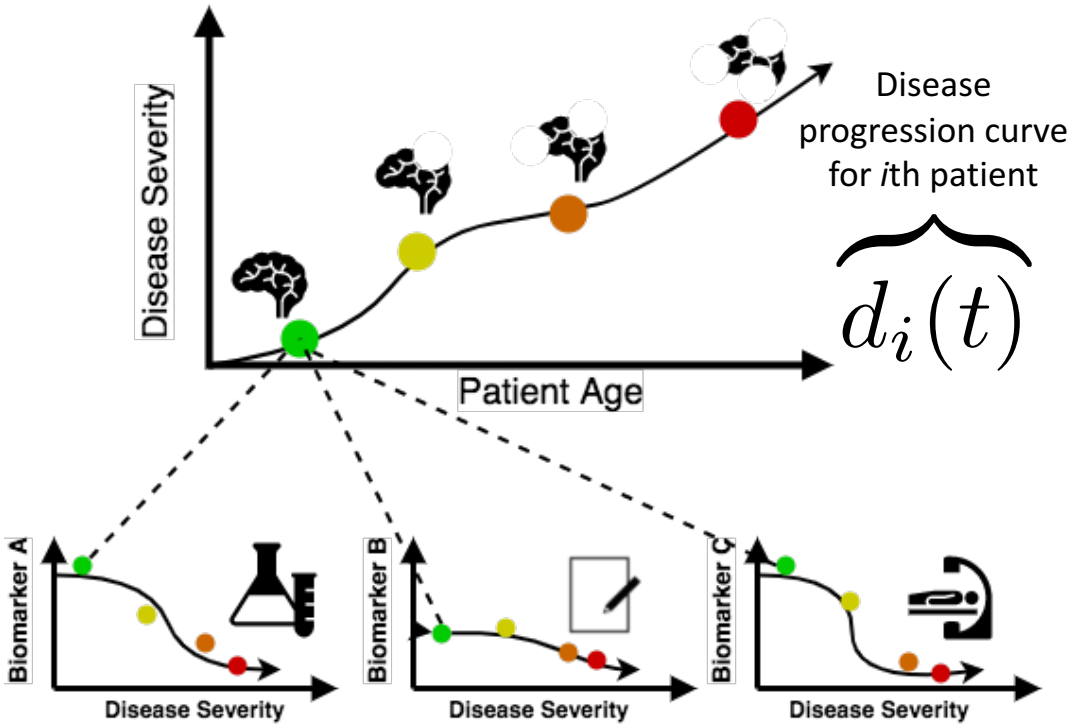
Capturing temporal progression tells us about different progression patterns

Logistic parameters are meaningful and teach us interpretable diagnostic criteria

They tell us when the biomarkers start degrading relative to one another.

They tell us how fast they will degrade.

$$\underbrace{y_{ij}(t)}_{\text{Observation of } j\text{th biomarker for patient } i \text{ at time } t} \sim \mathcal{N}\Big(\underbrace{f_j(d_i(t))}_{\frac{\alpha_j}{1+\exp(-\underbrace{\beta_j}_{\text{Biomarker progression rate relative to disease}} \cdot d_i(t) + \underbrace{\gamma_j}_{\text{Biomarker onset time}})} + \delta}, \sigma_j^2\Big)$$



