Discovery of Parkinson's disease states using machine learning and longitudinal data

Kristen A. Severson[1], Lana Chahine[2], Luba Smolensky[3], Murtaza Dhuliawala[1],

Mark Frasier[3], Kenney Ng[1], Soumya Ghosh[1], Jianying Hu[1]

September 27, 2021

1. IBM Research, Center for Computational Health
2. University of Pittsburgh, Department of Neurology
3. Michael J. Fox Foundation

Why Machine Learning?

- Key challenges in PD
  - Incomplete biological understanding
  - Heterogeneity in symptom presentation and progression
  - Desire for personalized patient management and prognostication
  - Noisy, varied, and possibly qualitative measurements
- Key advantages of machine learning
  - Ability to discover complex patterns
  - Opportunity to incorporate prior knowledge
  - Power to synthesize large amounts of data
  - Possibility to incorporate uncertainty estimates

Disease progression modeling

Team:

IBM, Machine Learning

University of Pittsburgh, Neurology

Michael J. Fox Foundation, Advisory

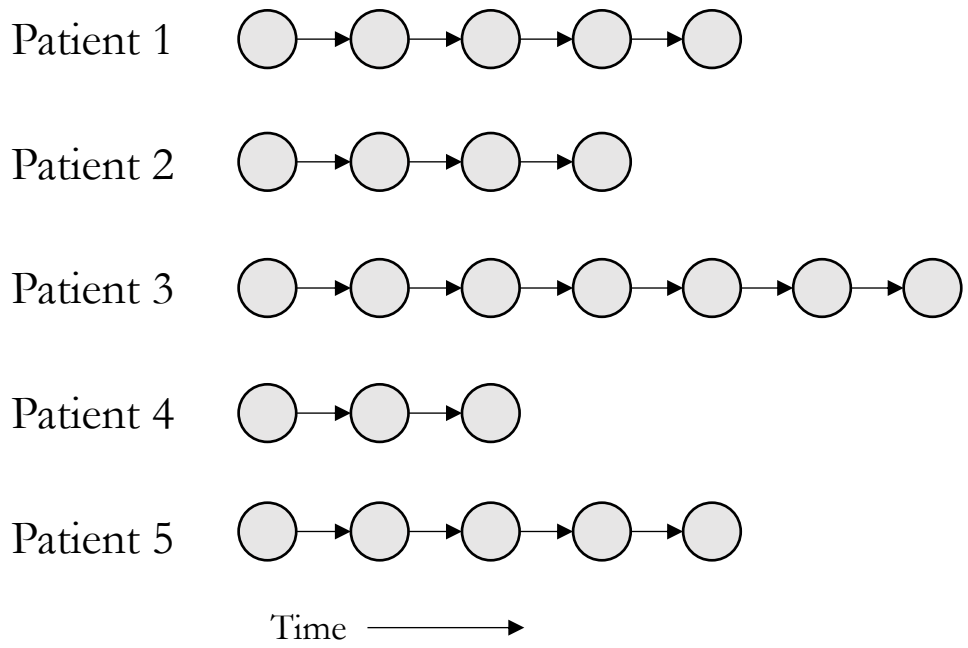
Disease progression modeling

Team:

IBM, Machine Learning

University of Pittsburgh, Neurology

Michael J. Fox Foundation, Advisory



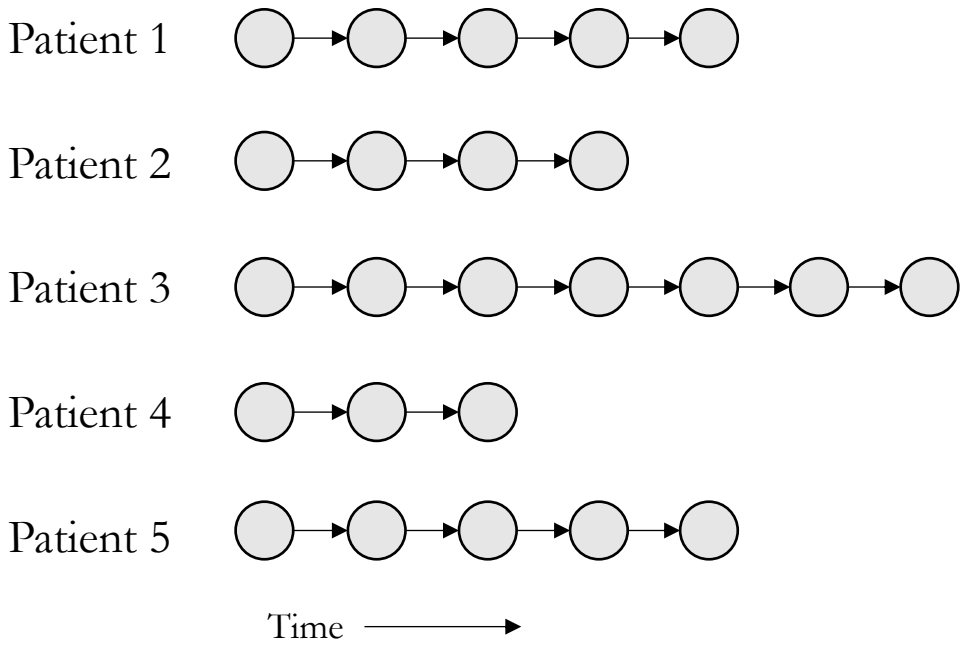
Disease progression modeling

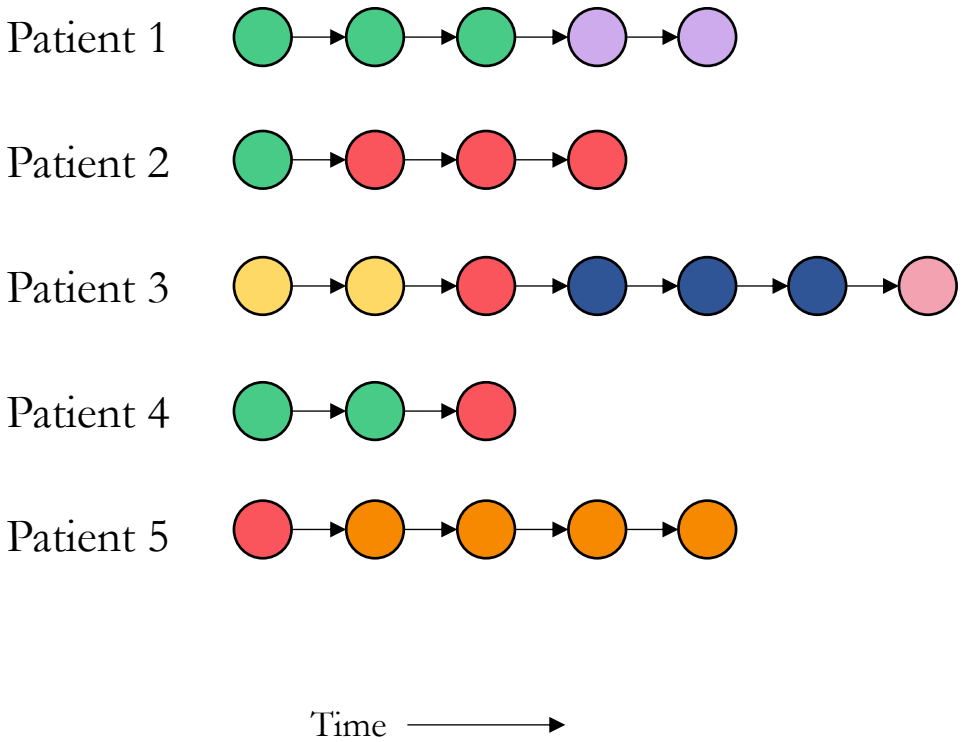
Team:

IBM, Machine Learning

University of Pittsburgh, Neurology

Michael J. Fox Foundation, Advisory





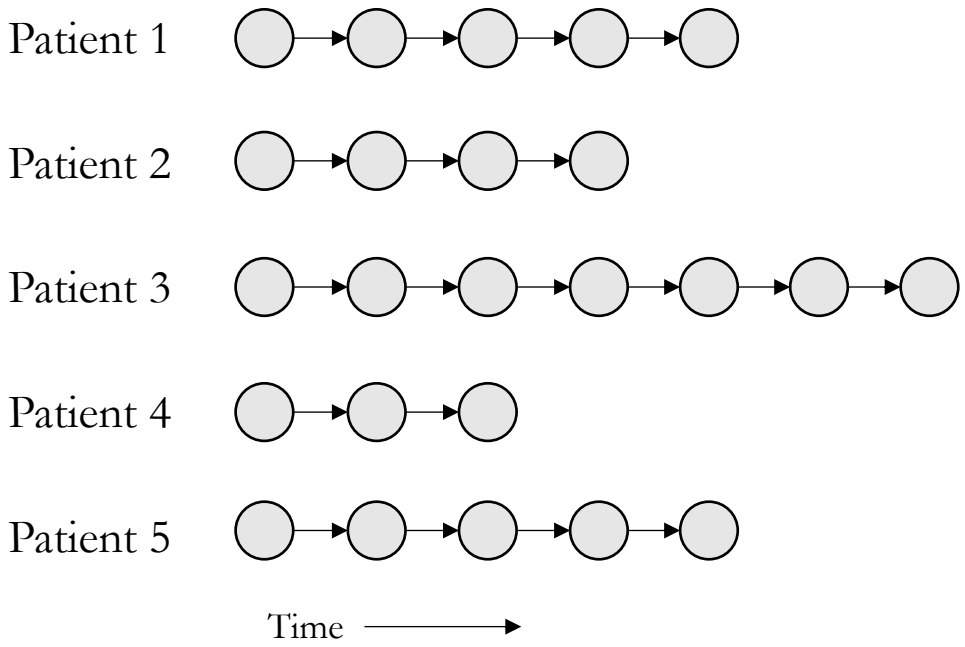
Disease progression modeling

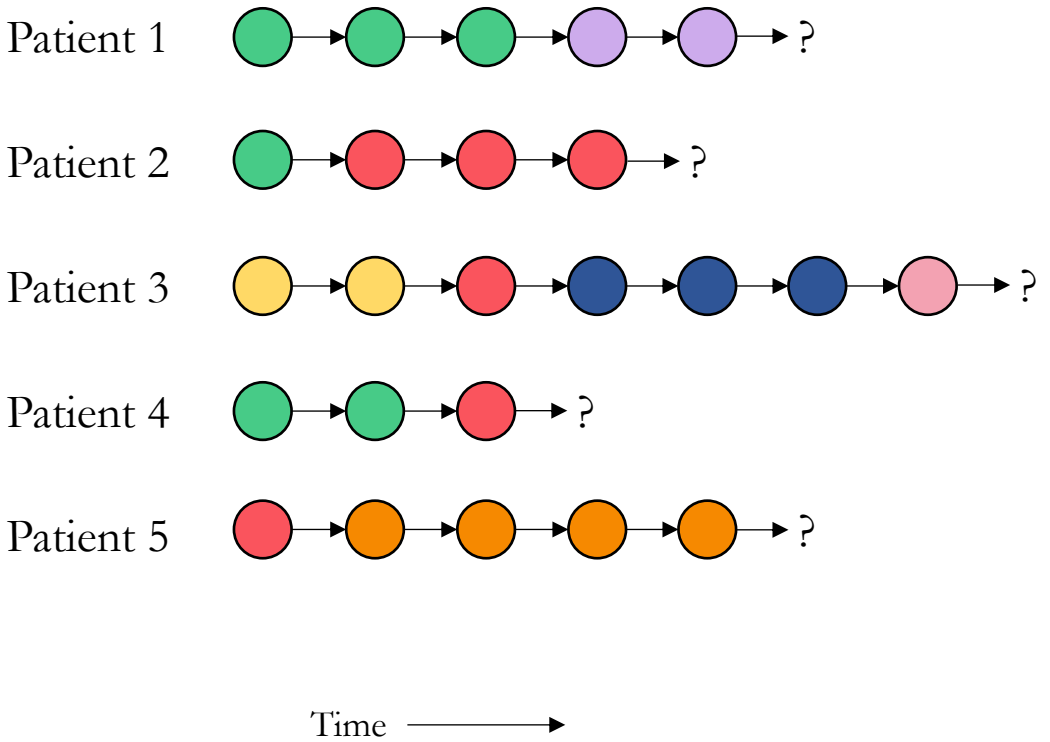
Team:

