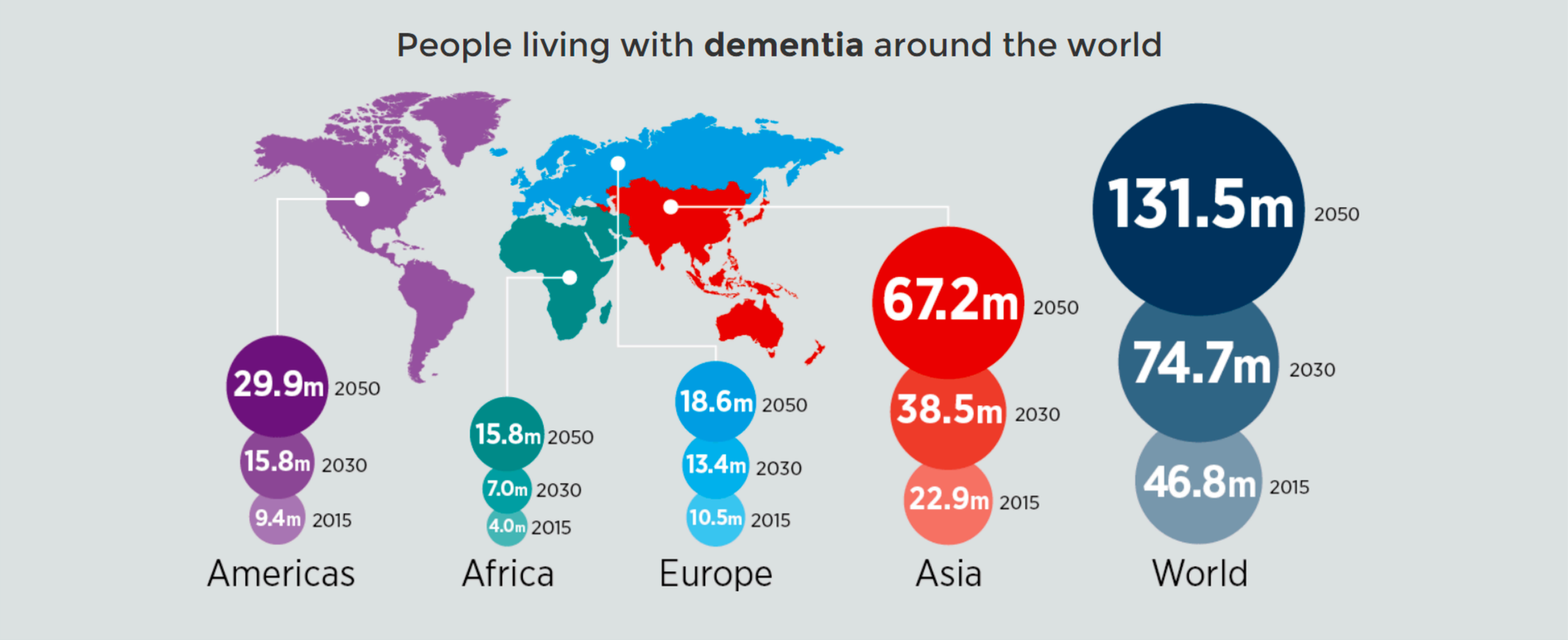
Transfer Learning from typical Alzheimer's disease to rare dementias using

Disease Knowledge Transfer

Razvan Marinescu

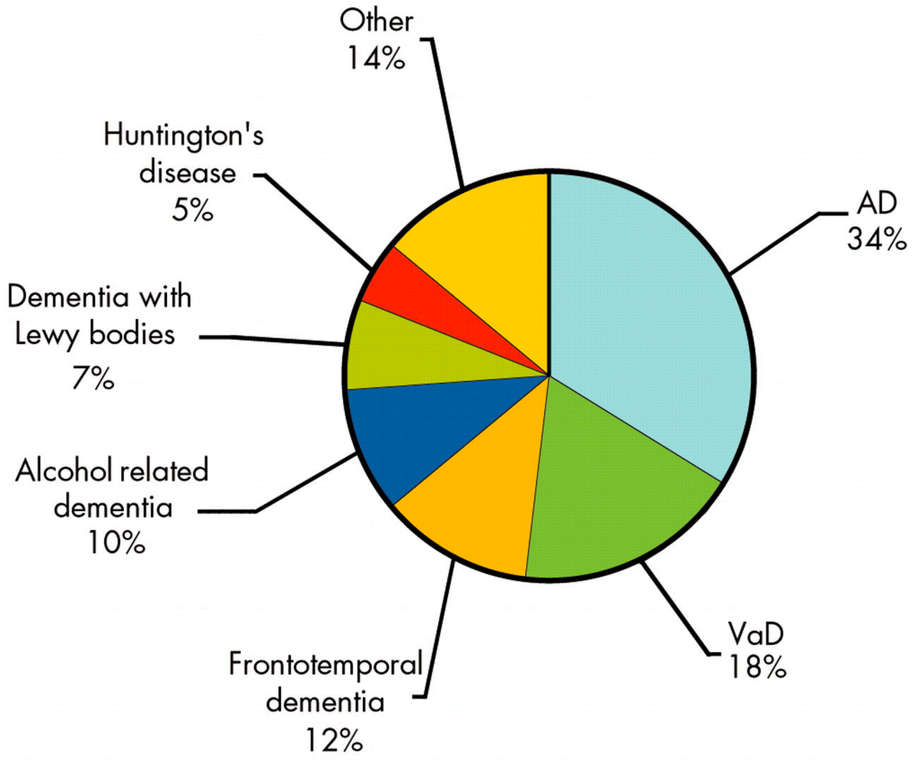
Alzheimer's disease is a devastating disease



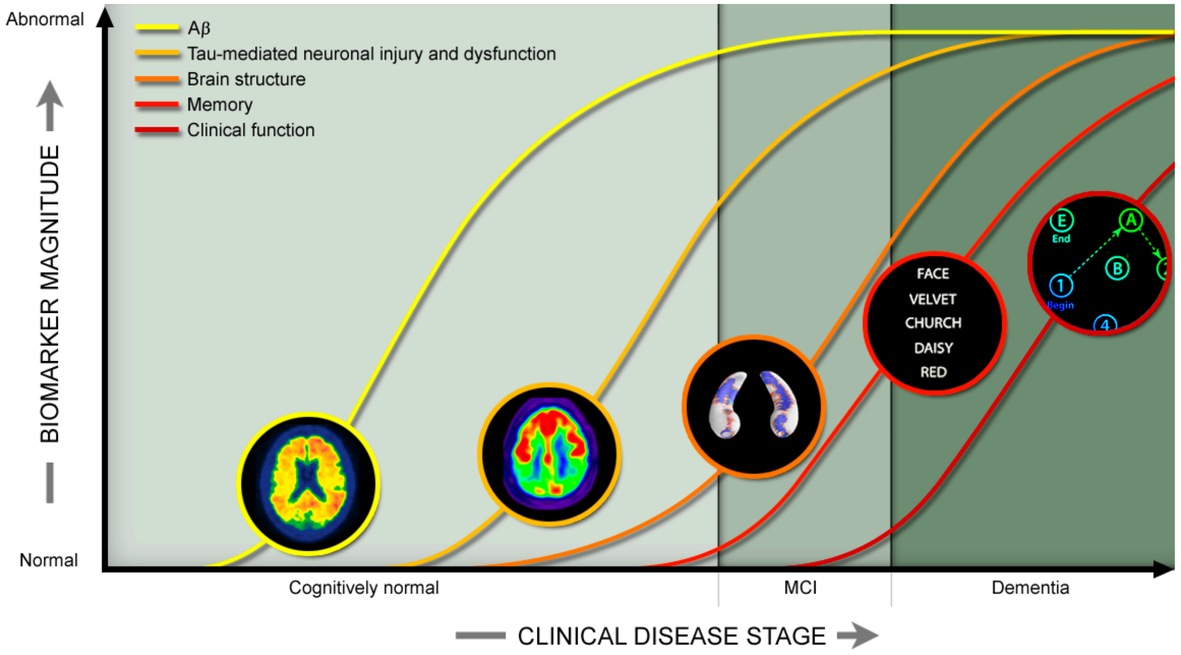
- Currently no treatment available that can stop, or at least slow down, cognitive decline

Neurodegenerative diseases other than Alzheimer's also affect many worldwide

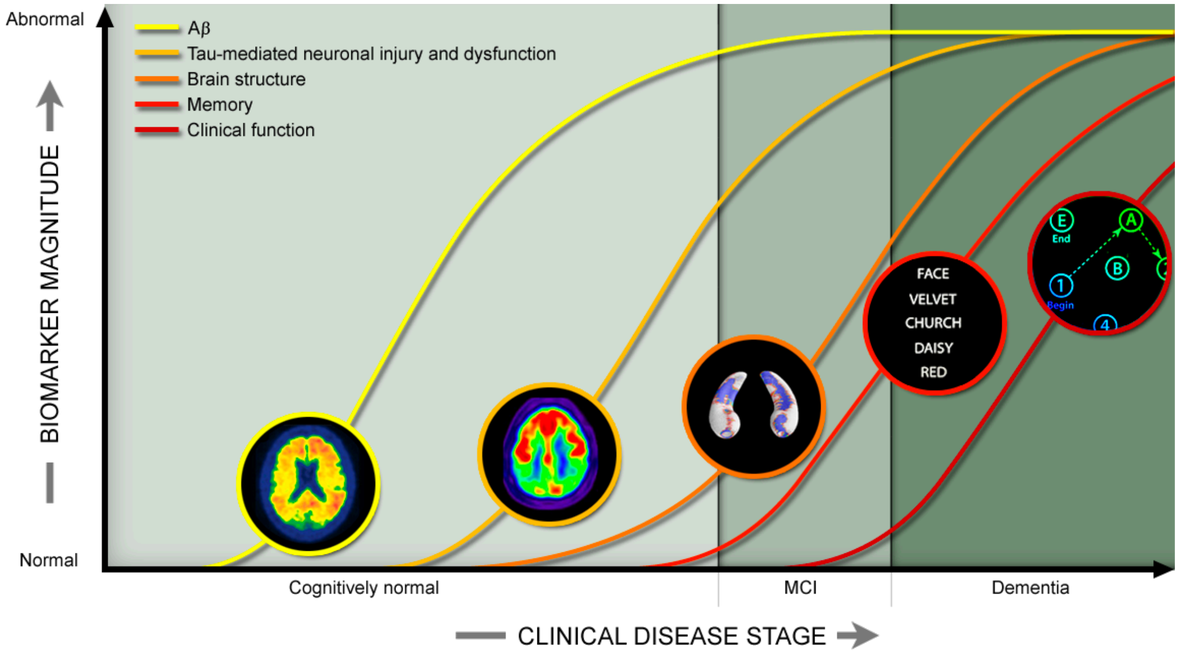
- **Posterior Cortical Atrophy > 1 million**
- Frontotemporal dementia > 6 million
  - All tauopathies …
- Dementia with Lewy bodies > 1.6 million
- Vascular dementia > 8 million
- Creutzfeld-Jacobs disease > 7000/year
- Parkinson's disease
- Huntington's disease