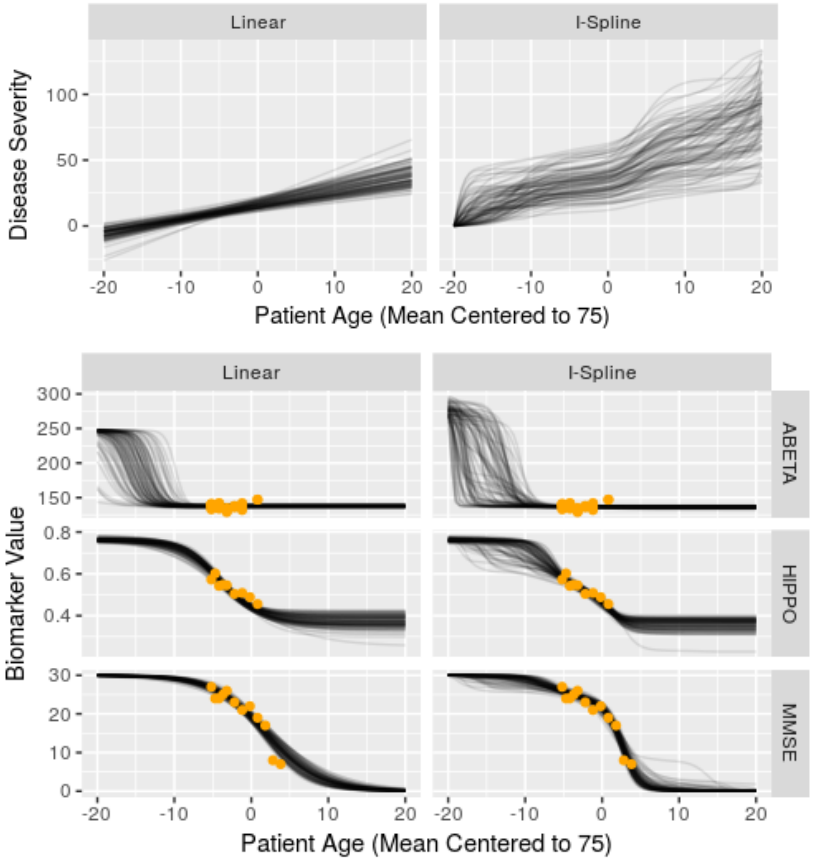
Jedynak et al. 2012

Expanding our model tells us when our understanding of the process we are modeling is too simplified

A nonlinear disease progression fit reveals when a patient's progression deviates from simple linear progression

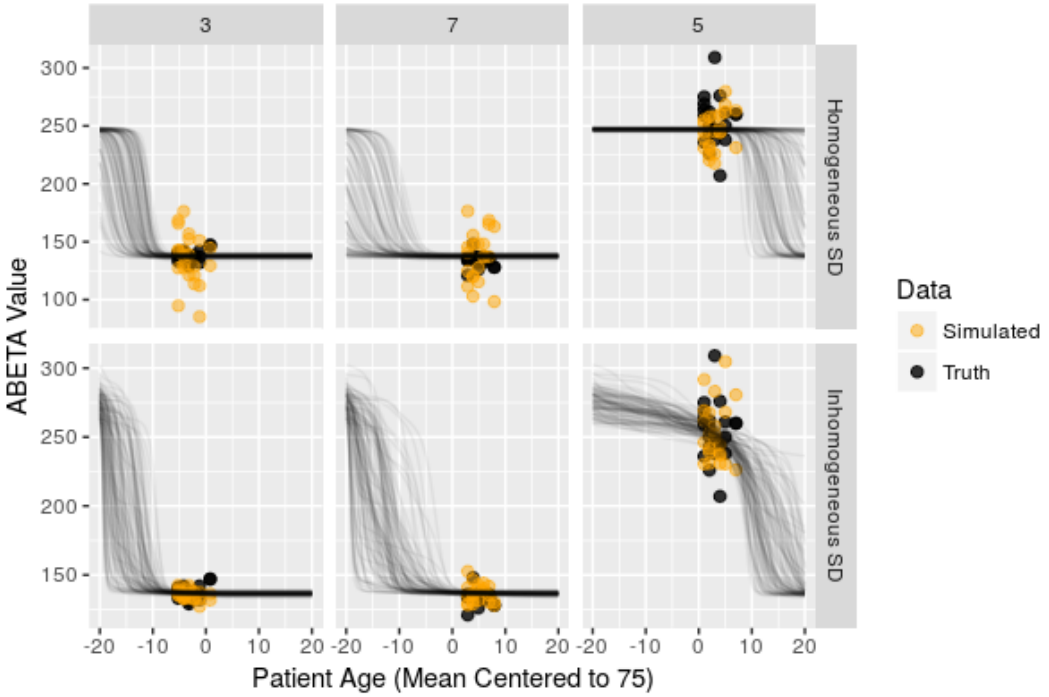
Knowing that a patient's disease severity is progressing nonlinearly provides insight in to the disease that physicians value