IBM, Machine Learning

University of Pittsburgh, Neurology

Michael J. Fox Foundation, Advisory





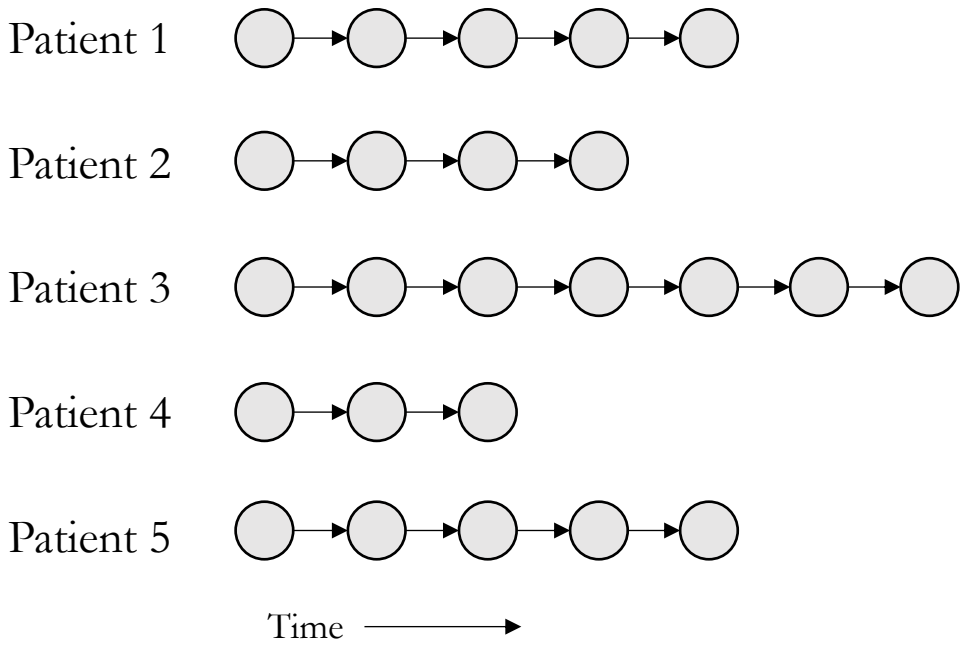
Disease progression modeling

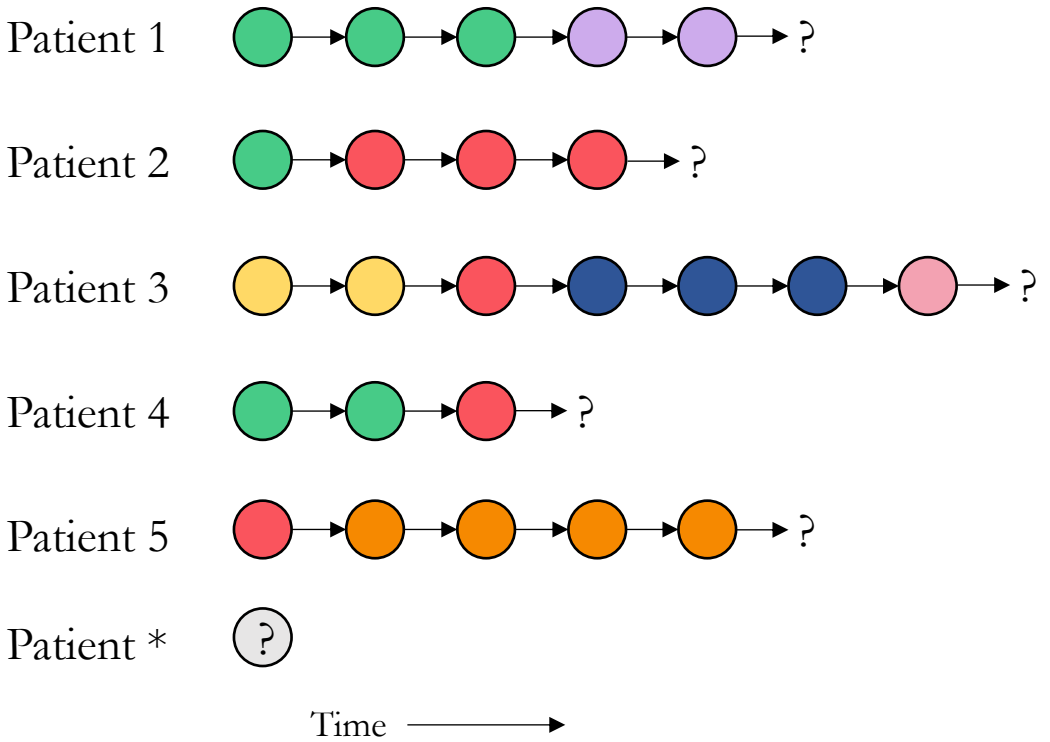
Team:

IBM, Machine Learning

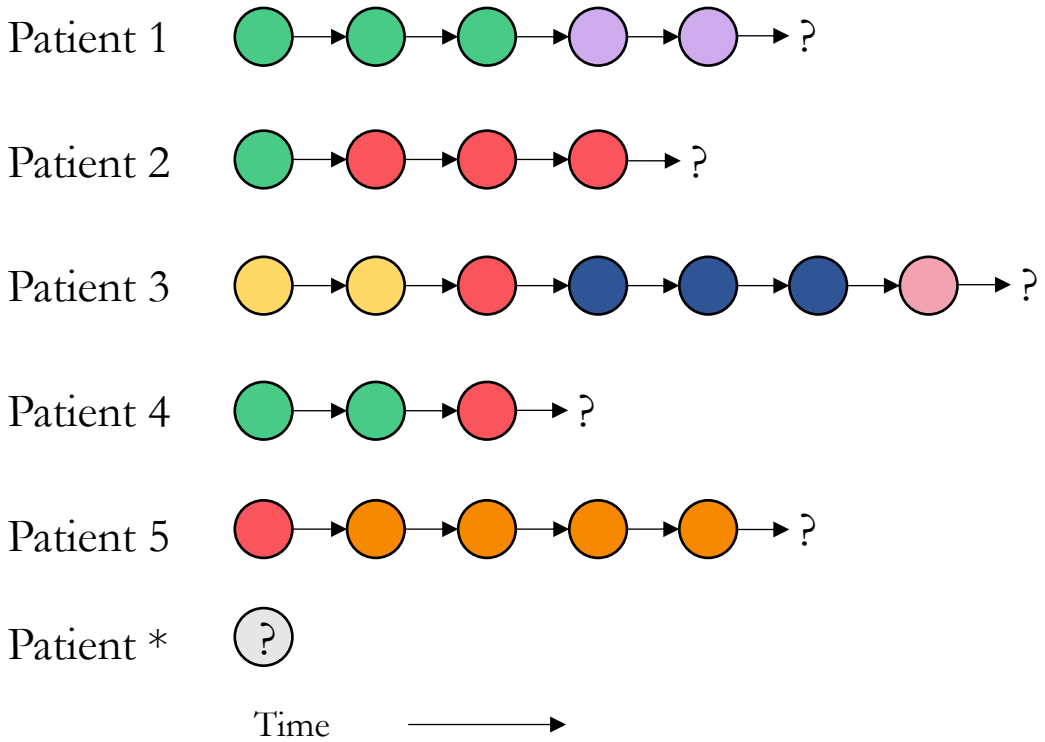
University of Pittsburgh, Neurology

Michael J. Fox Foundation, Advisory





Disease progression modeling



Patient management

Answer questions such as:

- How quickly will disease progress?
- What types of therapies will be most useful for me?
- When should I schedule my next visit?

Clinical trial design

Patient 1

Patient 2

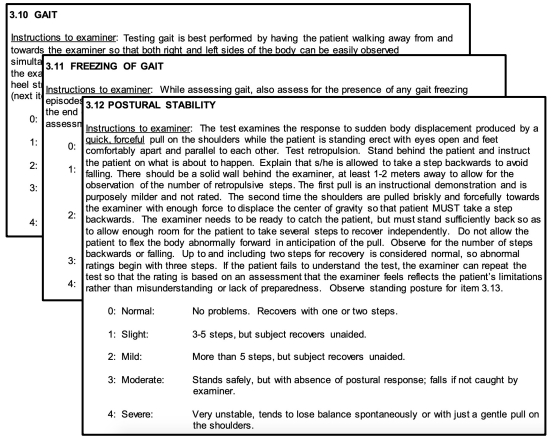
Patient 4

Past work

Expert opinion

Three subtypes:

- Postural instability and gait dominant
- Tremor dominant
- Indeterminant



3.10 GAIT

Instructions to examiner: Testing gait is best performed by having the patient walking away from and towards the examiner so that both right and left sides of the body can be easily observed

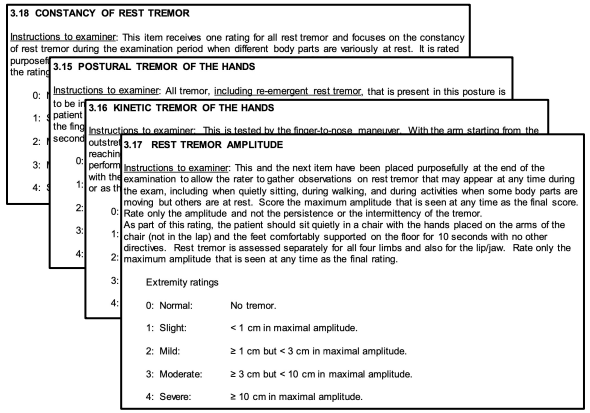
3.11 FREEZING OF GAIT

Instructions to examiner: While assessing gait, also assess for the presence of any gait freezing

3.12 POSTURAL STABILITY

Instructions to examiner: The test examines the response to sudden body displacement produced by a quick, forceful pull on the shoulders while the patient is standing erect with eyes open and feet comfortably apart and parallel to each other. Test retropulsion. Stand behind the patient and instruct the patient on what is about to happen. Explain that s/he is allowed to take a step backwards to avoid falling. There should be a solid wall behind the examiner, at least 1-2 meters away to allow for the observation of the number of retropulsive steps. The first pull is an instructional demonstration and is purposely milder and not rated. The second time the shoulders are pulled briskly and forcefully towards the examiner with enough force to displace the center of gravity so that patient MUST take a step backwards. The examiner needs to be ready to catch the patient, but must stand sufficiently back so as to allow enough room for the patient to take several steps to recover independently. Do not allow the patient to flex the body abnormally forward in anticipation of the pull. Observe for the number of steps backwards or falling. Up to and including two steps for recovery is considered normal, so abnormal ratings begin with three steps. If the patient fails to understand the test, the examiner can repeat the test so that the rating is based on an assessment that the examiner feels reflects the patient's limitations rather than misunderstanding or lack of preparedness. Observe standing posture for item 3.13.

0:	Normal:	No problems. Recovers with one or two steps.
1:	Slight:	3-5 steps, but subject recovers unaided.
2:	Mild:	More than 5 steps, but subject recovers unaided.
3:	Moderate:	Stands safely, but with absence of postural response; falls if not caught by examiner.
4:	Severe:	Very unstable, tends to lose balance spontaneously or with just a gentle pull on the shoulders.



3.18 CONSTANCY OF REST TREMOR

Instructions to examiner: This item receives one rating for all rest tremor and focuses on the constancy of rest tremor during the examination period when different body parts are variously at rest. It is rated

3.15 POSTURAL TREMOR OF THE HANDS

Instructions to examiner: All tremor, including re-emergent rest tremor, that is present in this posture is

3.16 KINETIC TREMOR OF THE HANDS

Instructions to examiner: This is tested by the finger-to-nose maneuver. With the arm starting from the

3.17 REST TREMOR AMPLITUDE

Instructions to examiner: This and the next item have been placed purposefully at the end of the examination to allow the rater to gather observations on rest tremor that may appear at any time during the exam, including when quietly sitting, during walking, and during activities when some body parts are moving but others are at rest. Score the maximum amplitude that is seen at any time as the final score. Rate only the amplitude and not the persistence or the intermittency of the tremor.

As part of this rating, the patient should sit quietly in a chair with the hands placed on the arms of the chair (not in the lap) and the feet comfortably supported on the floor for 10 seconds with no other directives. Rest tremor is assessed separately for all four limbs and also for the lip/jaw. Rate only the maximum amplitude that is seen at any time as the final rating.

Extremity ratings

0:	Normal:	No tremor.
1:	Slight:	< 1 cm in maximal amplitude.
2:	Mild:	≥ 1 cm but < 3 cm in maximal amplitude.
3:	Moderate:	≥ 3 cm but < 10 cm in maximal amplitude.
4:	Severe:	≥ 10 cm in maximal amplitude.

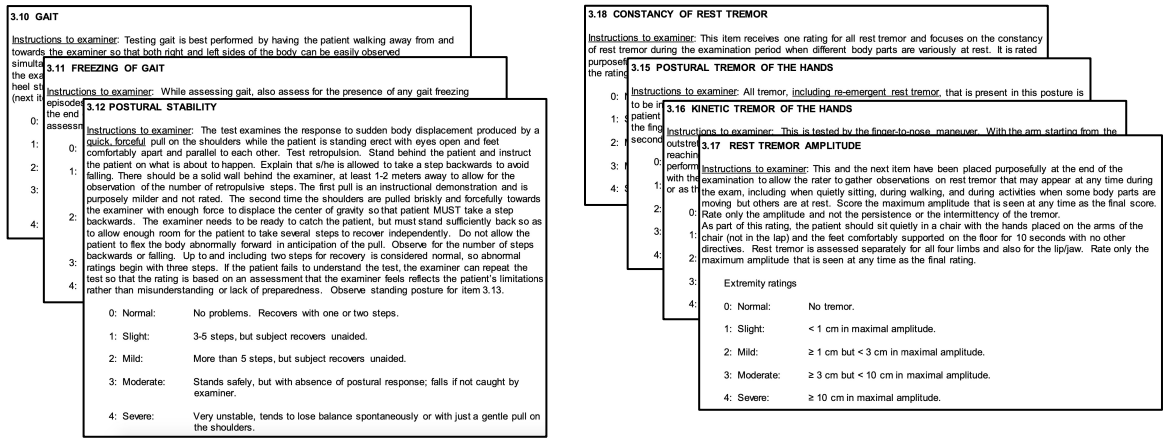
Jankovic. Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2008.