Progression of Alzheimer's disease *is known*



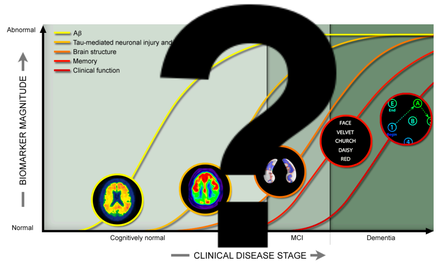
Progression of Alzheimer's disease *is known*



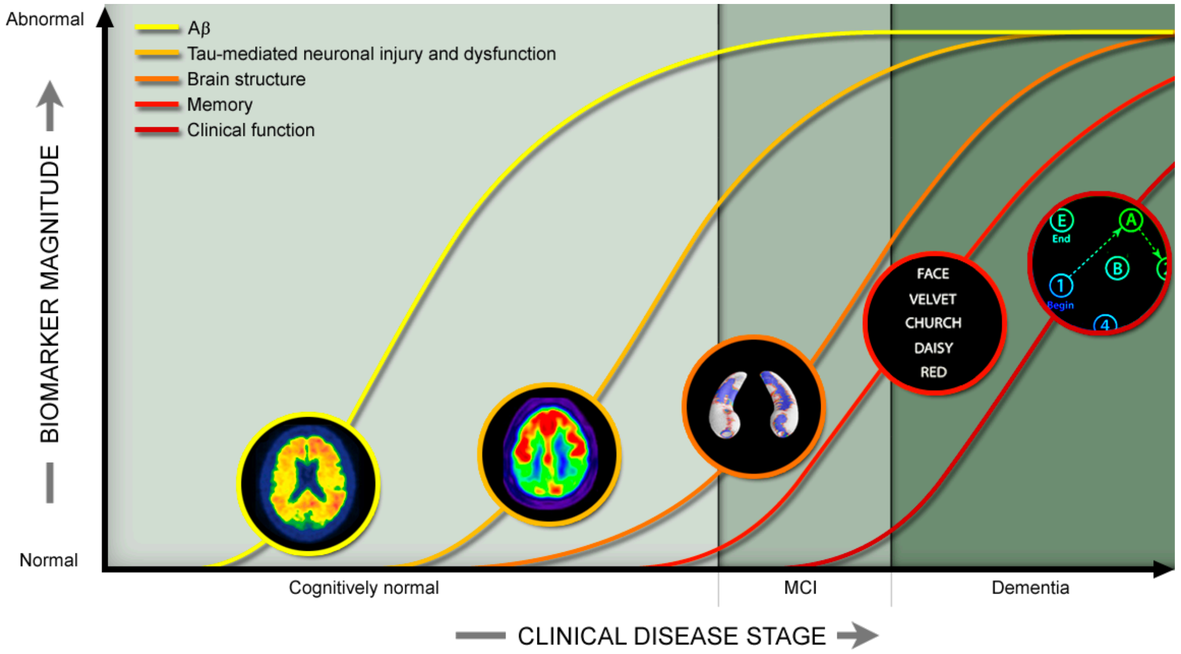
Progression of less common neurodegenerative diseases *is not known*

Lack of datasets that are:

- Large
- Longitudinal
- Multimodal



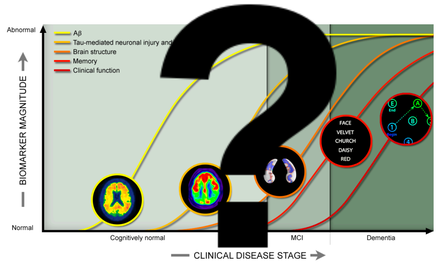
Progression of Alzheimer's disease *is known*

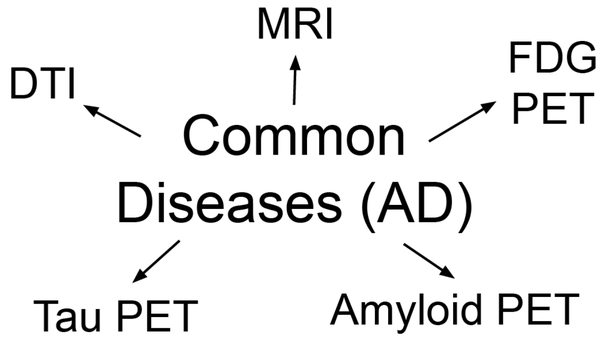


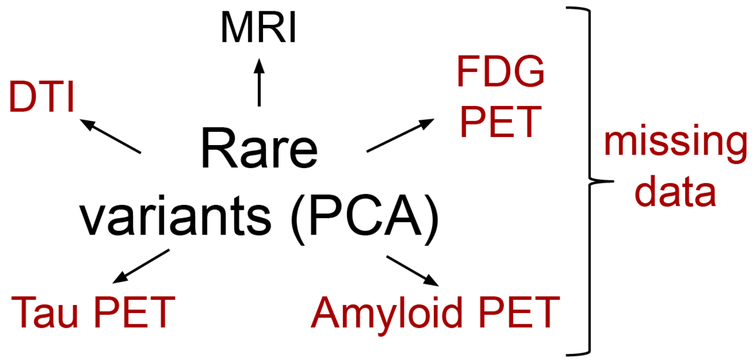
Progression of less common neurodegenerative diseases *is not known*

Lack of datasets that are:

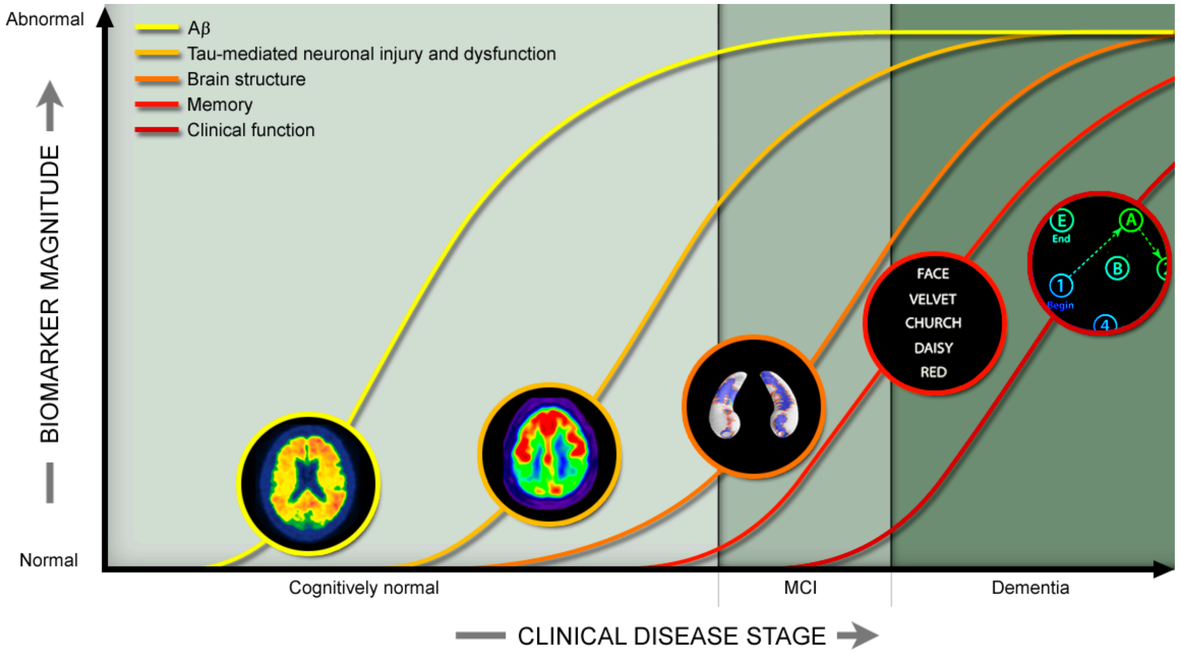
- Large
- Longitudinal
- Multimodal







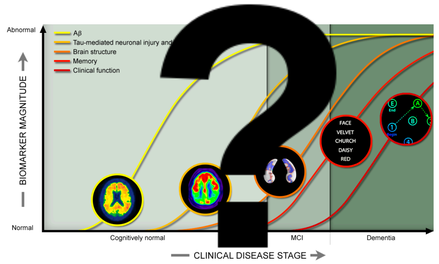
Progression of Alzheimer's disease *is known*