A step-like pattern is characteristic of vascular dementia



Simulating data from a model and comparing to real data reveals when our understanding is wrong

$$\underbrace{y_{ij}(t)}_{\text{Observation of } j\text{th biomarker for patient } i \text{ at time } t} \sim \mathcal{N}(\underbrace{f_j(d_i(t))}_{\text{4-parameter logistic function for } j\text{th biomarker evaluated at disease severity } i}, \sigma_j^2)$$



We need to diagnose the Bayesian way

1. We need a generative model to describe how Alzheimer's and biomarkers progress

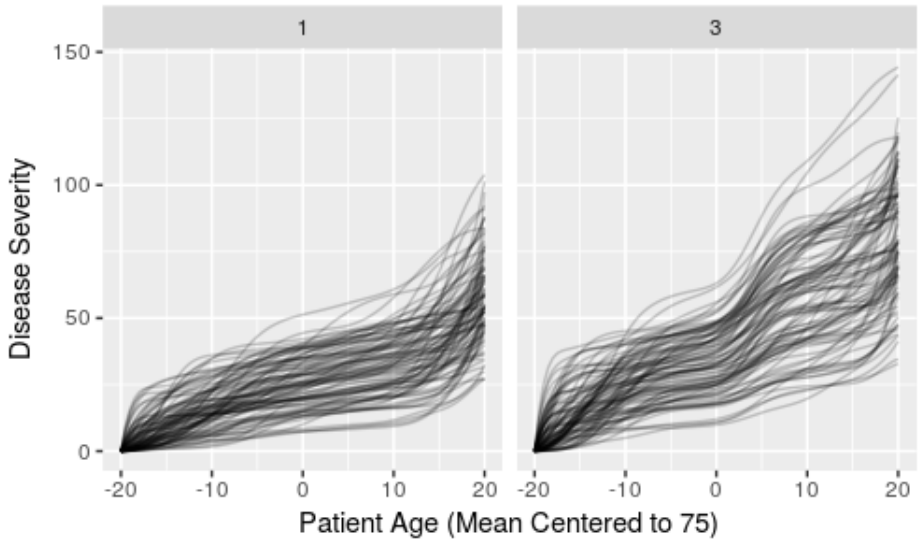
2. We need uncertainty estimates of a given patient's disease progression

3. We need multilevel models because every patient is different

Uncertainty is crucial in guiding treatment

Posterior disease progression curves represent the possible progression paths a patient could take given their observed biomarkers

If we know a patient has a low risk of progressing quickly we can be more cautious about administering new drugs with potential side-effects



Uncertainty estimates are necessary for making statistical claims about drug efficacy

The Bayesian framework and MCMC samples easily allow us to test the probability a drug works

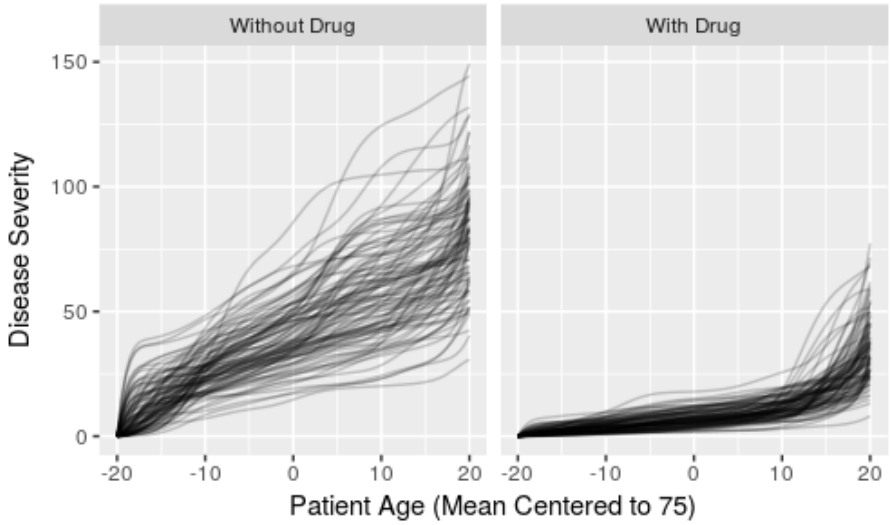
Disease severity w/ drug

Disease severity w/o drug

$$p\left(\overbrace{B(t)}^{\text{Disease severity w/ drug}} < \overbrace{A(t)}^{\text{Disease severity w/o drug}}\right)$$

$$= \mathbb{E}\left[I\left\{B^{\text{rep}}(t) < A^{\text{rep}}(t)\right\}\right]$$