Past work

Expert opinion

Three subtypes:

- Postural instability and gait dominant
- Tremor dominant
- Indeterminant



3.10 GAIT

Instructions to examiner: Testing gait is best performed by having the patient walking away from and towards the examiner so that both right and left sides of the body can be easily observed

3.11 FREEZING OF GAIT

Instructions to examiner: While assessing gait, also assess for the presence of any gait freezing

3.12 POSTURAL STABILITY

Instructions to examiner: The test examines the response to sudden body displacement produced by a quick, forceful pull on the shoulders while the patient is standing erect with eyes open and feet comfortably apart and parallel to each other. Test retropulsion. Stand behind the patient and instruct the patient on what is about to happen. Explain that s/he is allowed to take a step backwards to avoid falling. There should be a solid wall behind the examiner, at least 1-2 meters away to allow for the observation of the number of retropulsive steps. The first pull is an instructional demonstration and is purposely milder and not rated. The second time the shoulders are pulled briskly and forcefully towards the examiner with enough force to displace the center of gravity so that patient MUST take a step backwards. The examiner needs to be ready to catch the patient, but must stand sufficiently back so as to allow enough room for the patient to take several steps to recover independently. Do not allow the patient to flex the body abnormally forward in anticipation of the pull. Observe for the number of steps backwards or falling. Up to and including two steps for recovery is considered normal, so abnormal ratings begin with three steps. If the patient fails to understand the test, the examiner can repeat the test so that the rating is based on an assessment that the examiner feels reflects the patient's limitations rather than misunderstanding or lack of preparedness. Observe standing posture for item 3.13.

0: Normal: No problems. Recovers with one or two steps.

1: Slight: 3-5 steps, but subject recovers unaided.

2: Mild: More than 5 steps, but subject recovers unaided.

3: Moderate: Stands safely, but with absence of postural response; falls if not caught by examiner.

4: Severe: Very unstable, tends to lose balance spontaneously or with just a gentle pull on the shoulders.

3.18 CONSTANCY OF REST TREMOR

Instructions to examiner: This item receives one rating for all rest tremor and focuses on the constancy of rest tremor during the examination period when different body parts are variously at rest. It is rated

3.15 POSTURAL TREMOR OF THE HANDS

Instructions to examiner: All tremor, including re-emergent rest tremor, that is present in this posture is

3.16 KINETIC TREMOR OF THE HANDS

Instructions to examiner: This is tested by the finger-to-nose maneuver. With the arm starting from the

3.17 REST TREMOR AMPLITUDE

Instructions to examiner: This and the next item have been placed purposefully at the end of the examination to allow the rater to gather observations on rest tremor that may appear at any time during the exam, including when quietly sitting, during walking, and during activities when some body parts are moving but others are at rest. Score the maximum amplitude that is seen at any time as the final score. Rate only the amplitude and not the persistence or the intermittency of the tremor.

As part of this rating, the patient should sit quietly in a chair with the hands placed on the arms of the chair (not in the lap) and the feet comfortably supported on the floor for 10 seconds with no other directives. Rest tremor is assessed separately for all four limbs and also for the lip/jaw. Rate only the maximum amplitude that is seen at any time as the final rating.

Extremity ratings

0: Normal: No tremor.

1: Slight: < 1 cm in maximal amplitude.

2: Mild: ≥ 1 cm but < 3 cm in maximal amplitude.

3: Moderate: ≥ 3 cm but < 10 cm in maximal amplitude.

4: Severe: ≥ 10 cm in maximal amplitude.

Machine-learning based approaches

Three subtypes:

- Mild motor-predominant
- Diffuse malignant
- Intermediate

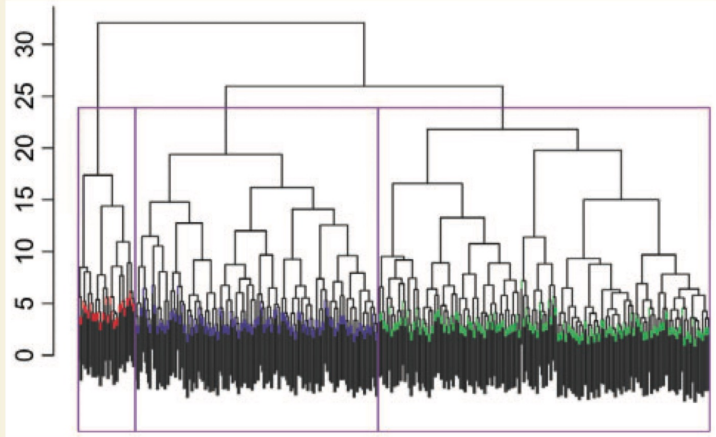


Figure 1 Dendrogram of the final hierarchical cluster solution in the PPMI population. Green = mild motor-predominant; blue = intermediate; red = diffuse malignant.

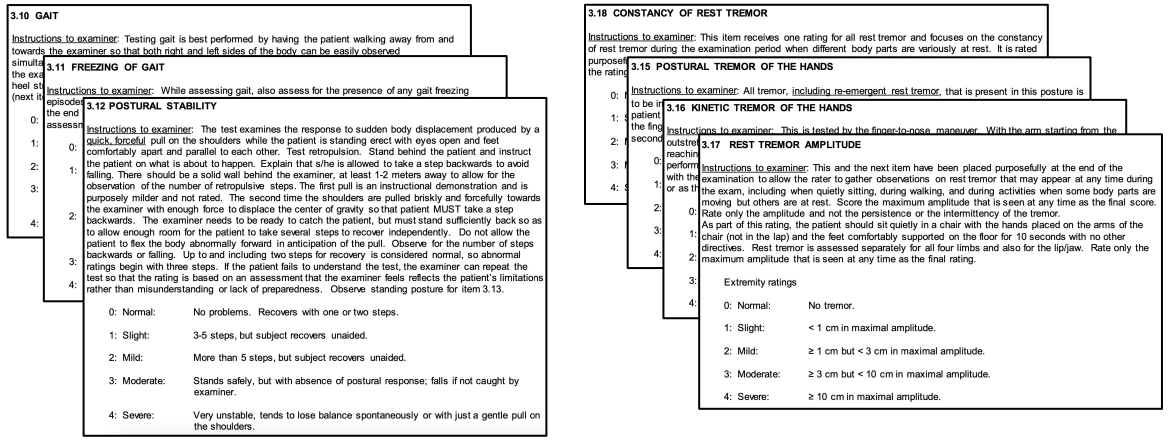
Fereshtehnejad, Zeighami, Dagher, and Postuma. Clinical criteria for subtyping Parkinson's disease: Biomarkers and longitudinal progression. *Brain*. 2017.

Past work

Expert opinion

Three subtypes:

- Postural instability and gait dominant
- Tremor dominant
- Indeterminant



3.10 GAIT

Instructions to examiner: Testing gait is best performed by having the patient walking away from and towards the examiner so that both right and left sides of the body can be easily observed

3.11 FREEZING OF GAIT

Instructions to examiner: While assessing gait, also assess for the presence of any gait freezing

3.12 POSTURAL STABILITY

Instructions to examiner: The test examines the response to sudden body displacement produced by a quick, forceful pull on the shoulders while the patient is standing erect with eyes open and feet comfortably apart and parallel to each other. Test retropulsion. Stand behind the patient and instruct the patient on what is about to happen. Explain that s/he is allowed to take a step backwards to avoid falling. There should be a solid wall behind the examiner, at least 1-2 meters away to allow for the observation of the number of retropulsive steps. The first pull is an instructional demonstration and is purposely milder and not rated. The second time the shoulders are pulled briskly and forcefully towards the examiner with enough force to displace the center of gravity so that patient MUST take a step backwards. The examiner needs to be ready to catch the patient, but must stand sufficiently back so as to allow enough room for the patient to take several steps to recover independently. Do not allow the patient to flex the body abnormally forward in anticipation of the pull. Observe for the number of steps backwards or falling. Up to and including two steps for recovery is considered normal, so abnormal ratings begin with three steps. If the patient fails to understand the test, the examiner can repeat the test so that the rating is based on an assessment that the examiner feels reflects the patient's limitations rather than misunderstanding or lack of preparedness. Observe standing posture for item 3.13.

0:	Normal:	No problems. Recovers with one or two steps.
1:	Slight:	3-5 steps, but subject recovers unaided.
2:	Mild:	More than 5 steps, but subject recovers unaided.
3:	Moderate:	Stands safely, but with absence of postural response; falls if not caught by examiner.
4:	Severe:	Very unstable, tends to lose balance spontaneously or with just a gentle pull on the shoulders.

3.18 CONSTANCY OF REST TREMOR

Instructions to examiner: This item receives one rating for all rest tremor and focuses on the constancy of rest tremor during the examination period when different body parts are variously at rest. It is rated

3.15 POSTURAL TREMOR OF THE HANDS

Instructions to examiner: All tremor, including re-emergent rest tremor, that is present in this posture is

3.16 KINETIC TREMOR OF THE HANDS

Instructions to examiner: This is tested by the finger-to-nose maneuver. With the arm starting from the

3.17 REST TREMOR AMPLITUDE

Instructions to examiner: This and the next item have been placed purposefully at the end of the examination to allow the rater to gather observations on rest tremor that may appear at any time during the exam, including when quietly sitting, during walking, and during activities when some body parts are moving but others are at rest. Score the maximum amplitude that is seen at any time as the final score. Rate only the amplitude and not the persistence or the intermittency of the tremor.

As part of this rating, the patient should sit quietly in a chair with the hands placed on the arms of the chair (not in the lap) and the feet comfortably supported on the floor for 10 seconds with no other directives. Rest tremor is assessed separately for all four limbs and also for the lip/jaw. Rate only the maximum amplitude that is seen at any time as the final rating.

Extremity ratings

0:	Normal:	No tremor.
1:	Slight:	< 1 cm in maximal amplitude.
2:	Mild:	≥ 1 cm but < 3 cm in maximal amplitude.
3:	Moderate:	≥ 3 cm but < 10 cm in maximal amplitude.
4:	Severe:	≥ 10 cm in maximal amplitude.

Machine-learning based approaches

Four subtypes:

- Fast motor progression (1)
- Mild motor and non-motor disease (2)
- Severe motor disease, poor psychological well-being and sleep (3)
- Slow motor progression (4)

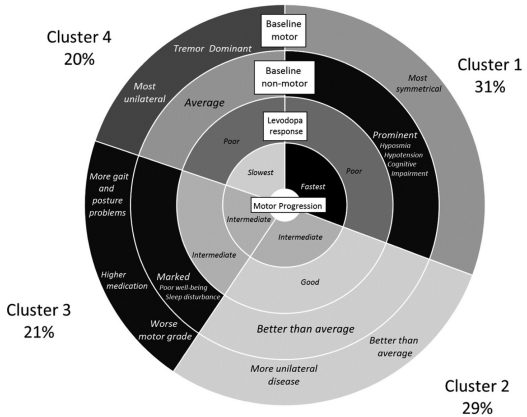


Figure 1 Important salient clinical features of the four clusters across the two cohorts where the percentages within each cluster are from the Tracking Parkinson's cohort.

Lawton et al. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *Journal of Neurology, Neurosurgery, and Psychiatry.* 2018.

Analysis Goals

Analysis Goals

- Particularly interested in discovering *disease states*

Analysis Goals

- Particularly interested in discovering *disease states*

Disease state definition: *discrete label* of a patient with two primary characteristics: a transition model which describes the probability of changing states and an observation model which describes the distribution of clinical measures associated with a state

Analysis Goals

- Particularly interested in discovering *disease states*
- Account for *medications* which affect clinical presentation but do not affect disease course

Analysis Goals

- Particularly interested in discovering *disease states*
- Account for *medications* which affect clinical presentation but do not affect disease course
- Allow for *personalized* deviations in symptom presentation and medication effects

Key Model Assumptions

Key Model Assumptions

- Pre-process the data using a latent variable model

Severson, Ghosh and Ng. Unsupervised learning with contrastive latent variable models. *AAAI.* 2019.

Key Model Assumptions

- Pre-process the data using a latent variable model
- Model progression using a Markov model, a specific type of probabilistic model

Severson et al. Personalized input-output hidden Markov models for disease progression modeling. *Machine Learning for Healthcare*. 2020.

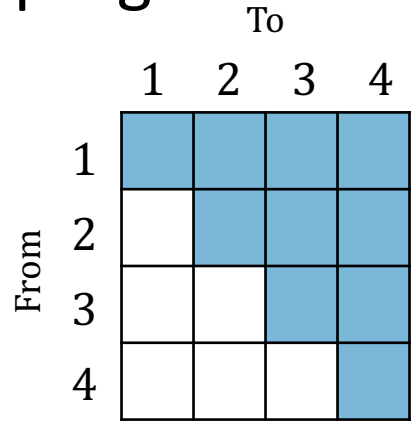
Key Model Assumptions

- Pre-process the data using a latent variable model
- Model progression using a Markov model, a specific type of probabilistic model
- Medications effect symptoms but not underlying disease progression

Severson et al. Personalized input-output hidden Markov models for disease progression modeling. *Machine Learning for Healthcare*. 2020.