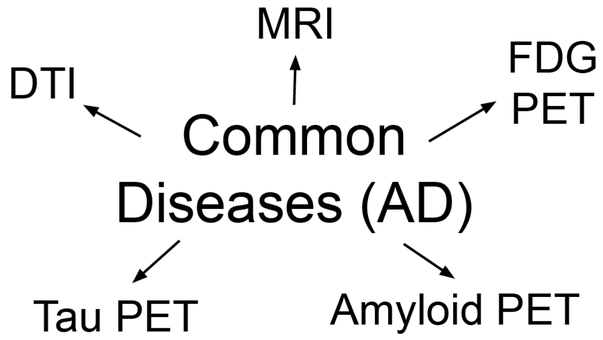
Progression of less common neurodegenerative diseases *is not known*

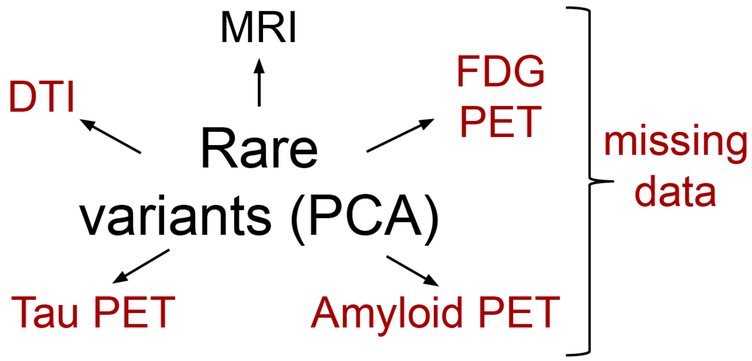
Lack of datasets that are:

- Large
- Longitudinal
- Multimodal



Transfer learning provides a key solution towards characterizing rare diseases





Previous literature on Transfer Learning for neurodegenerative diseases

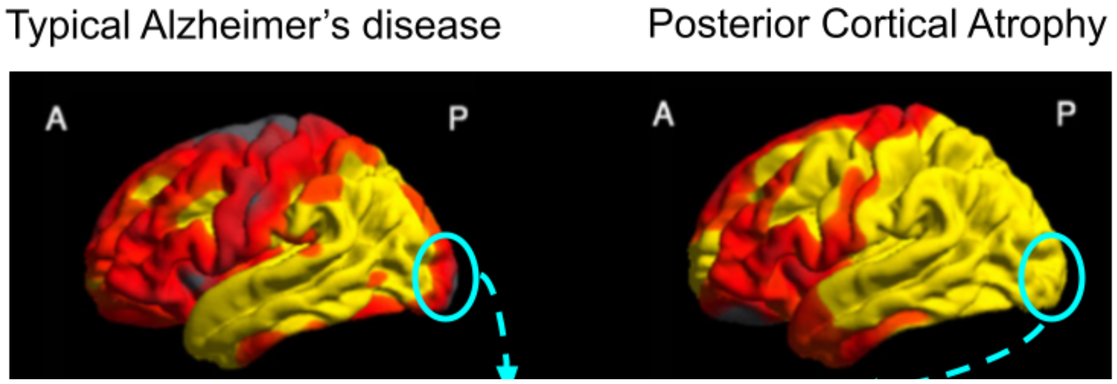
- Hon and Khan 2017, Nanni et. al. 2020 - transfer from computer vision datasets to medical datasets
- Cheng, Zhang and Shen 2012, Wachinger and Reuter, 2017, Guerrero et. al. 2014, Hofer et. al. 2017 - transfer learning across Alzheimer's disease diagnoses (e.g. CN vs MCI -> MCI vs AD)
- Methods are supervised on clinical diagnosis, which is **unreliable without post-mortem neuropathology**
- No work tried to use transfer learning to improve predictions on **rarer** neurodegenerative diseases

Reference	Topic	Task	Domain	Transfer type
Brain				
Zhang and Shen (2012)	MCI conversion prediction	different	same	feature, multi-task
Guerrero et al. (2014)	AD classification	same	different	instance, align
Wachinger and Reuter (2016)	AD classification	same	different	instance, weight
Hofer et al. (2017)	AD classification	same	different	instance, align
Hon and Khan (2017)	AD classification	different	different	feature, pretraining

Survey of transfer learning in Alzheimer's research (Cheplygina et. al., 2019)

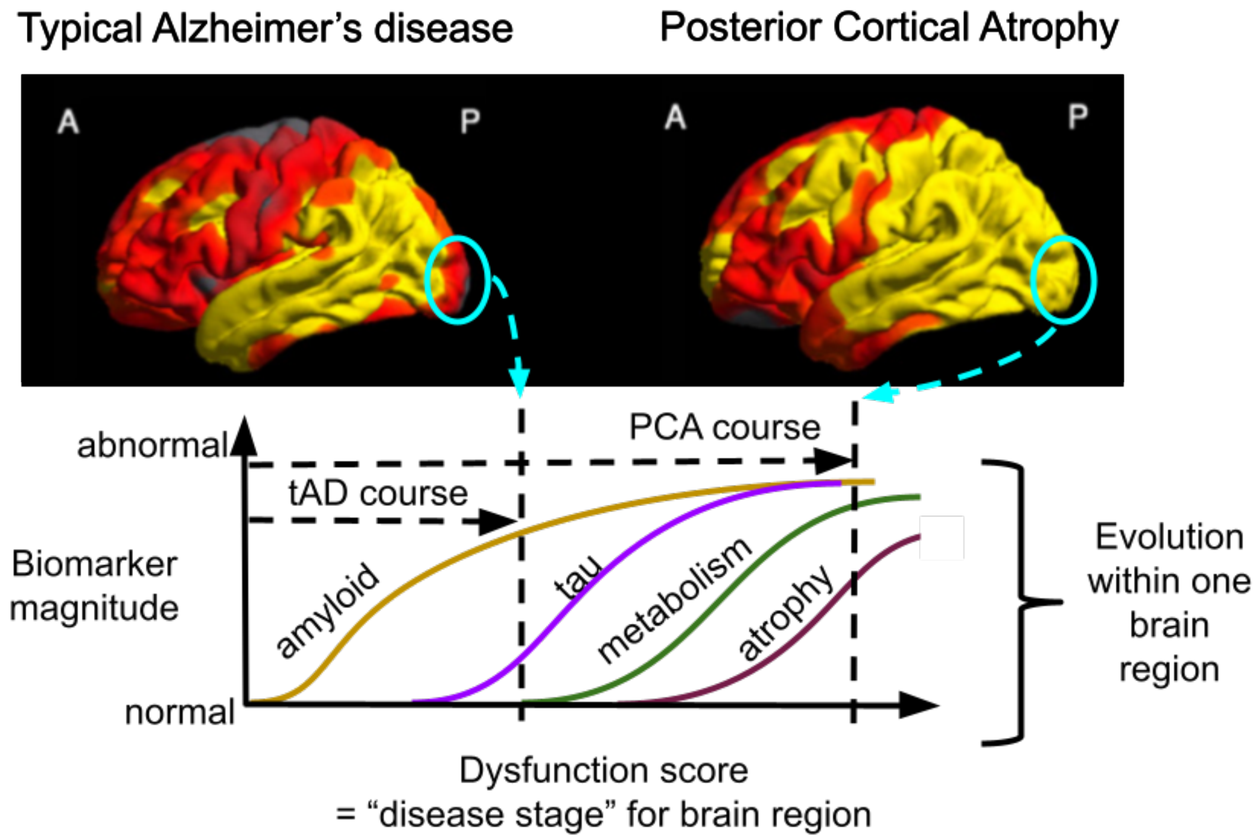
Transfer Learning intuition: sharing the disease progression template but not the extent of damage

- Two diseases such as typical AD and Posterior Cortical Atrophy (PCA) affect the brain at different spatial locations



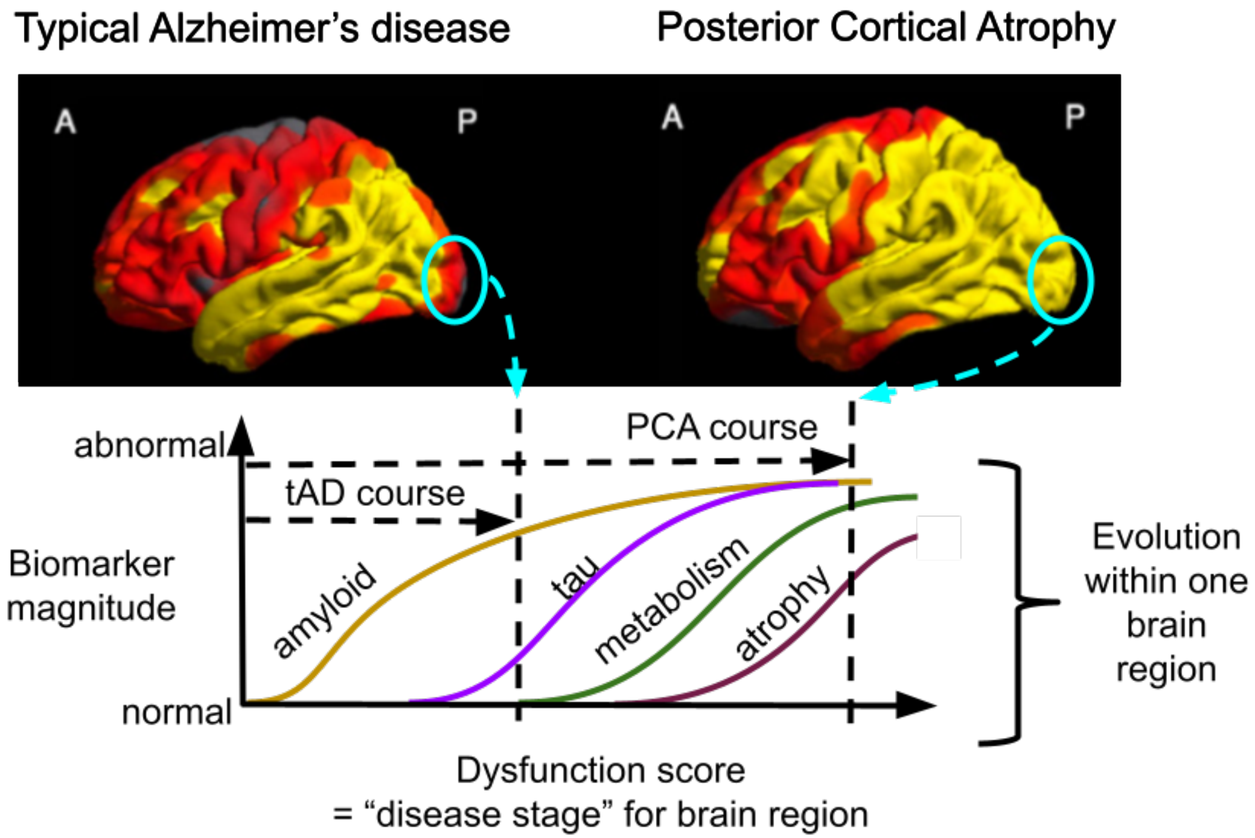
Transfer Learning intuition: sharing the disease progression template but not the extent of damage