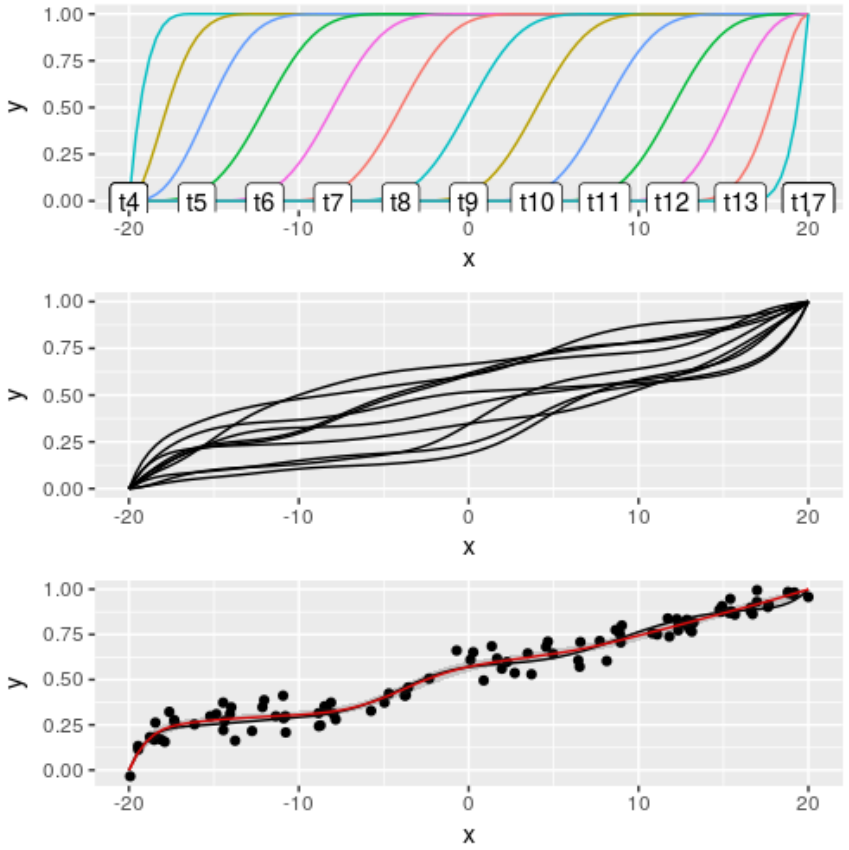
$$\approx \frac{1}{S}\sum_{s=1}^{S} I\left\{B_s^{\text{rep}}(t) < A_s^{\text{rep}}(t)\right\}$$



Capturing disease progression uncertainty is easy with I-Splines in Stan

```
data {
 int N;      //num obs
 int D;      //num splines
 real x[N];
 real y[N];
}
parameters {
 //increase func bounded by 1
 simplex[D] beta;

 //increase unbounded func
 //vector<lower=0>[D] beta;
}
model {
 matrix[N,D] Phi = ispline(x);
 y ~ normal(Phi*beta, 1);
}
```