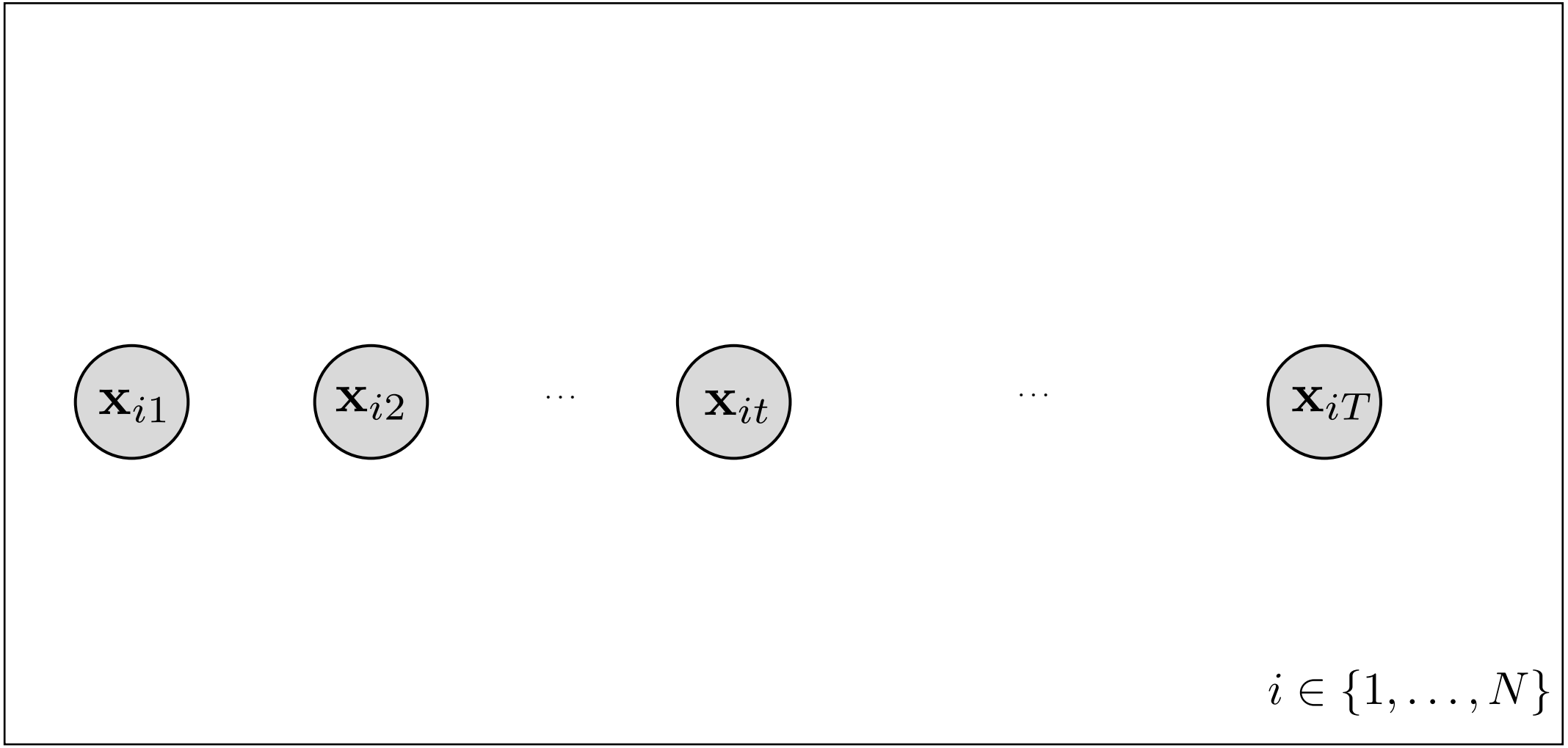
Key Model Assumptions

- Pre-process the data using a latent variable model
- Model progression using a Markov model, a specific type of probabilistic model
- Medications effect symptoms but not underlying disease progression
- Patients can only move progressively through states

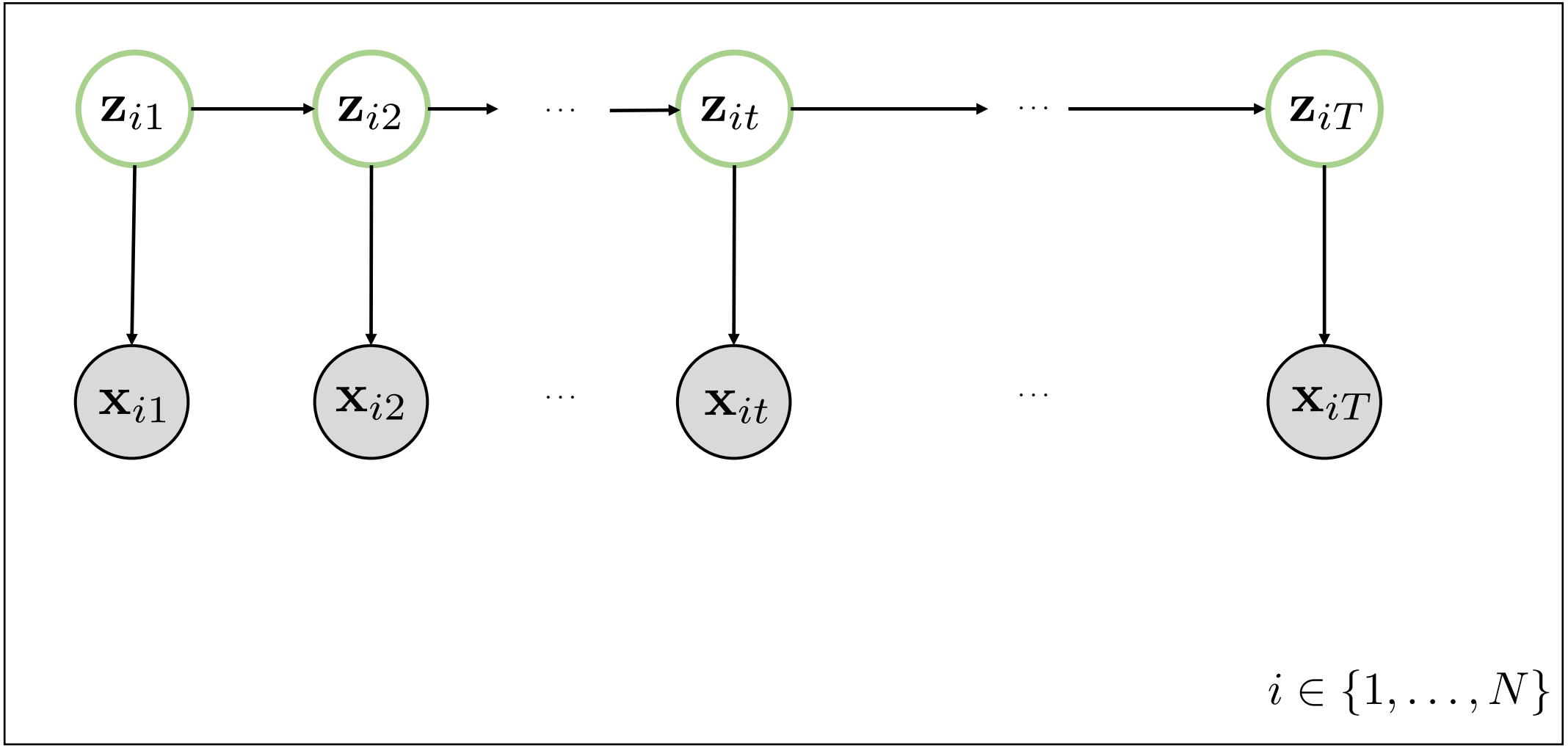


Severson et al. Personalized input-output hidden Markov models for disease progression modeling. *Machine Learning for Healthcare*. 2020.

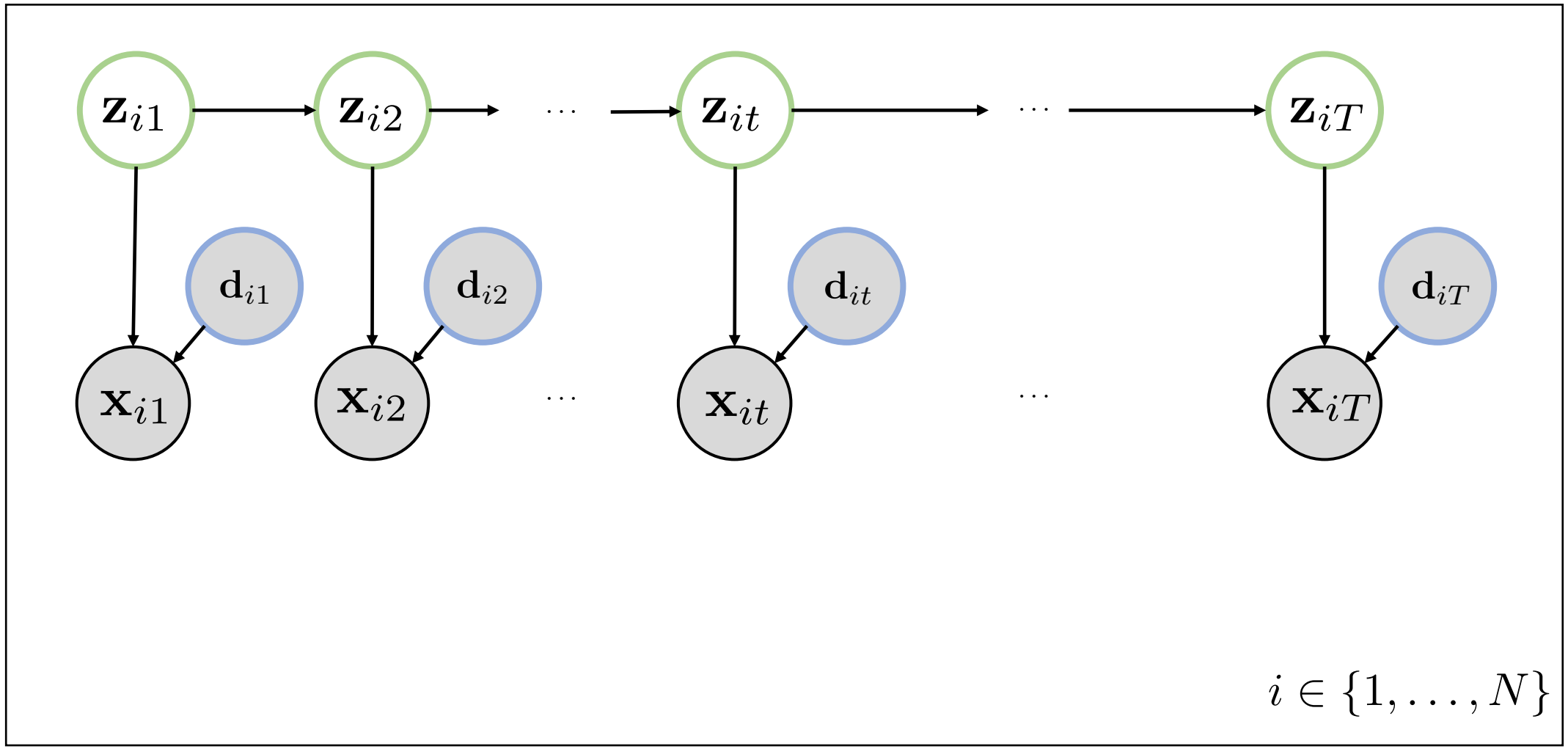
Personalized input output hidden Markov models



Personalized input output hidden Markov models

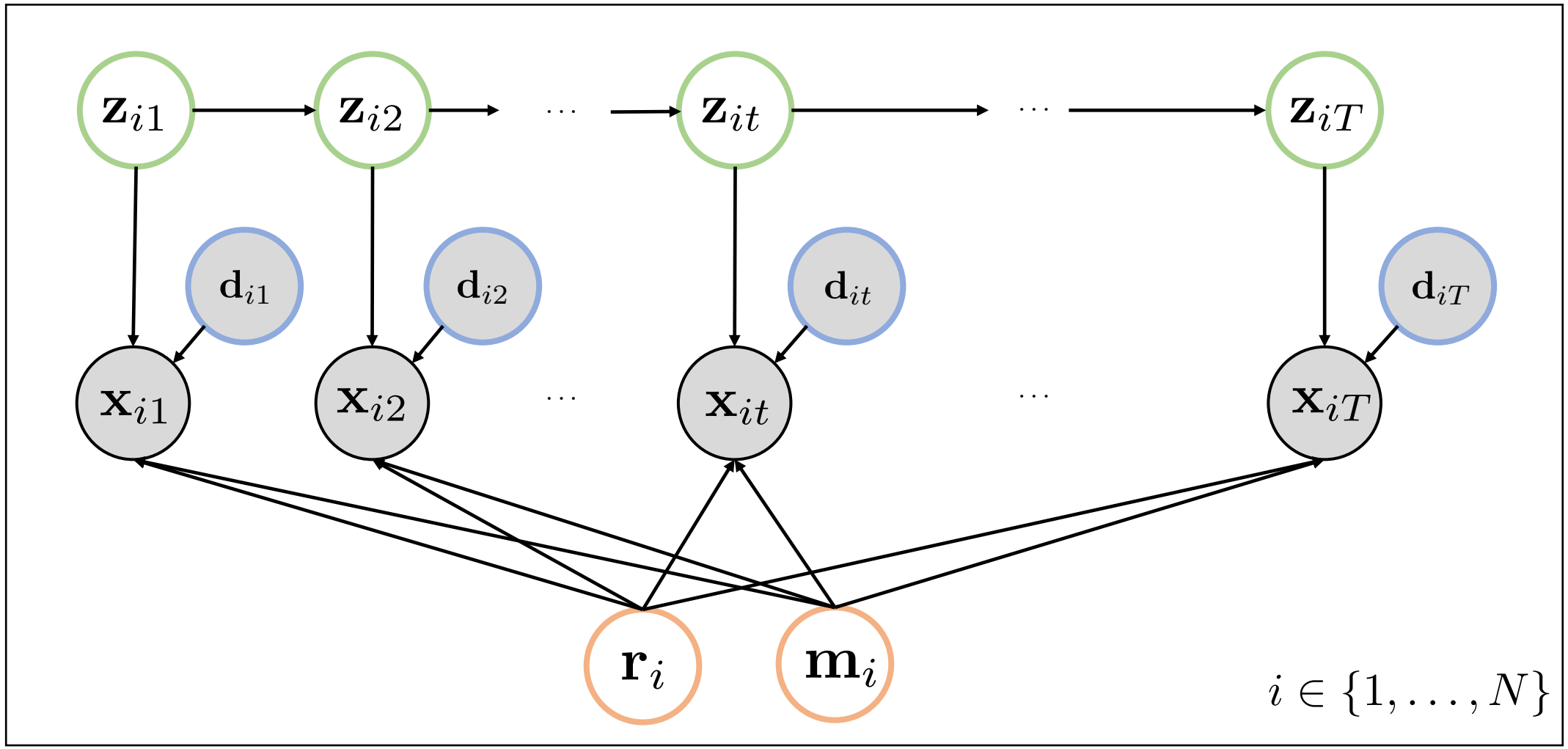


Personalized input output hidden Markov models



Bengio and Frasconi. An input output HMM architecture. *NIPS* 1995.