- Two diseases such as typical AD and Posterior Cortical Atrophy (PCA) affect the brain at different spatial locations
- Each region, e.g. occipital lobe, is believed to follow a certain cascade of events



Transfer Learning intuition: sharing the disease progression template but not the extent of damage

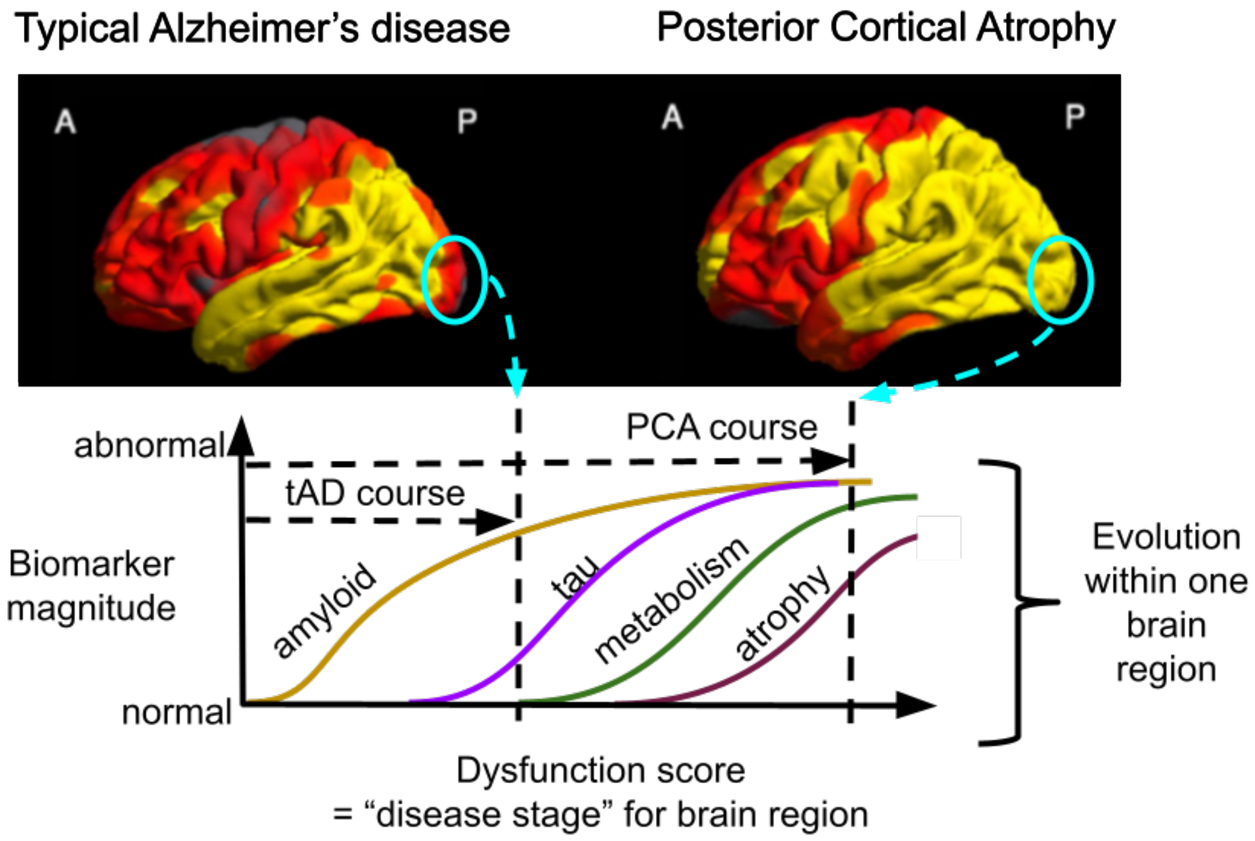
- Two diseases such as typical AD and Posterior Cortical Atrophy (PCA) affect the brain at different spatial locations
- Each region, e.g. occipital lobe, is believed to follow a certain cascade of events
- We propose that each region ***follows the same multimodal trajectories*** for both typical AD and PCA
- Difference between typical AD vs PCA is ***the extent of pathology*** along the trajectory



Marinescu et al., MICCAI, 2019

Transfer Learning intuition: sharing the disease progression template but not the extent of damage

- Two diseases such as typical AD and Posterior Cortical Atrophy (PCA) affect the brain at different spatial locations
- Each region, e.g. occipital lobe, is believed to follow a certain cascade of events
- We propose that each region ***follows the same multimodal trajectories*** for both typical AD and PCA
- Difference between typical AD vs PCA is ***the extent of pathology*** along the trajectory
- Current understanding: PCA, as a different syndrome, is modeled separately from tAD

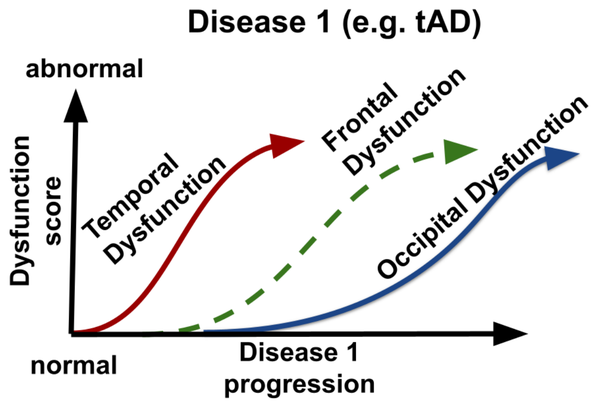


Marinescu et al., MICCAI, 2019

The Disease Knowledge Transfer (DKT) framework

The Disease Knowledge Transfer (DKT) framework

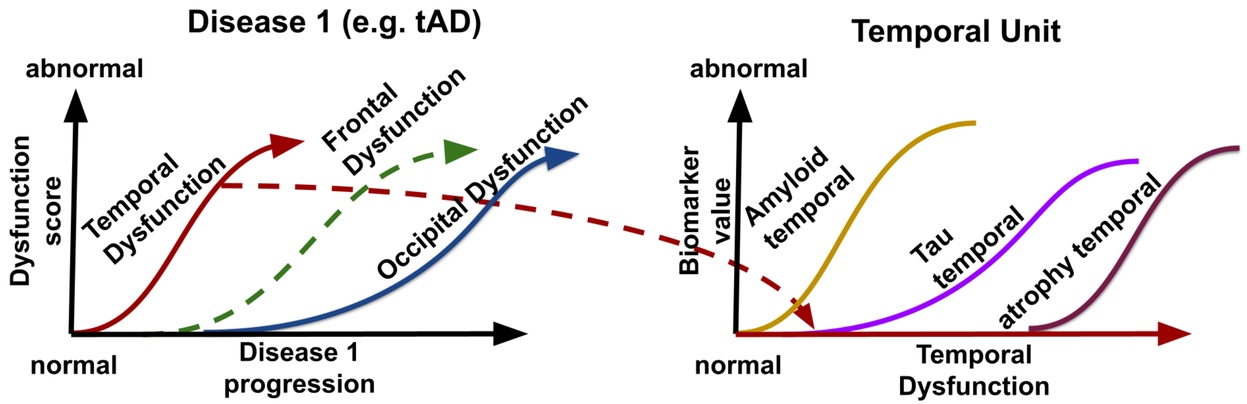
- Model disease as progression of composite dysfunction scores for each brain region:
  - Typical AD: temporal first
  - PCA: occipital first