We need to diagnose the Bayesian way

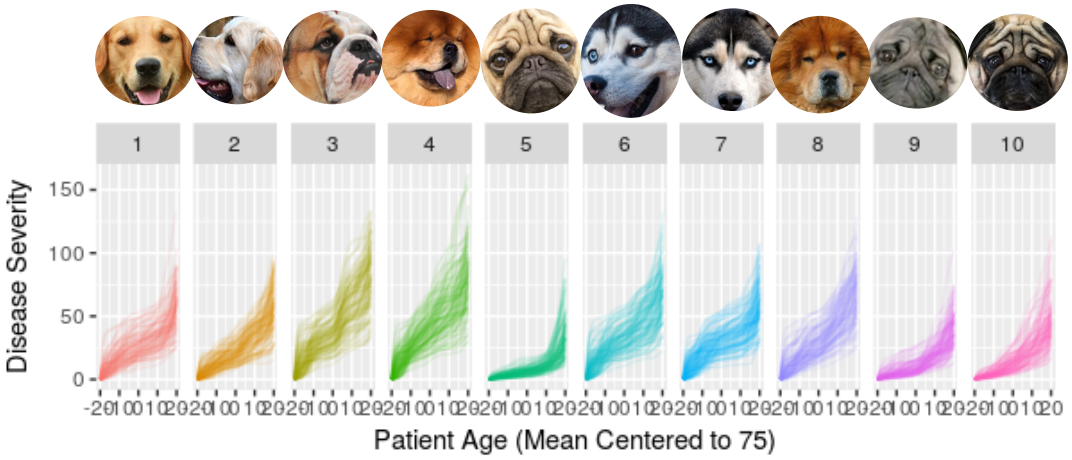
1. We need a generative model to describe how Alzheimer's and biomarkers progress

2. We need uncertainty estimates of a given patient's disease progression

3. We need multilevel models because every patient is different

No two patients have identical trajectories

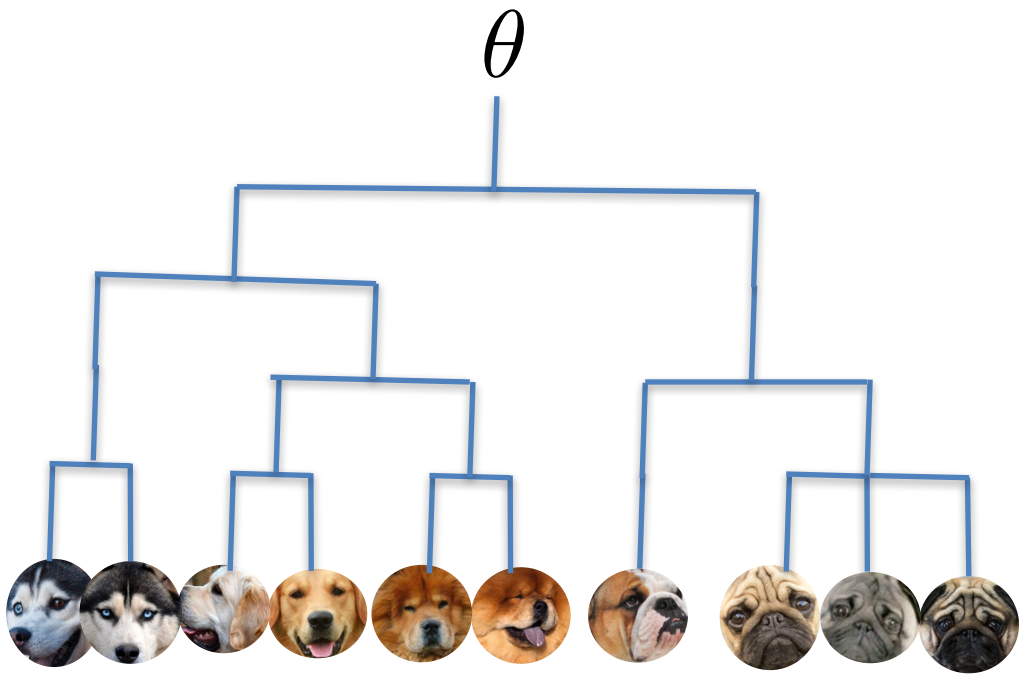
Every patient should have their own unique I-Spline coefficients



Patients should have unique trajectories,
but similar patients should be informative of one another

```
parameters {
 vector<lower=0>[D] theta;
 simplex[D] beta[N];
}
model {
 . . .
 beta ~ dirichlet(theta);
 . . .
}
```

Hierarchical priors allows us to leverage this similarity, and "borrow" information when biomarkers are sparse



We need to diagnose the Bayesian way

1. We need a generative model to describe how Alzheimer's and biomarkers progress

2. We need uncertainty estimates of a given patient's disease progression

3. We need multilevel models because every patient is different