Personalized input output hidden Markov models



Severson et al. Personalized input-output hidden Markov models for disease progression modeling. *Machine Learning for Healthcare*. 2020.

Personalized input output hidden Markov models

- The graphical model on the previous slide is specified using the following distributions:

$$z_{i,1} \sim \mathrm{Cat}(\pi), \quad z_{i,t}|z_{i,t-1} = j \sim \mathrm{Cat}(A_j)$$

$$x_{i,t}|z_{i,t} = k, d_{i,t} \sim \mathcal{N}(\mu_k + r_i + (v_k + m_i)d_{i,t}, \Sigma_k)$$

$$m_i \sim \mathcal{N}(0, \sigma_m^2 \mathbf{I}_D), \quad r_i \sim \mathcal{N}(0, \sigma_r^2 \mathbf{I}_D), \quad v_k \sim \mathcal{N}(0, \sigma_v^2, \mathbf{I}_D)$$

Severson et al. Personalized input-output hidden Markov models for disease progression modeling. *Machine Learning for Healthcare*. 2020.

Learning the model

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

Optimization objective:

$$\mathcal{L}(\theta, \lambda) = \mathbb{E}_{q(\mathbf{Z},\mathbf{M},\mathbf{R}|\mathbf{X},\lambda)}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{M}, \mathbf{R}|\theta)] + \mathbb{H}[q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)]$$

Two remaining pieces to the learning algorithm for PIOHMM:

1. How to choose $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$?
2. Given the posterior approximation, can we evaluate the required expectations?

Choosing the variational approximation

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

- Need to specify $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$

Choosing the variational approximation

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

- Need to specify $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$
- Mean field approximation

$$q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda) = \prod_{i=1}^{N} \left[q(m_i|\lambda_{m_i}) q(r_i|\lambda_{r_i}) \prod_{t=1}^{T} q(z_{it}|\lambda_{z_{it}}) \right]$$

Choosing the variational approximation

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

- Need to specify $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$
- Mean field approximation

$$q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda) = \prod_{i=1}^{N} \left[q(m_i|\lambda_{m_i}) q(r_i|\lambda_{r_i}) \prod_{t=1}^{T} q(z_{it}|\lambda_{z_{it}}) \right]$$

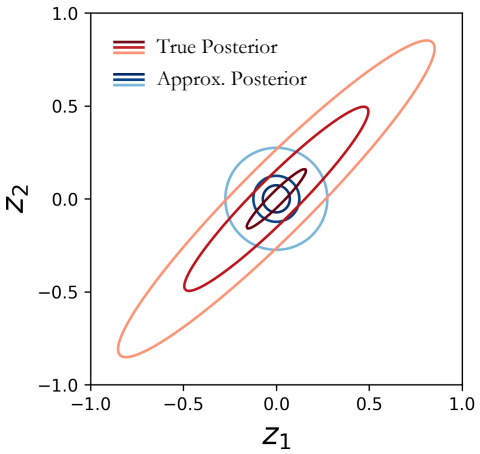


Image based on: Bishop. Pattern Recognition and Machine Learning. 2006.

Choosing the variational approximation

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

- Need to specify $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$
- Mean field approximation

$$q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda) = \prod_{i=1}^{N} \left[q(m_i|\lambda_{m_i}) q(r_i|\lambda_{r_i}) \prod_{t=1}^{T} q(z_{it}|\lambda_{z_{it}}) \right]$$

- Leverage structure of hidden states

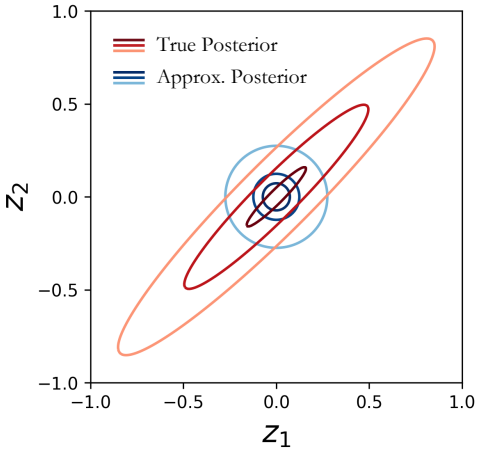


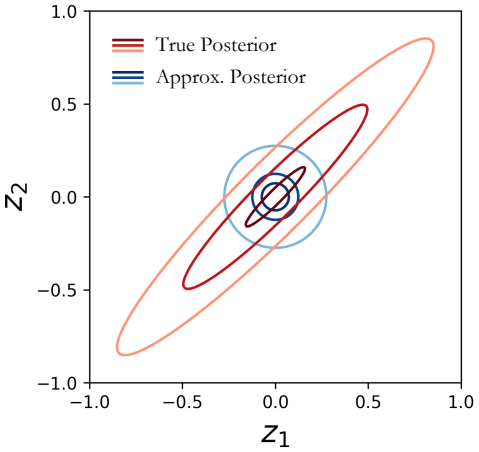
Image based on: Bishop. Pattern Recognition and Machine Learning. 2006.

Choosing the variational approximation

X: Observed data
Z: Disease states
R: Personalized state
M: Personalized medication effects

- Need to specify $q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)$
- Mean field approximation

$$q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda) = \prod_{i=1}^{N} \left[q(m_i|\lambda_{m_i}) q(r_i|\lambda_{r_i}) \prod_{t=1}^{T} q(z_{it}|\lambda_{z_{it}}) \right]$$



- Leverage structure of hidden states

$$q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda) = \prod_{i=1}^{N} \left[q(m_i|\lambda_{m_i}) q(r_i|\lambda_{r_i}) q(z_{i1}|r_i, m_i, x_i) \prod_{t=2}^{T} q(z_{it}|z_{i(t-1)}, r_i, m_i, x_i) \right]$$

Image based on: Bishop. Pattern Recognition and Machine Learning. 2006.

Evaluating the expectations

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

Optimization objective:

$$\mathcal{L}(\theta, \lambda) = \mathbb{E}_{q(\mathbf{Z},\mathbf{M},\mathbf{R}|\mathbf{X},\lambda)}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{M}, \mathbf{R}|\theta)] + \mathbb{H}[q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)]$$

Evaluating the expectations

$\mathbf{X}$: Observed data
$\mathbf{Z}$: Disease states
$\mathbf{R}$: Personalized state
$\mathbf{M}$: Personalized medication effects

Optimization objective:

$$\mathcal{L}(\theta, \lambda) = \mathbb{E}_{q(\mathbf{Z},\mathbf{M},\mathbf{R}|\mathbf{X},\lambda)}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{M}, \mathbf{R}|\theta)] + \mathbb{H}[q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)]$$

$$\mathbb{E}_{q(z_i,m_i,r_i|x_i,\lambda)}[\log p(x_i|z_i, m_i, r_i, \theta)]$$

Evaluating the expectations

X: Observed data
Z: Disease states
R: Personalized state
M: Personalized medication effects

Optimization objective:

$$\mathcal{L}(\theta, \lambda) = \mathbb{E}_{q(\mathbf{Z},\mathbf{M},\mathbf{R}|\mathbf{X},\lambda)}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{M}, \mathbf{R}|\theta)] + \mathbb{H}[q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)]$$

Monte Carlo approximation:

$$\mathbb{E}_{q(z_i,m_i,r_i|x_i,\lambda)}[\log p(x_i|z_i, m_i, r_i, \theta)] \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{E}_{q(z_i|x_i,m_i=m_i^s,r_i=r_i^s)}[\log p(x_i|z_i, m_i = m_i^s, r_i = r_i^s, \theta)]$$

Evaluating the expectations

X: Observed data
Z: Disease states
R: Personalized state
M: Personalized medication effects

Optimization objective:

$$\mathcal{L}(\theta, \lambda) = \mathbb{E}_{q(\mathbf{Z},\mathbf{M},\mathbf{R}|\mathbf{X},\lambda)}[\log p(\mathbf{X}, \mathbf{Z}, \mathbf{M}, \mathbf{R}|\theta)] + \mathbb{H}[q(\mathbf{Z}, \mathbf{M}, \mathbf{R}|\mathbf{X}, \lambda)]$$

Monte Carlo approximation:

$$\mathbb{E}_{q(z_i,m_i,r_i|x_i,\lambda)}[\log p(x_i|z_i, m_i, r_i, \theta)] \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{E}_{q(z_i|x_i,m_i=m_i^s,r_i=r_i^s)}[\log p(x_i|z_i, m_i = m_i^s, r_i = r_i^s, \theta)]$$

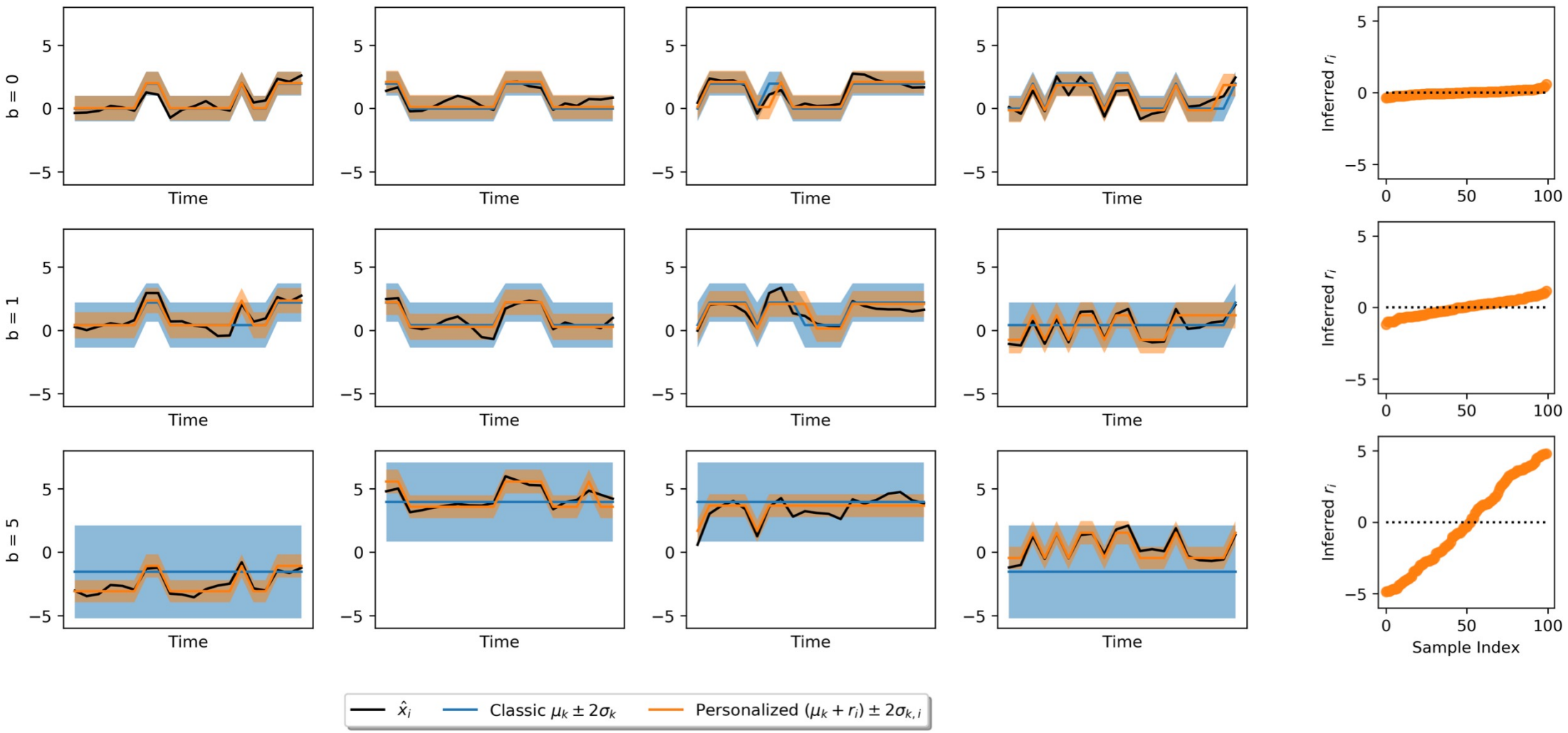
"Reparameterization trick":

$$\epsilon_{m_i}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad m_i^s = \hat{L}_{m_i}\epsilon_{m_i}^s + \hat{\mu}_{m_i}$$