Marinescu et al., MICCAI, 2019

The Disease Knowledge Transfer (DKT) framework

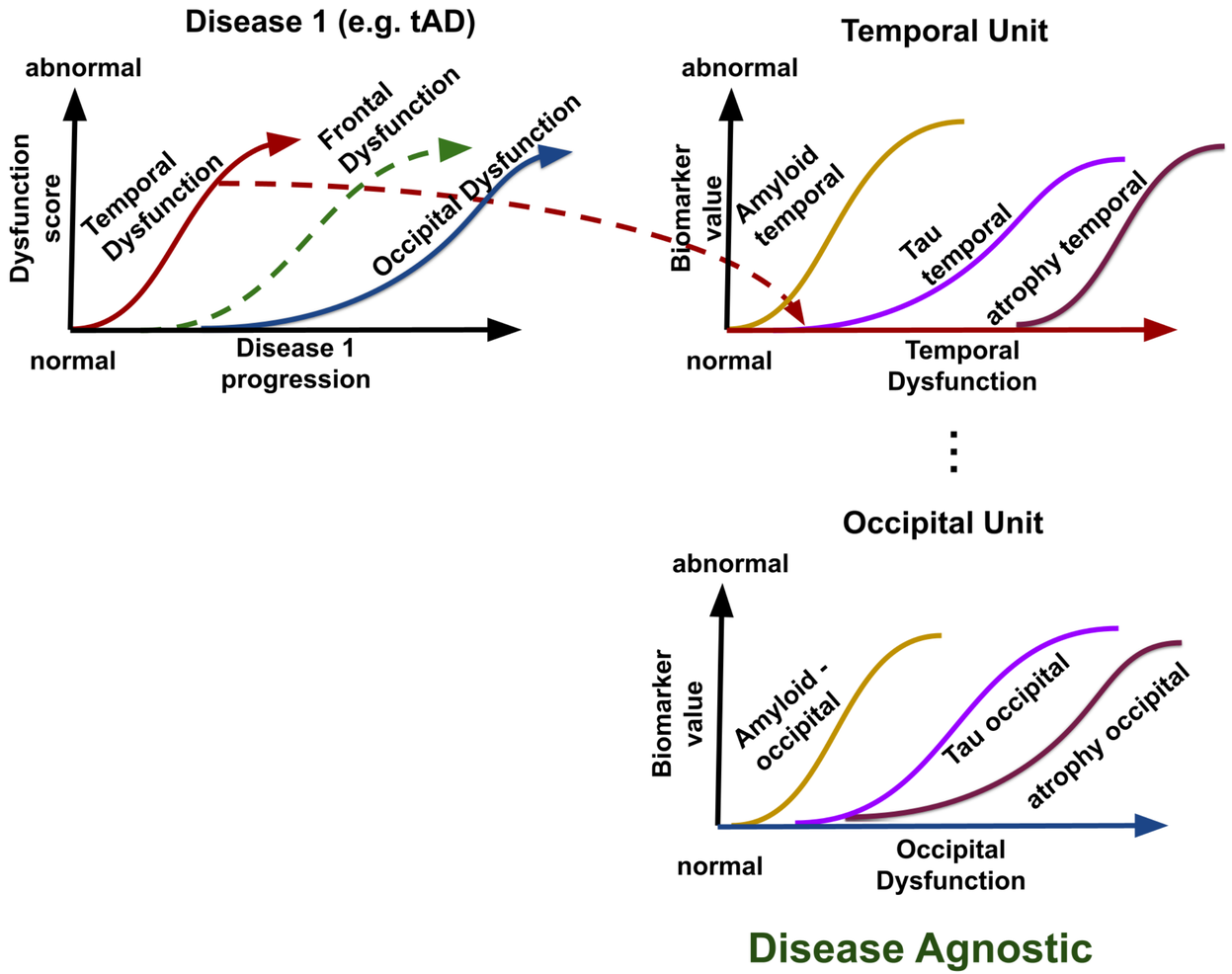
- Model disease as progression of composite dysfunction scores for each brain region:
  - Typical AD: temporal first
  - PCA: occipital first
- Model dysfunction scores as “aggregate pathology” from multiple modalities (e.g. amyloid + tau + atrophy)



Marinescu et al., MICCAI, 2019

The Disease Knowledge Transfer (DKT) framework

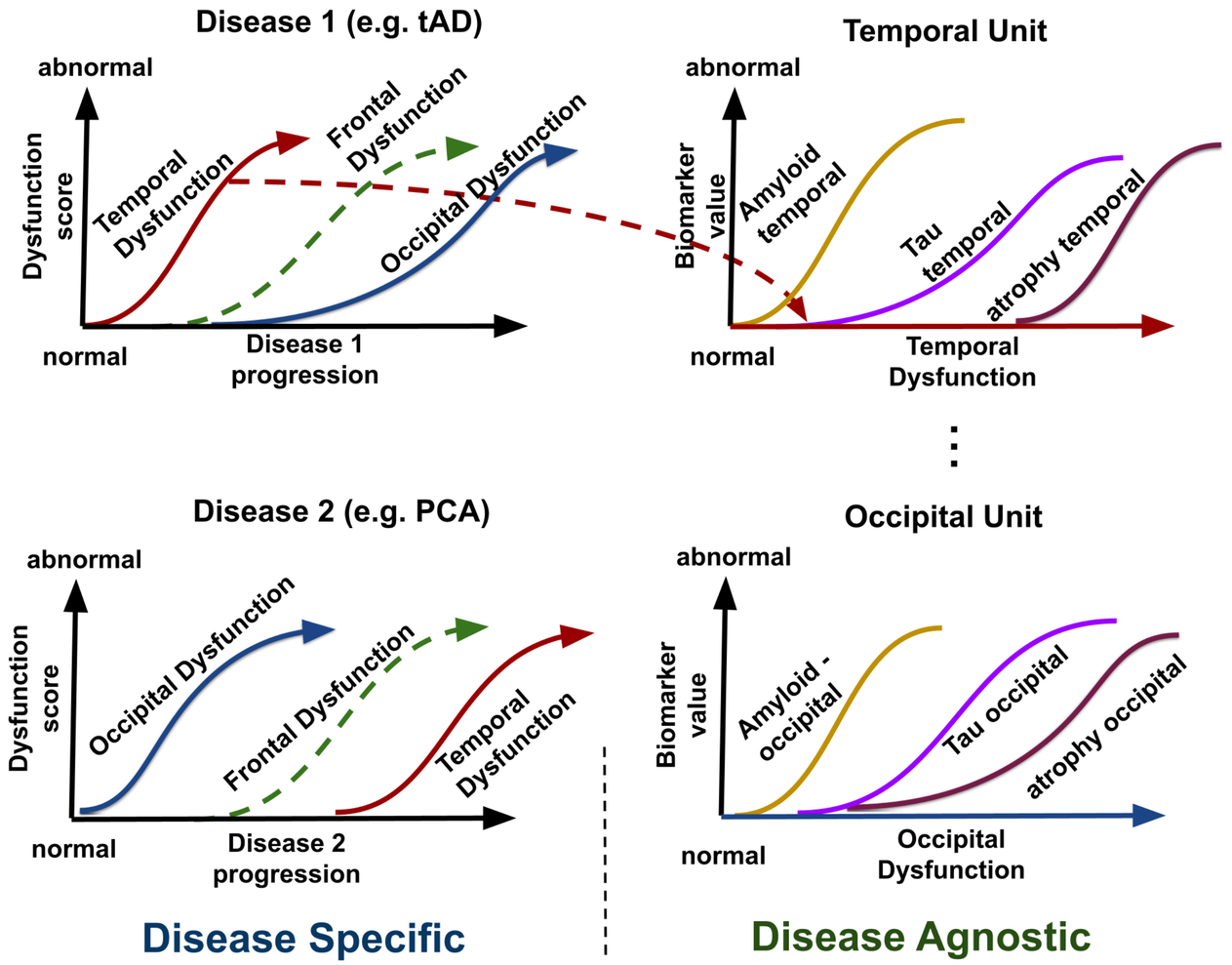
- Model disease as progression of composite dysfunction scores for each brain region:
  - Typical AD: temporal first
  - PCA: occipital first
- Model dysfunction scores as “aggregate pathology” from multiple modalities (e.g. amyloid + tau + atrophy)
- Extend dysfunction modeling to all brain regions



Marinescu et al., MICCAI, 2019

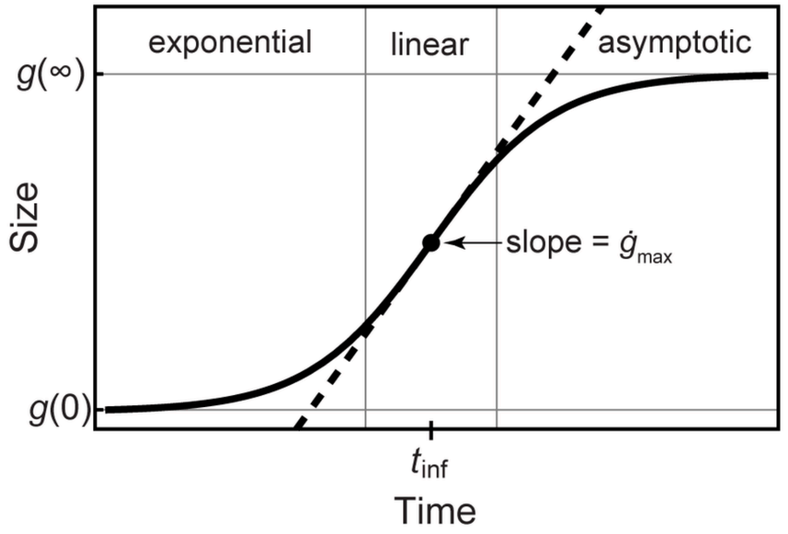
The Disease Knowledge Transfer (DKT) framework

- Model disease as progression of composite dysfunction scores for each brain region:
  - Typical AD: temporal first
  - PCA: occipital first
- Model dysfunction scores as “aggregate pathology” from multiple modalities (e.g. amyloid + tau + atrophy)
- Extend dysfunction modeling to all brain regions
- A new disease, e.g. Posterior Cortical Atrophy (PCA) will have **different** dysfunction progression across the brain (disease specific), but **similar** progression within individual regions (disease agnostic)