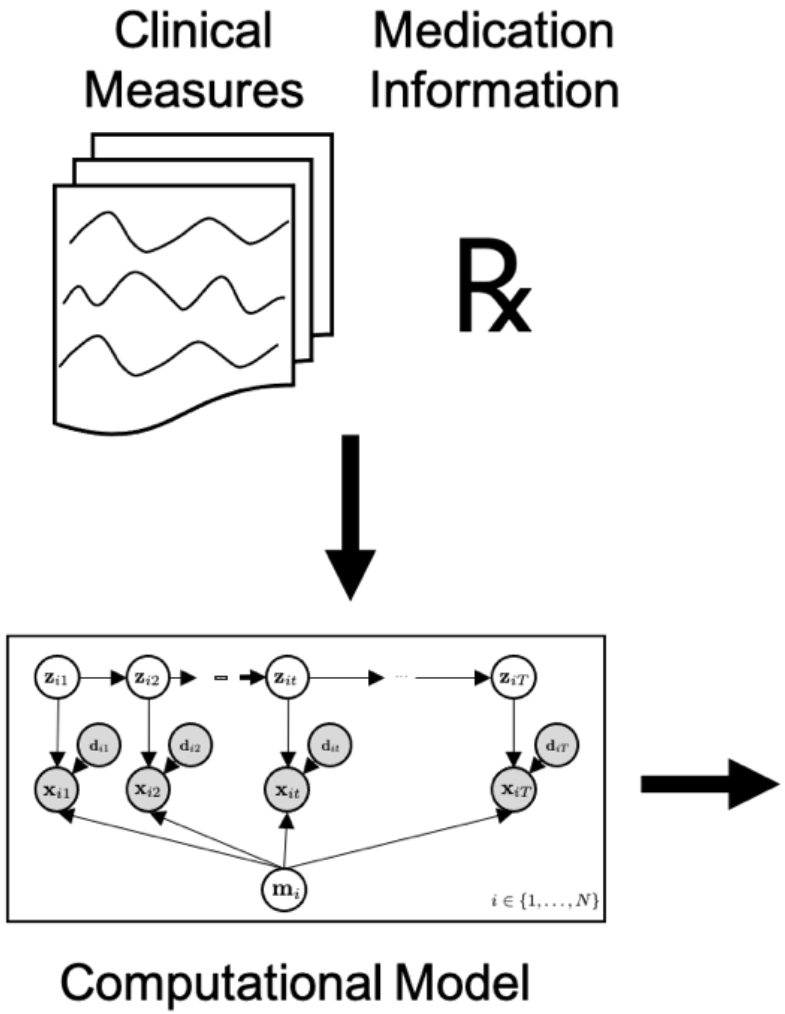
$$\epsilon_{r_i}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad r_i^s = \hat{L}_{r_i}\epsilon_{r_i}^s + \hat{\mu}_{r_i}$$

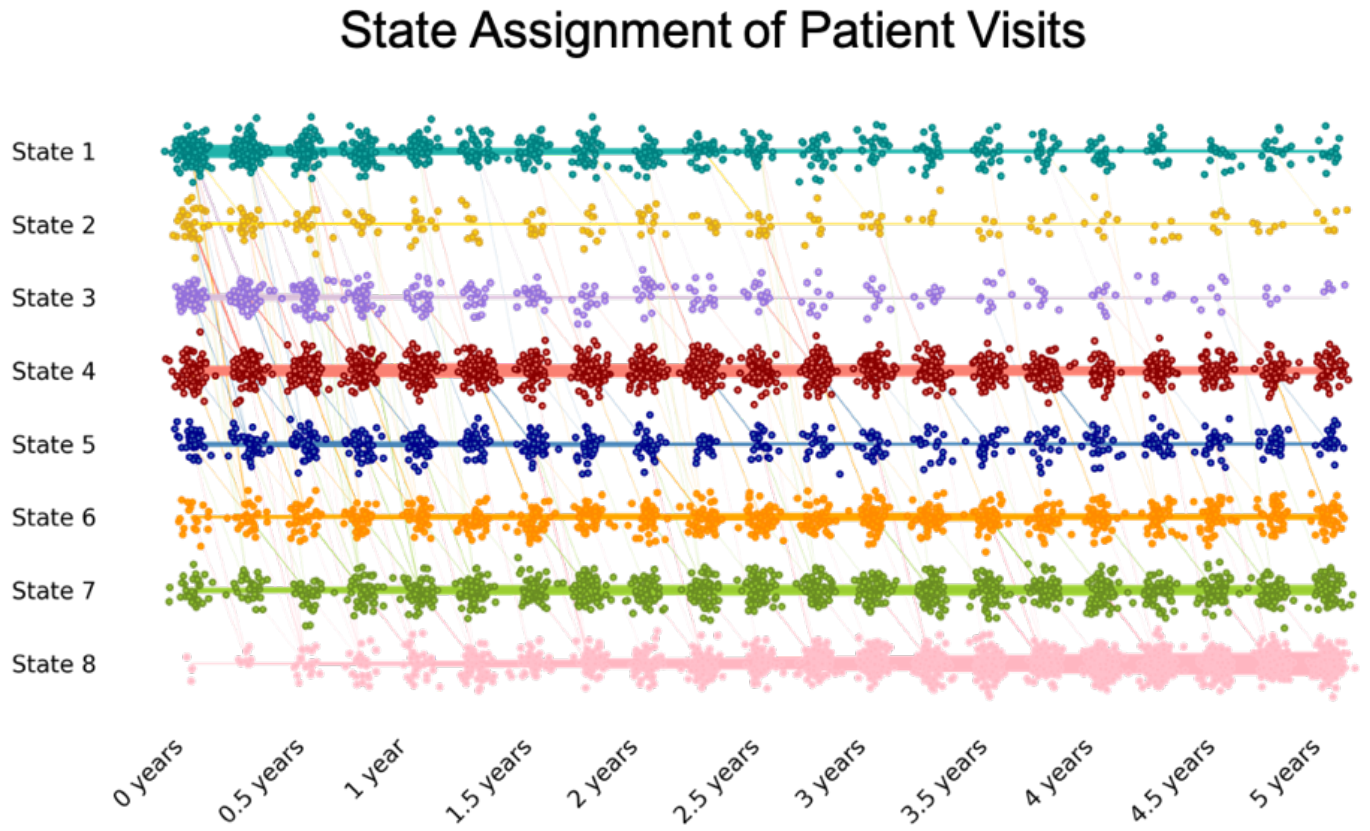
Kingma and Welling. Auto-encoding variational Bayes. ICLR 2014.

Synthetic proof of concept



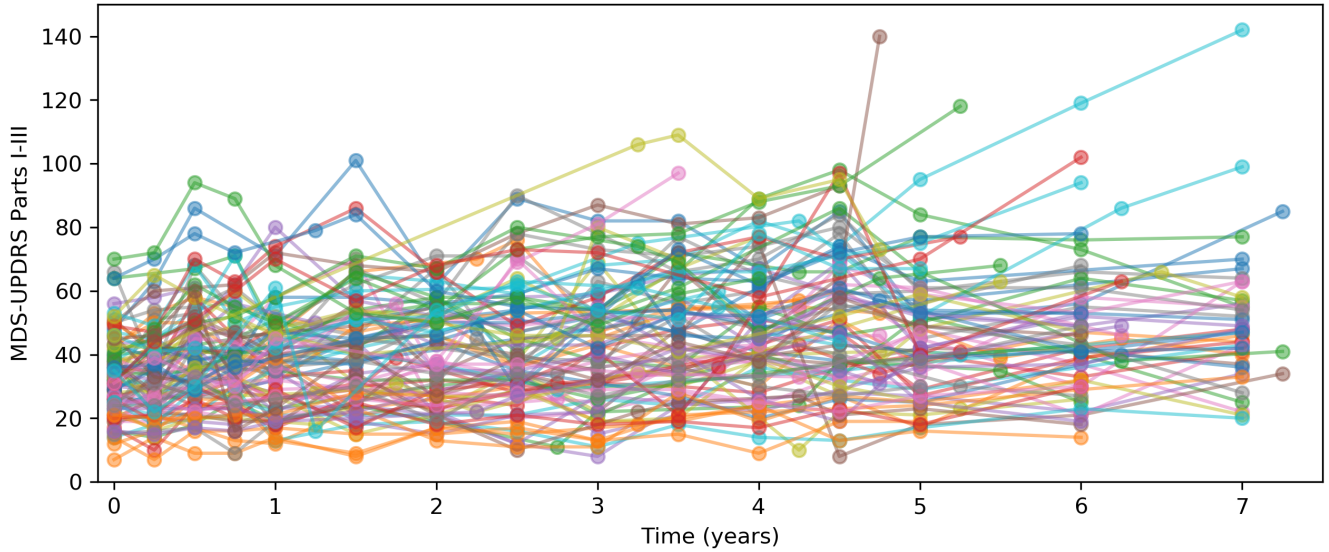
Summarizing the project





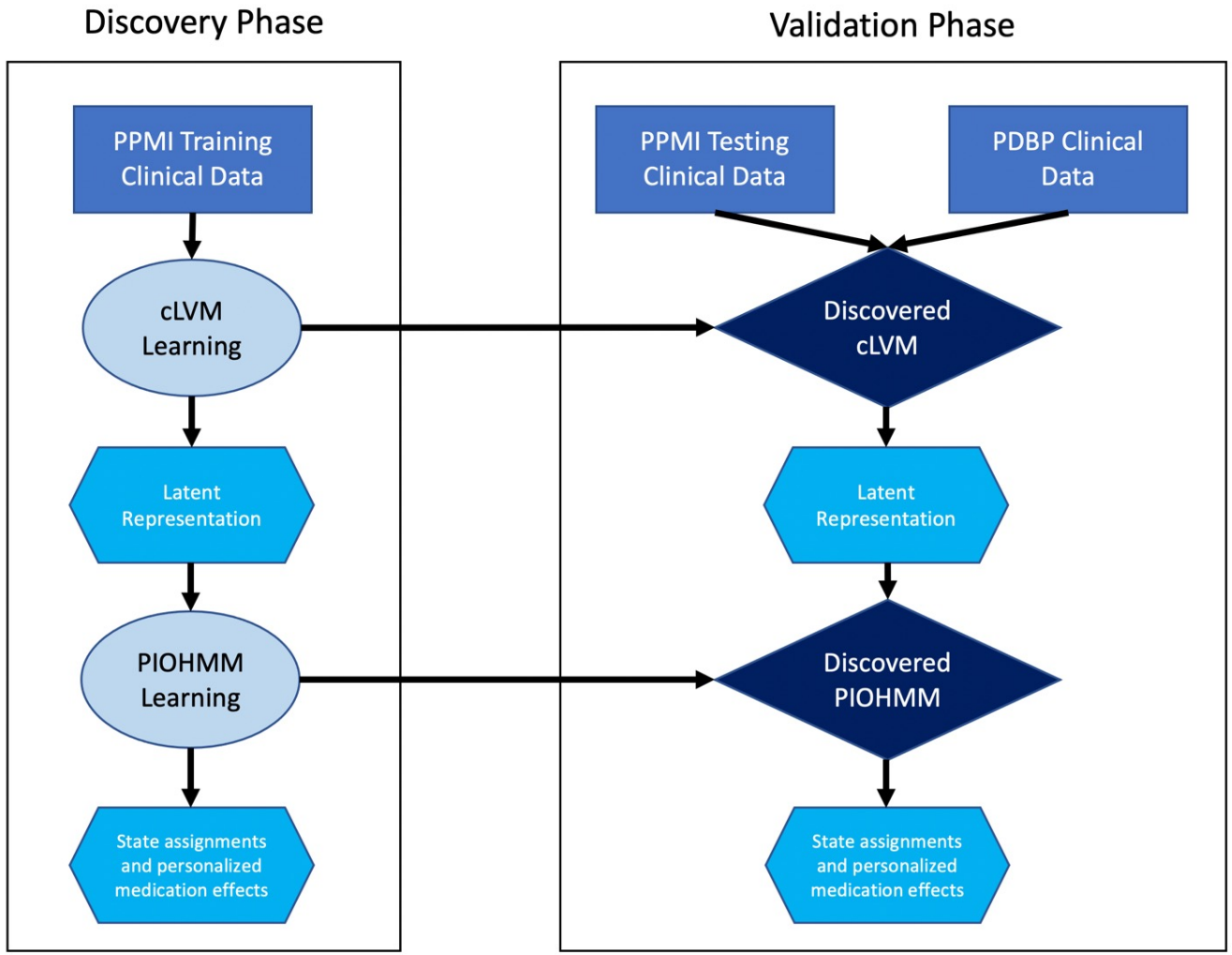
Dataset for analysis

- Longitudinal data from the Parkinson's Progression Markers Initiative (PPMI) and Parkinson's Disease Biomarker Program (PDBP)
- 82 variables measuring motor and non-motor symptoms over time
- 423 PD patients from PPMI and 610 PD patients from PDBP



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Analysis workflow



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Key take-aways

- The learning procedure resulted in an 8 state model where states are primarily differentiated by functional impairment, tremor, bradykinesia, and neuropsychiatric measures

Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Key take-aways

- The learning procedure resulted in an 8 state model where states are primarily differentiated by functional impairment, tremor, bradykinesia, and neuropsychiatric measures
- The predictive model discovers non-sequential overlapping disease progression trajectories

Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

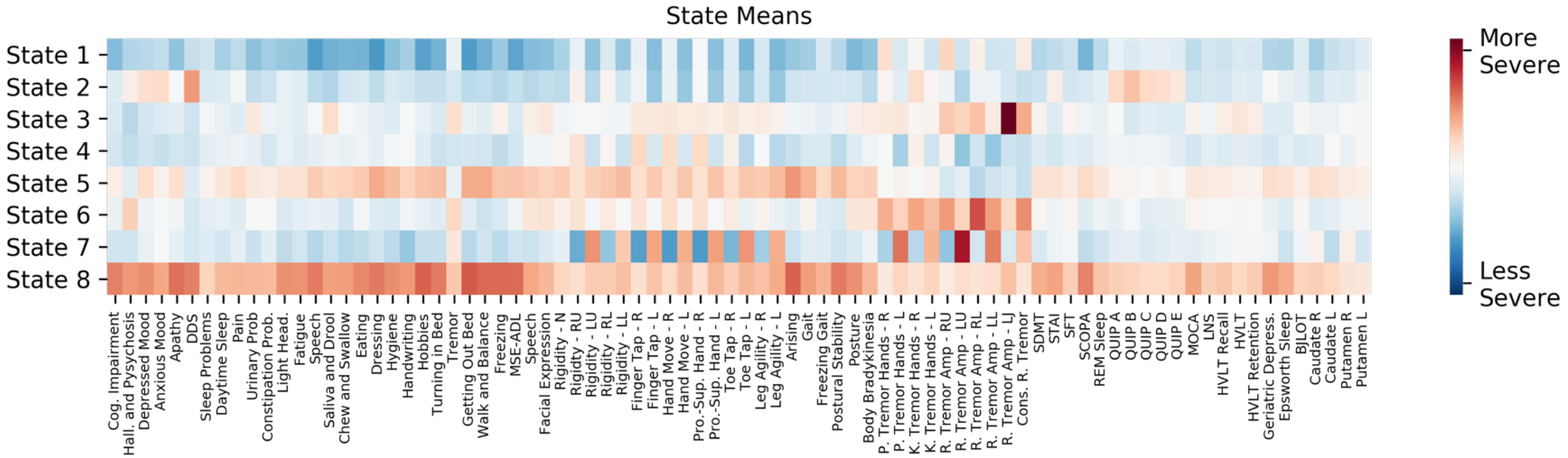
Key take-aways

- The learning procedure resulted in an 8 state model where states are primarily differentiated by functional impairment, tremor, bradykinesia, and neuropsychiatric measures
- The predictive model discovers non-sequential overlapping disease progression trajectories
- The addition of personalized medication effects leads to a more descriptive model

Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Model parameters – State means

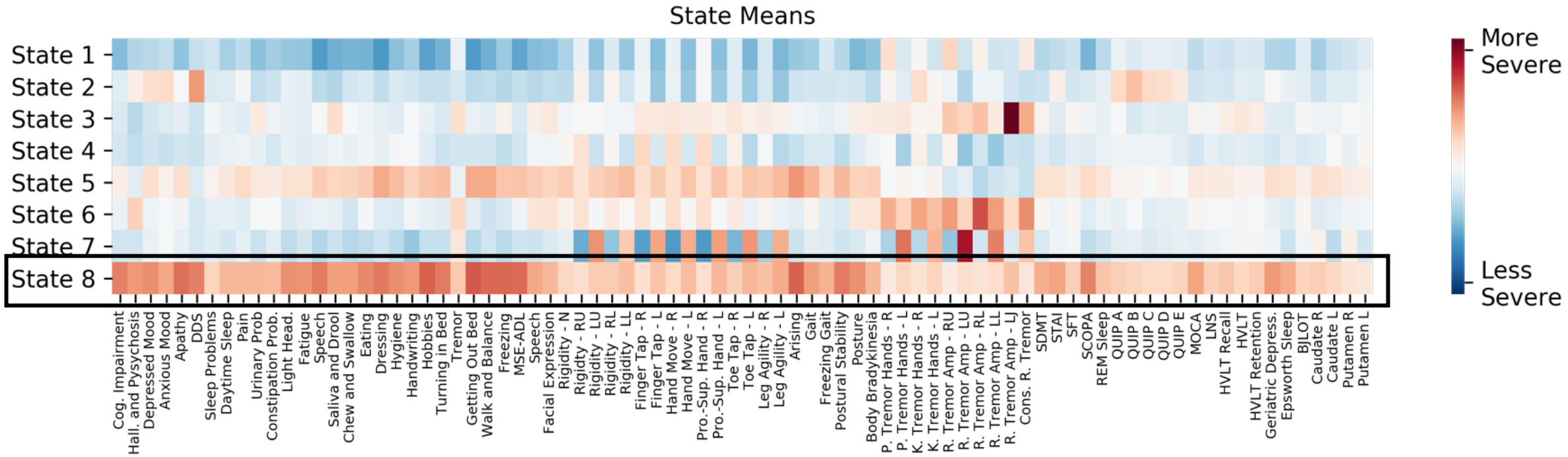
How can we interpret the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Model parameters – State means

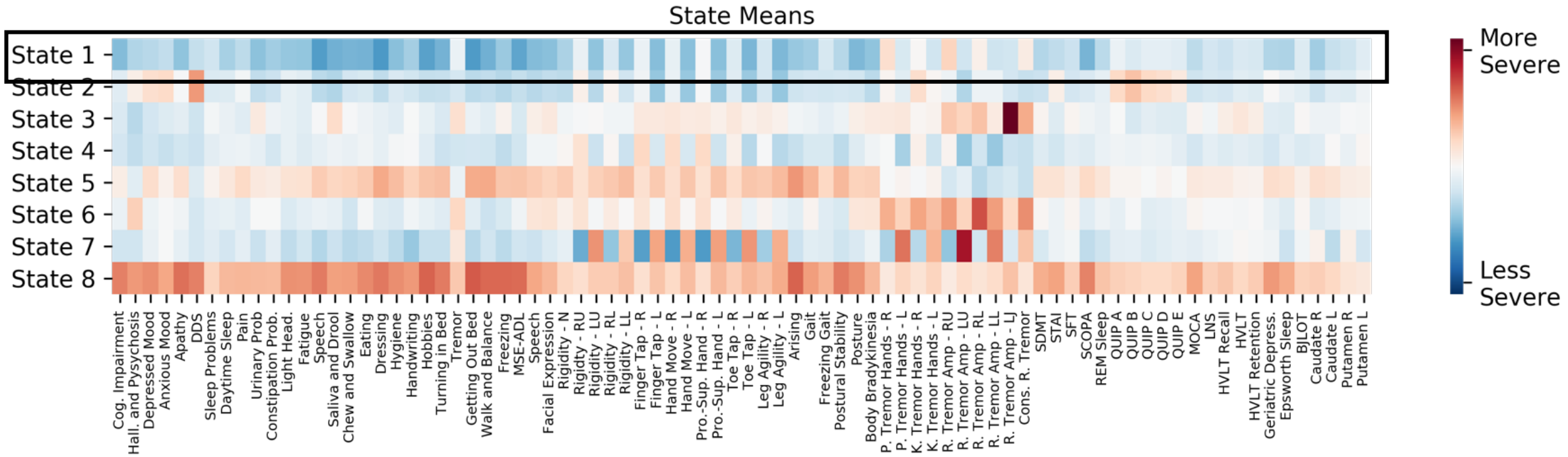
How can we interpret the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Model parameters – State means

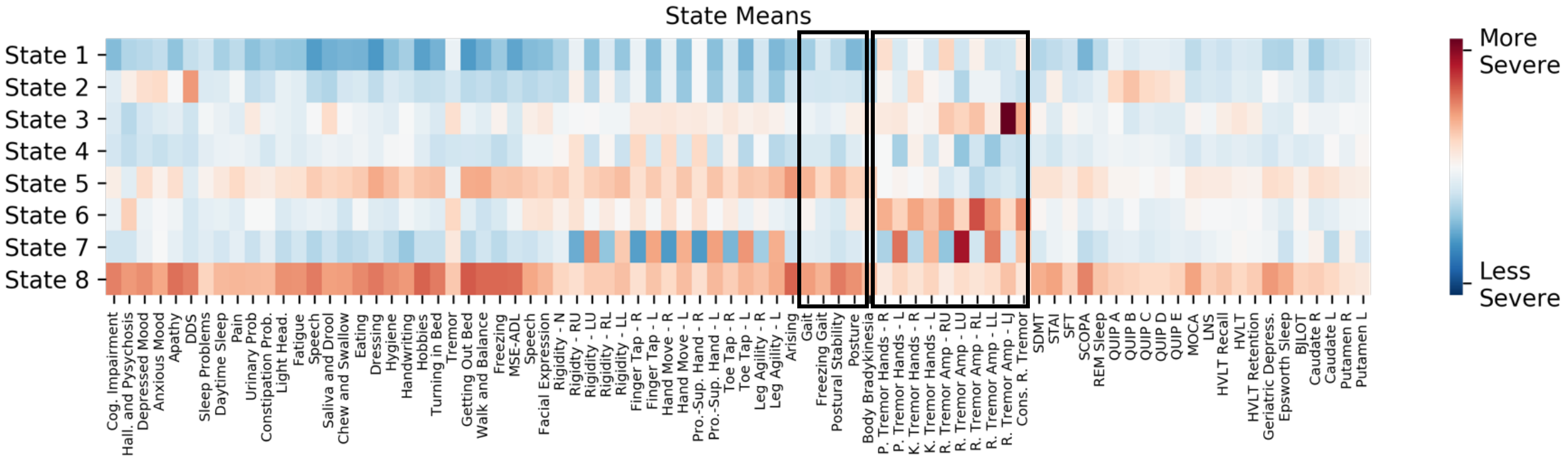
How can we interpret the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Model parameters – State means

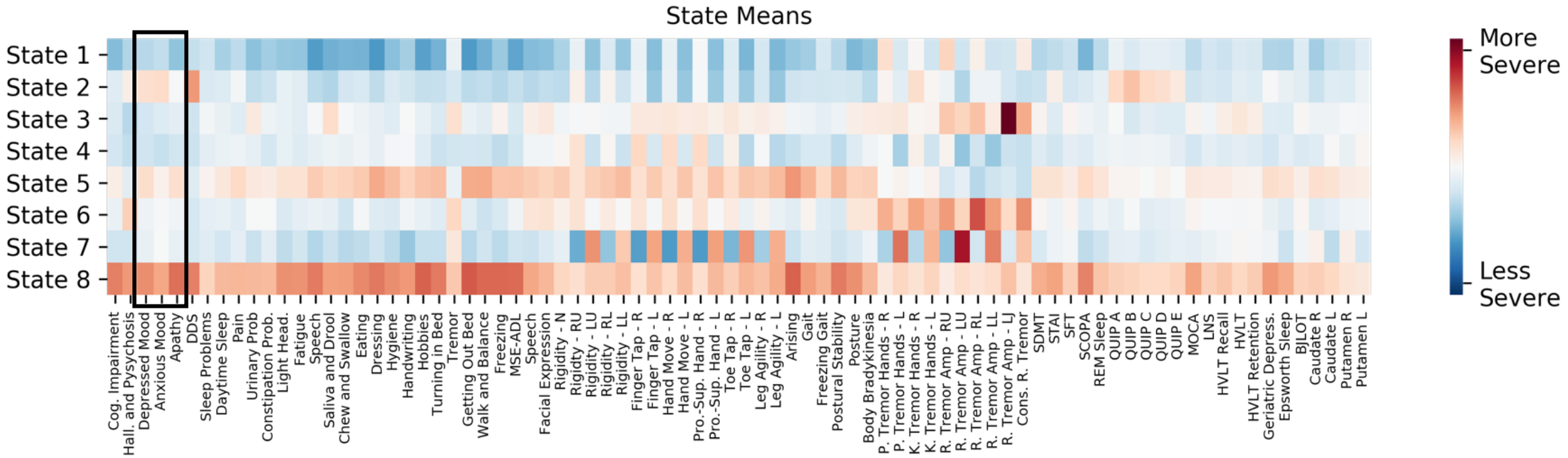
How can we interpret the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Model parameters – State means

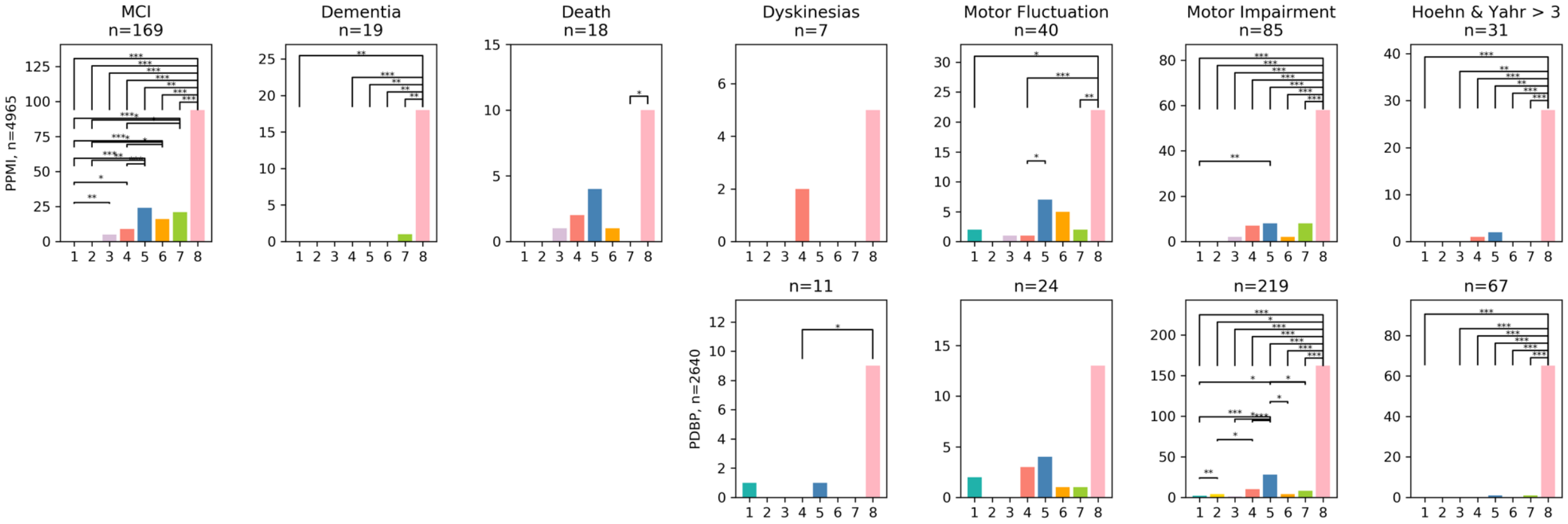
How can we interpret the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Outcome measures

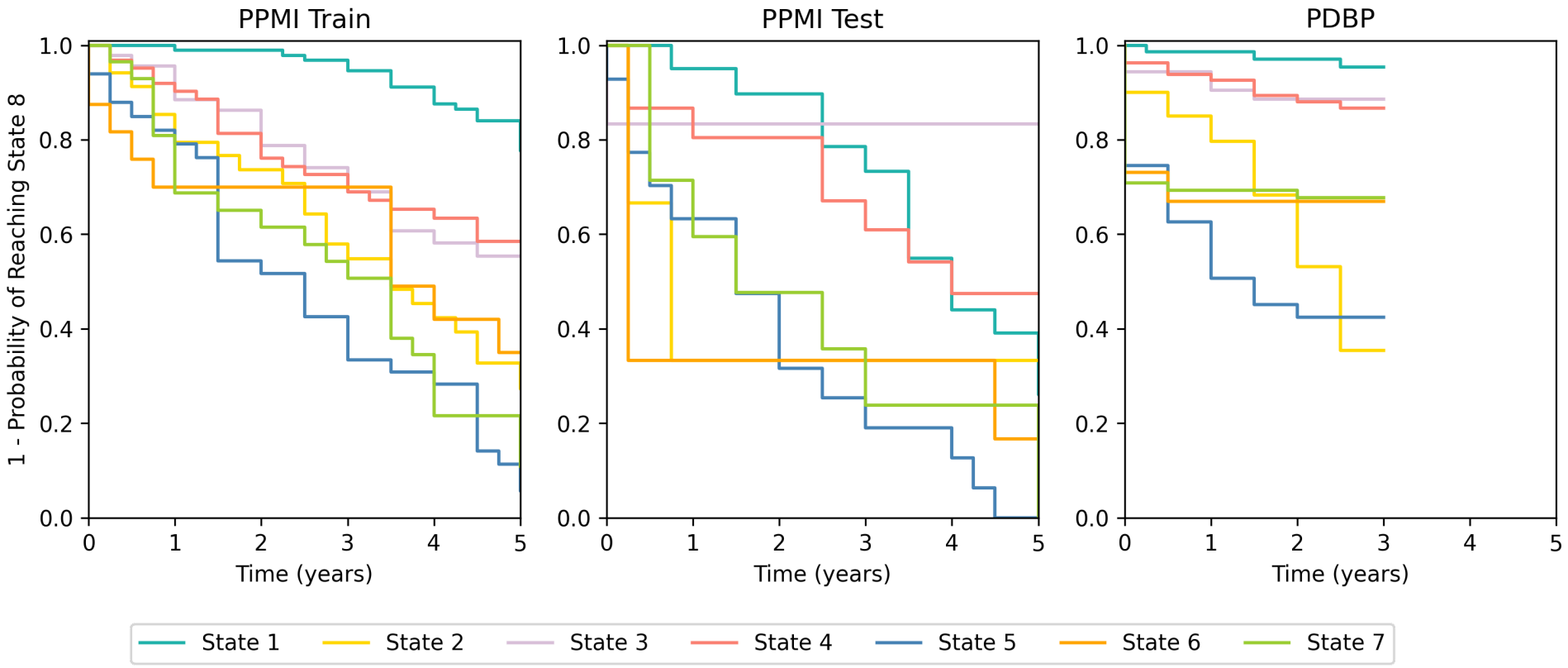
What are the clinical implications of the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Progression pathways

What can we say about the progression of the states?



Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Predictive capability

- What can we say about the predictive capability of the model?
- Predictive question: how well can we predict disease state at year 2 given data from baseline?
- Prediction in this case is defined as the most probable state assignment

PPMI Testing	Condition Positive	Condition Negative	
Predicted Positive	**10**	7	Precision = 0.59
Predicted Negative	14	**52**	NPV = 0.79
	Recall = 0.42	Specificity = 0.88	Accuracy = 0.75 F1 Score = 0.49

PDBP	Condition Positive	Condition Negative	
Predicted Positive	**91**	62	Precision = 0.59
Predicted Negative	52	**405**	NPV = 0.87
	Recall = 0.64	Specificity = 0.87	Accuracy = 0.81 F1 Score = 0.61

Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. 2021.

Conclusions

- Application of the disease progression resulted in novel Parkinson's disease insights and strong predictive performance
- Disease state was found to vary across motor and non-motor symptoms
- The model discovered non-sequential, overlapping disease progression trajectories, supporting the use of non-deterministic disease progression models

Acknowledgements

Project team

IBM, Machine Learning


Kristen Severson


Soumya Ghosh


Jianying Hu


Murtaza Dhuliawala


Kenney Ng

University of Pittsburgh, Neurology


Lana Chahine

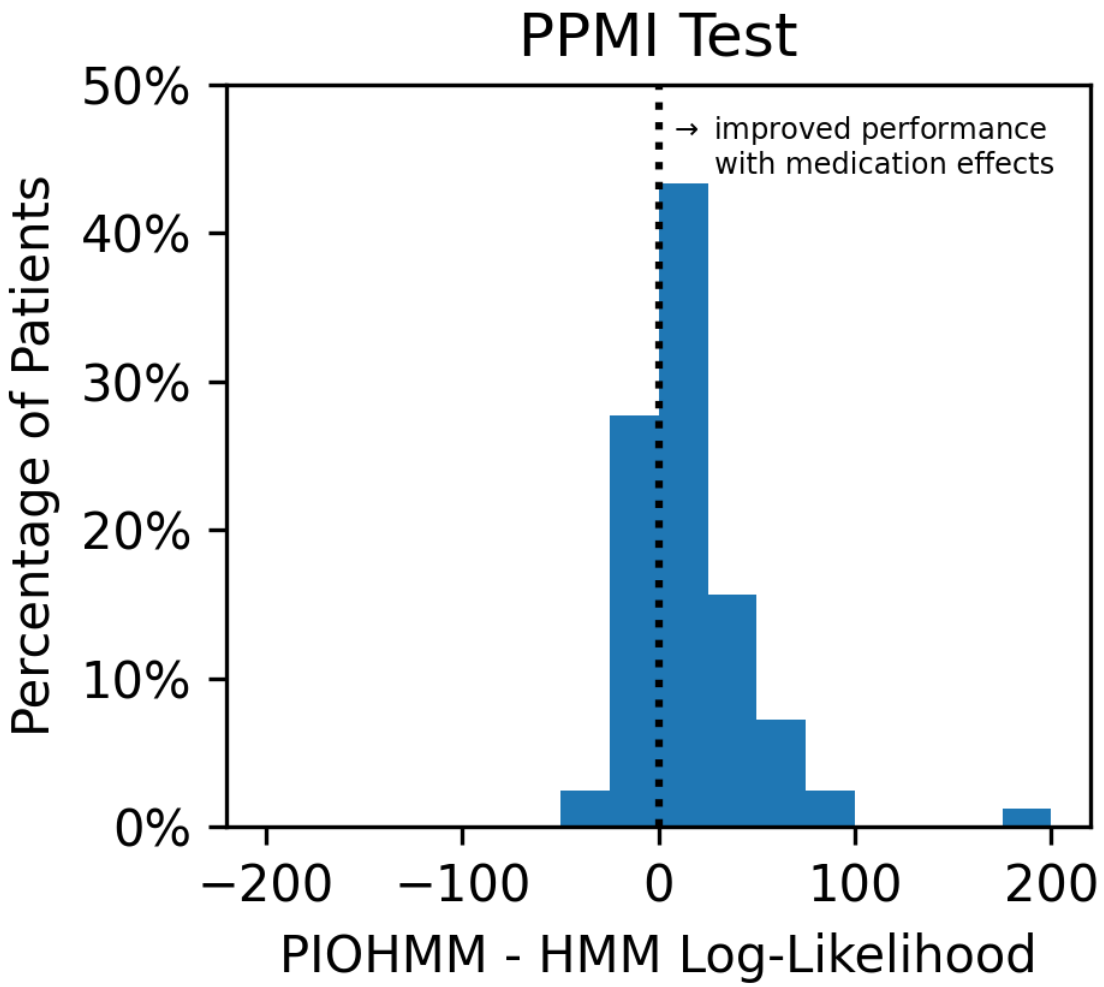
Michael J. Fox Foundation, Advisory

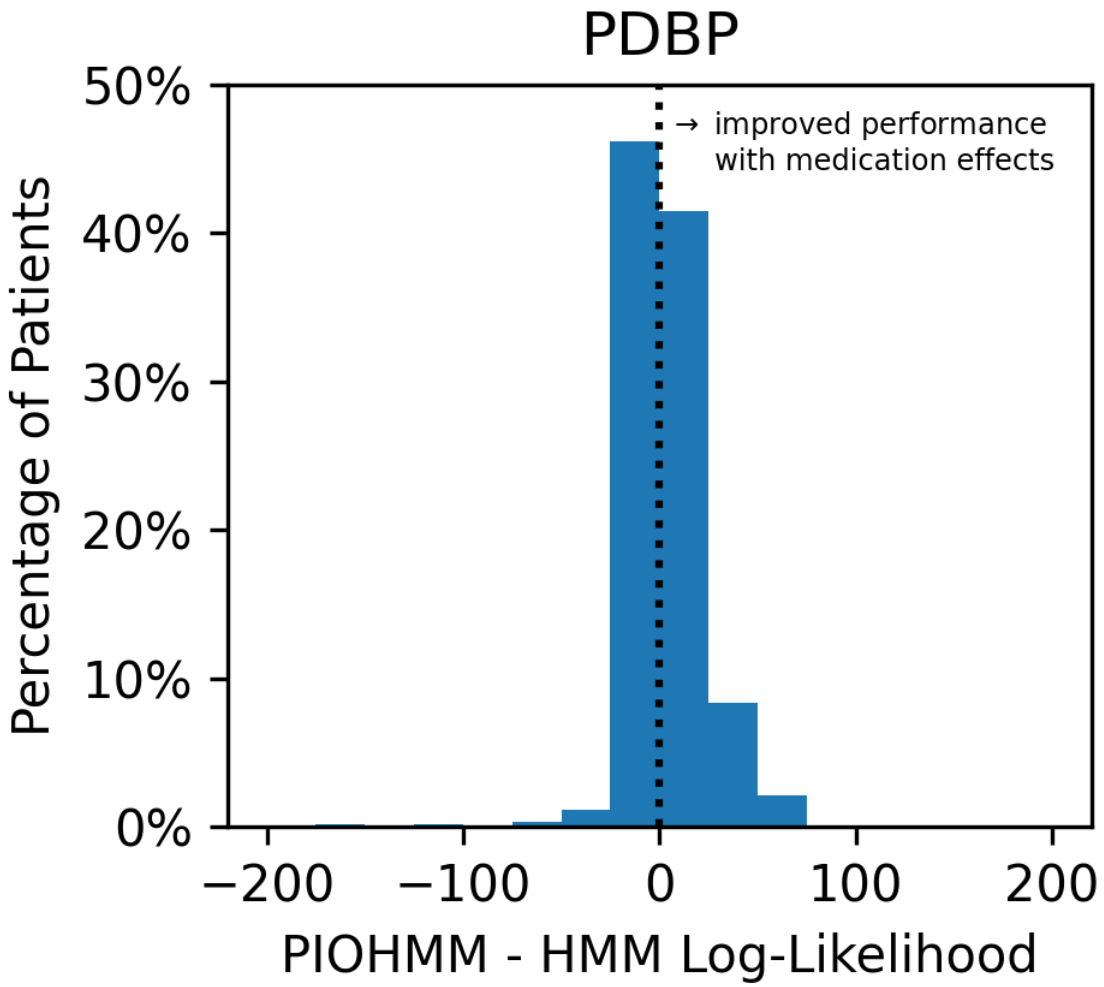

Mark Frasier


Luba Smolensky

Funding provided by the Michael J. Fox Foundation

Comparison to a standard hidden Markov model





Severson et al. Discovery of Parkinson's disease states using machine learning and longitudinal data. *Lancet Digital Health*. Accepted.