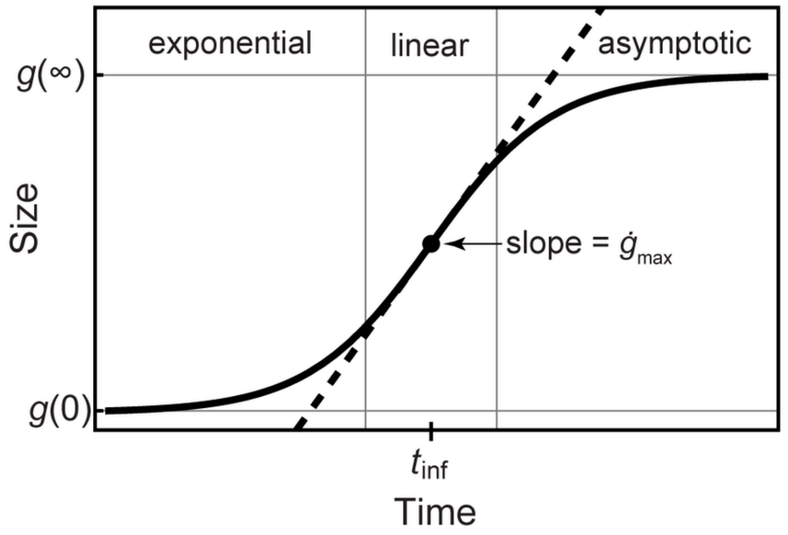
Marinescu et al., MICCAI, 2019

DKT is a bayesian hierarchical model



DKT is a bayesian hierarchical model

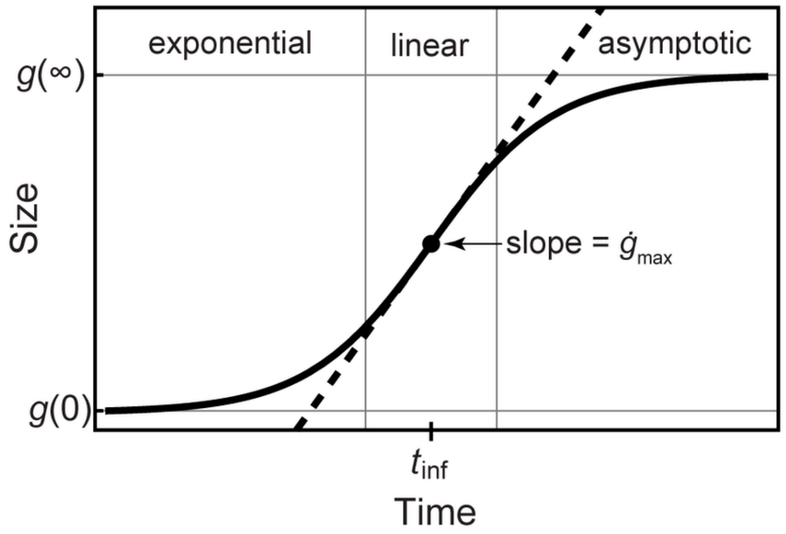
- Define subject disease stage:



DKT is a bayesian hierarchical model

- Define subject disease stage:

$$\beta_i + m_{ij}$$

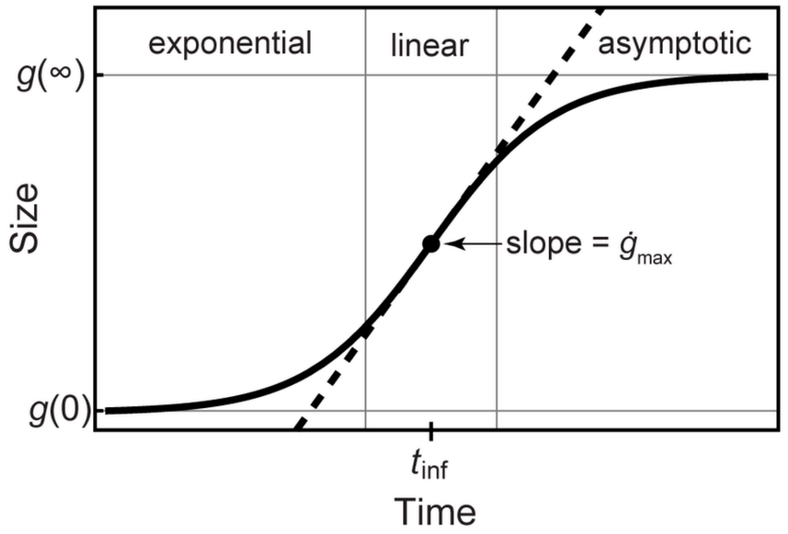


DKT is a bayesian hierarchical model

- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:



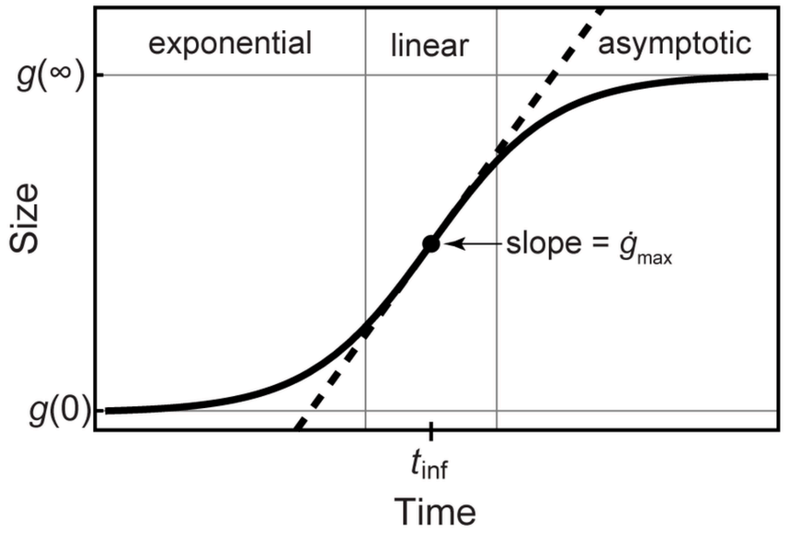
DKT is a bayesian hierarchical model

- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:

$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$



DKT is a bayesian hierarchical model

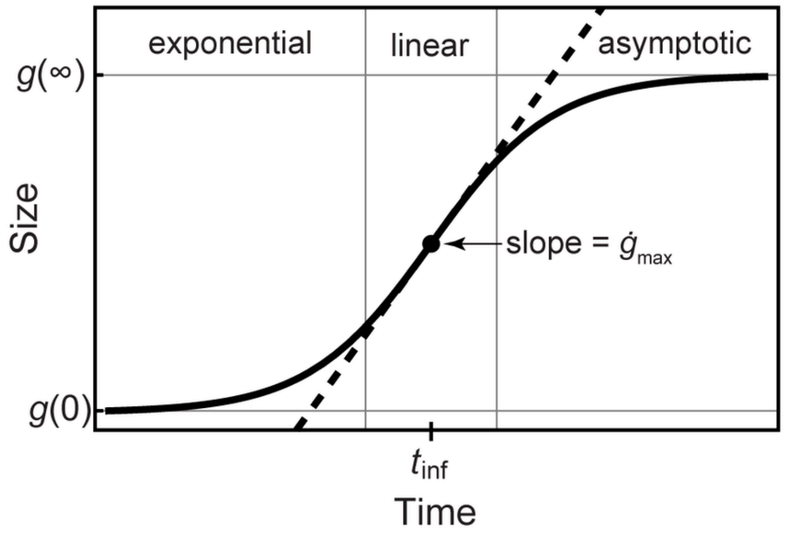
- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:

$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$

- Estimate value of a biomarker of a particular modality given the dysfunction score:



DKT is a bayesian hierarchical model

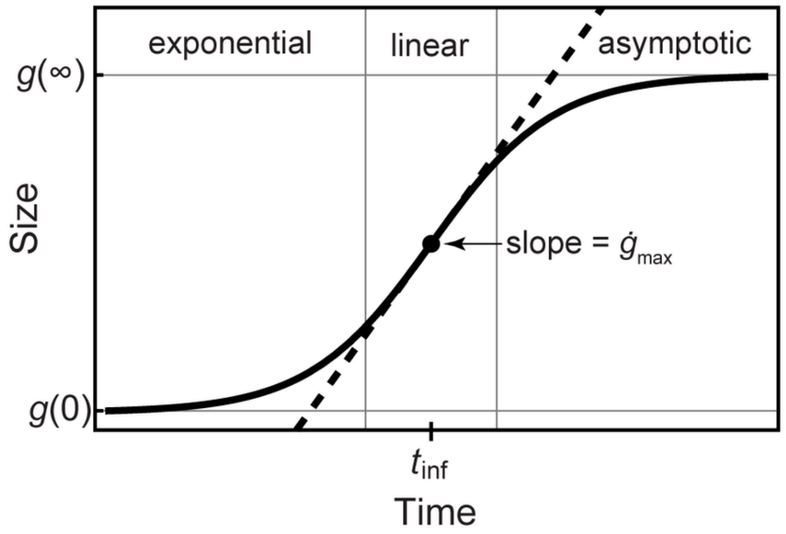
- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:

$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$

- Estimate value of a biomarker of a particular modality given the dysfunction score:



DKT is a bayesian hierarchical model

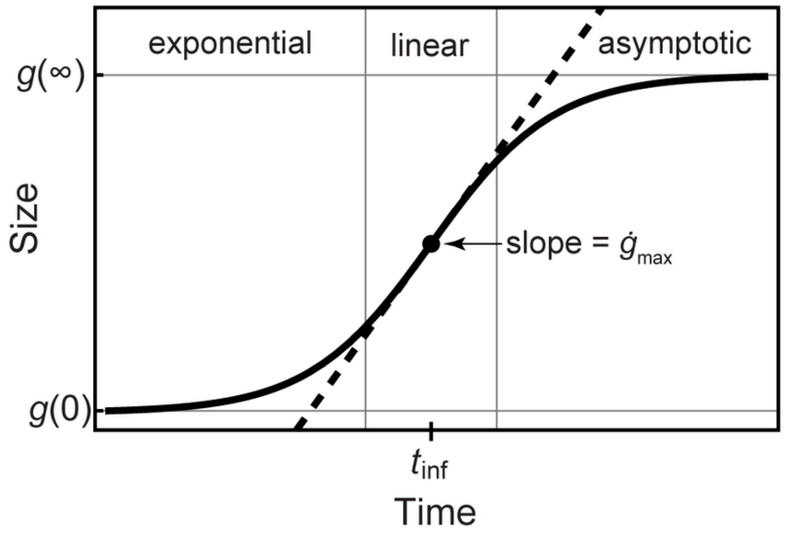
- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:

$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$

- Estimate value of a biomarker of a particular modality given the dysfunction score:



- Parameters are estimated through loopy belief propagation

DKT is a bayesian hierarchical model

- Define subject disease stage:

$$\beta_i + m_{ij}$$

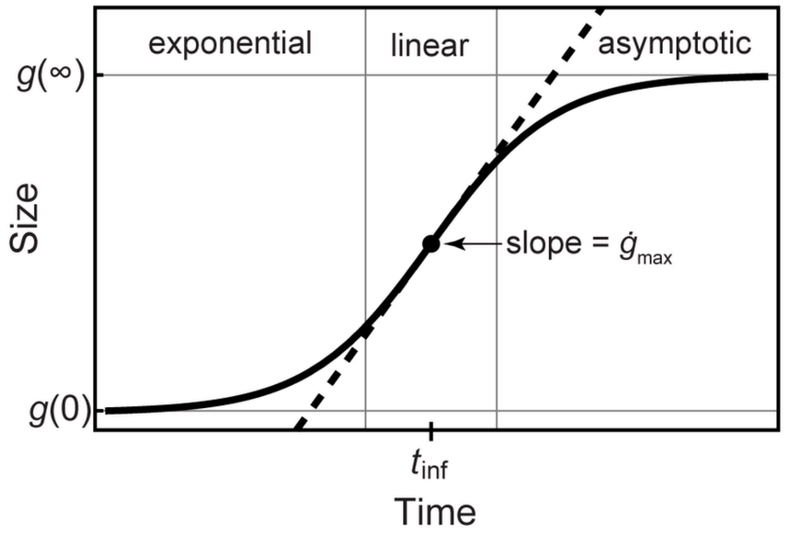
- Estimate dysfunction score of a particular brain ROI:

$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$

- Estimate value of a biomarker of a particular modality given the dysfunction score:

$$y_{ijk} \sim g(f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)}); \theta_k) + \epsilon_k$$

- Parameters are estimated through loopy belief propagation



DKT is a bayesian hierarchical model

- Define subject disease stage:

$$\beta_i + m_{ij}$$

- Estimate dysfunction score of a particular brain ROI:

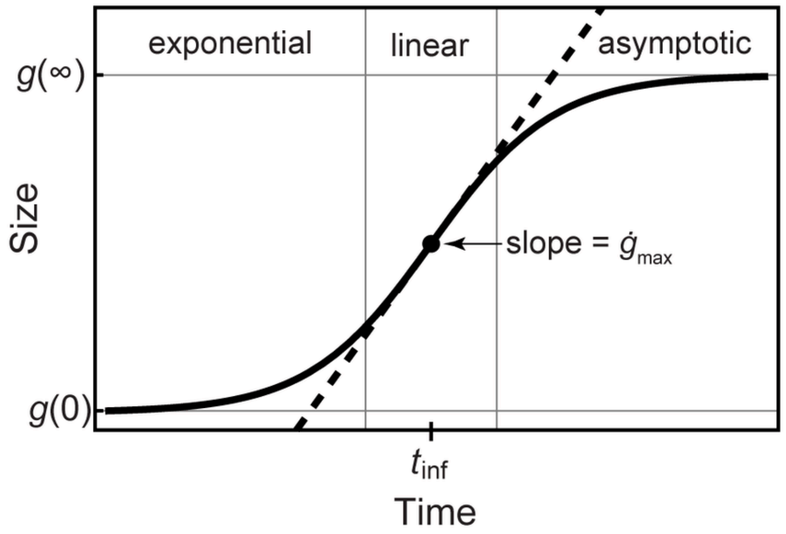
$$f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)})$$

- Estimate value of a biomarker of a particular modality given the dysfunction score:

$$y_{ijk} \sim g(f(\beta_i + m_{ij}; \lambda_{d_i}^{\psi(k)}); \theta_k) + \epsilon_k$$

- Parameters are estimated through loopy belief propagation

- Functions *f* and *g* are parameterized using sigmoidal curves



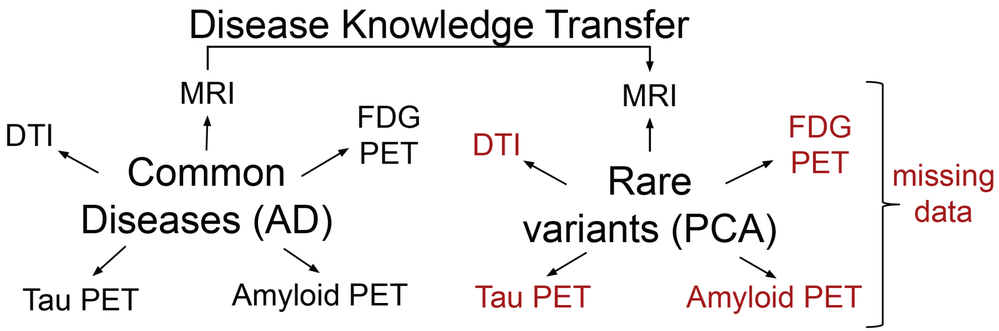
Outline of Results

- Results on simulated data
- Results on patient data from ADNI and the Dementia Research Center UK
- Quantitative evaluation

DKT works well on simulated data

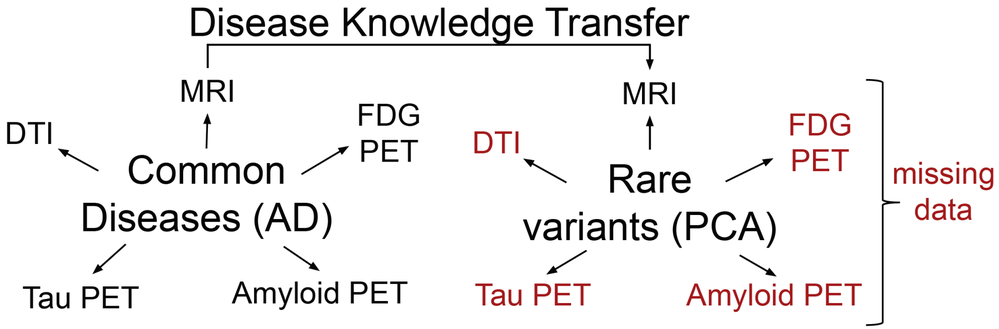
- Simulated 100 subjects with two diseases: synAD & synPCA

- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers



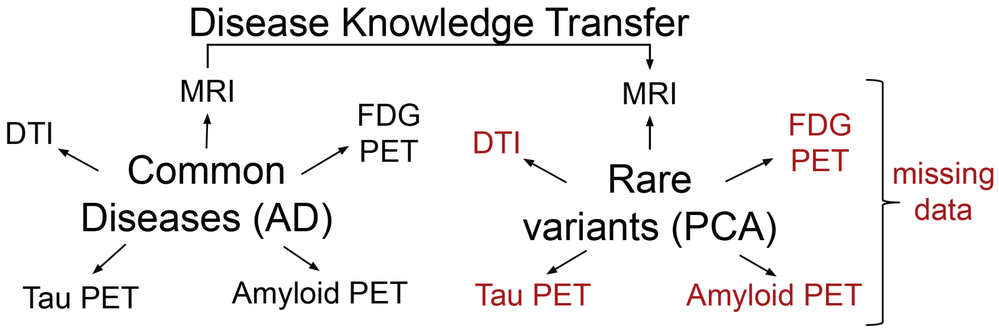
DKT works well on simulated data

- Simulated 100 subjects with two diseases: synAD & synPCA
- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers
- DKT was able to:



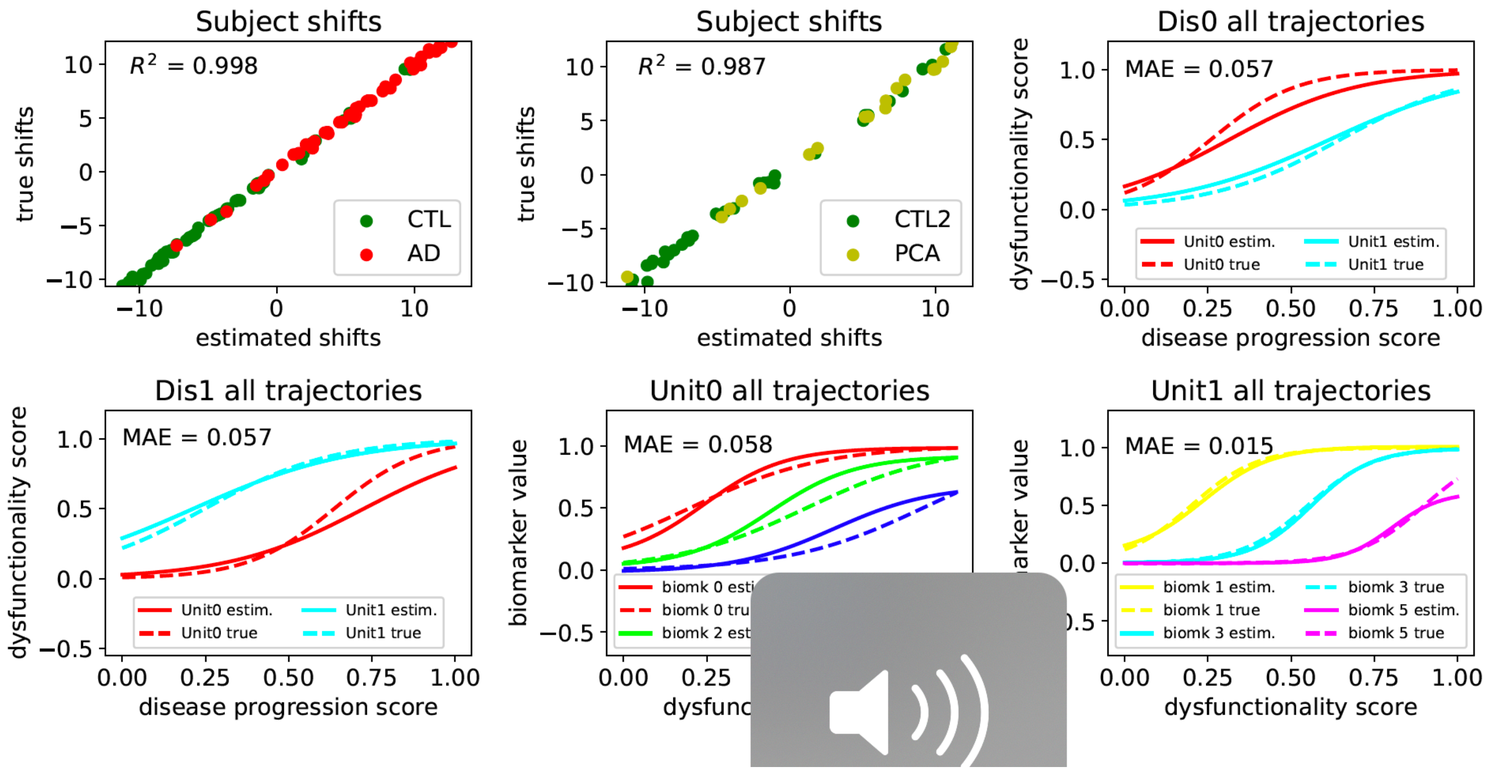
DKT works well on simulated data

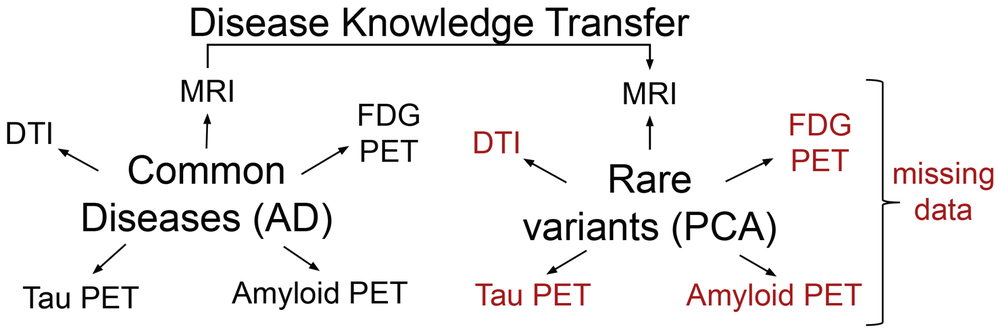
- Simulated 100 subjects with two diseases: synAD & synPCA
- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers
- DKT was able to:
  - Reliably fit the data



DKT works well on simulated data

- Simulated 100 subjects with two diseases: synAD & synPCA
- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers
- DKT was able to:
  - Reliably fit the data

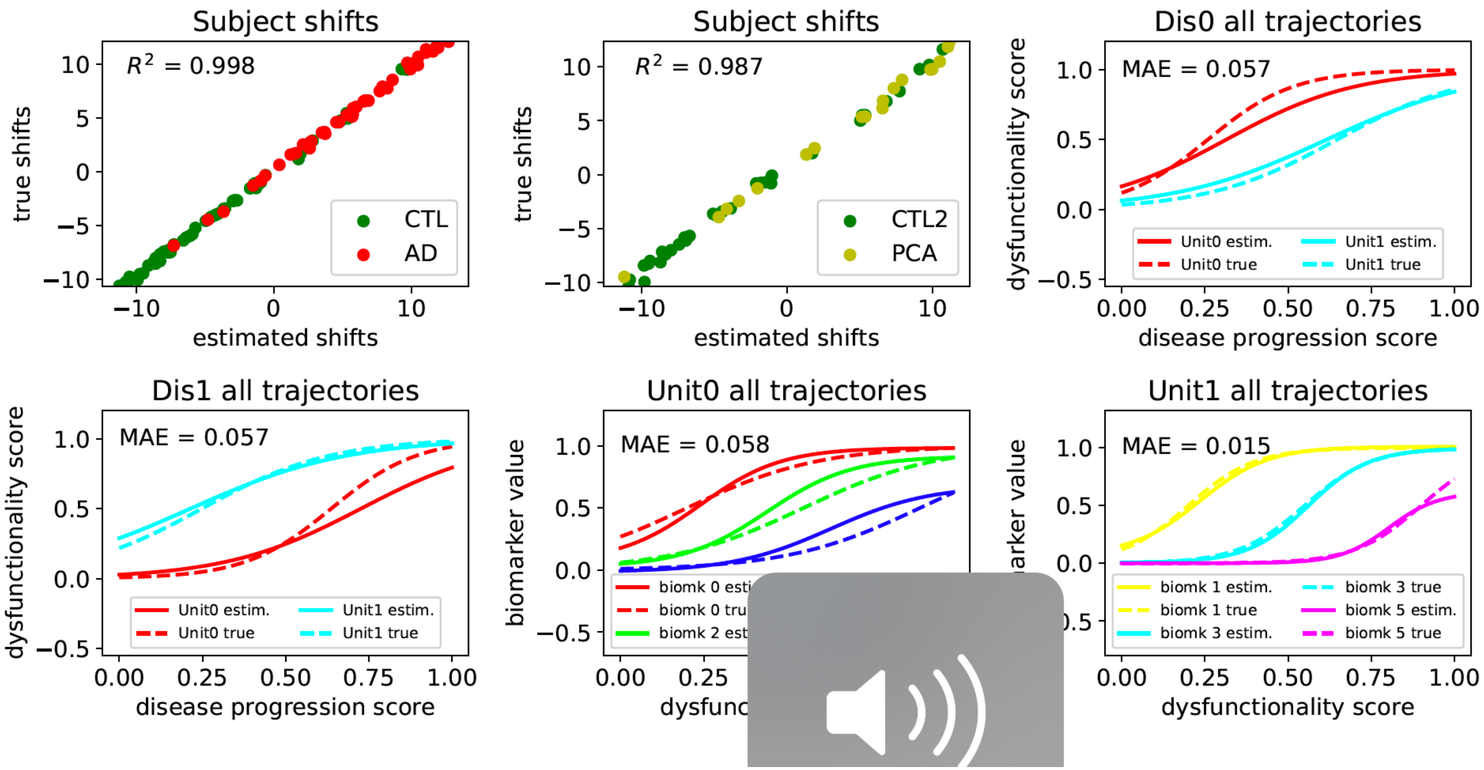


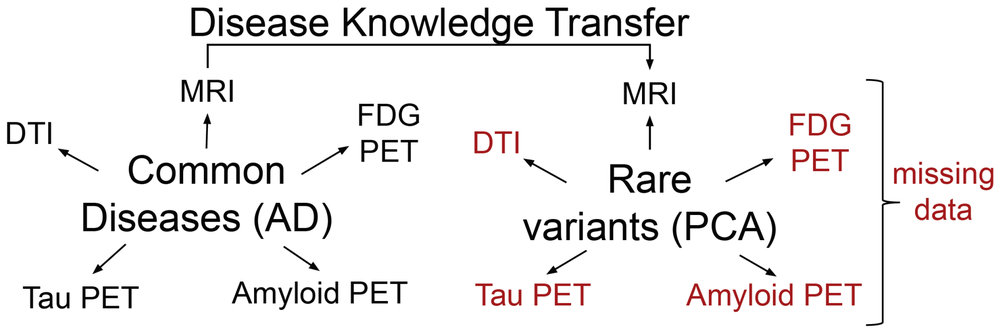


Marinescu et al., MICCAI, 2019

DKT works well on simulated data

- Simulated 100 subjects with two diseases: synAD & synPCA
- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers
- DKT was able to:
  - Reliably fit the data
  - Infer the missing biomarkers in synPCA

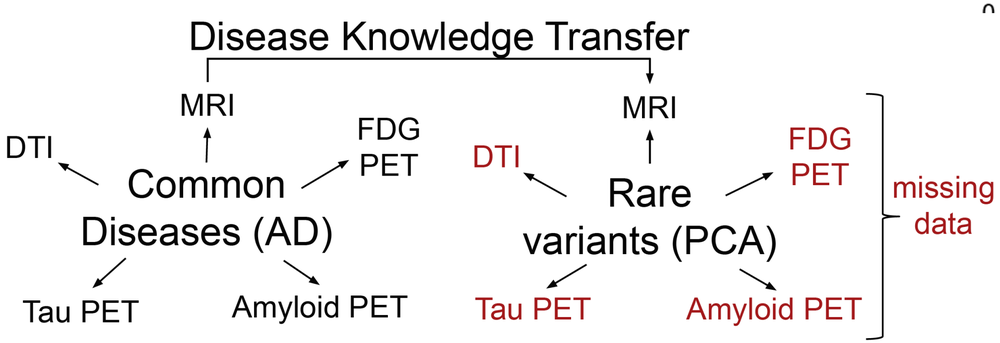


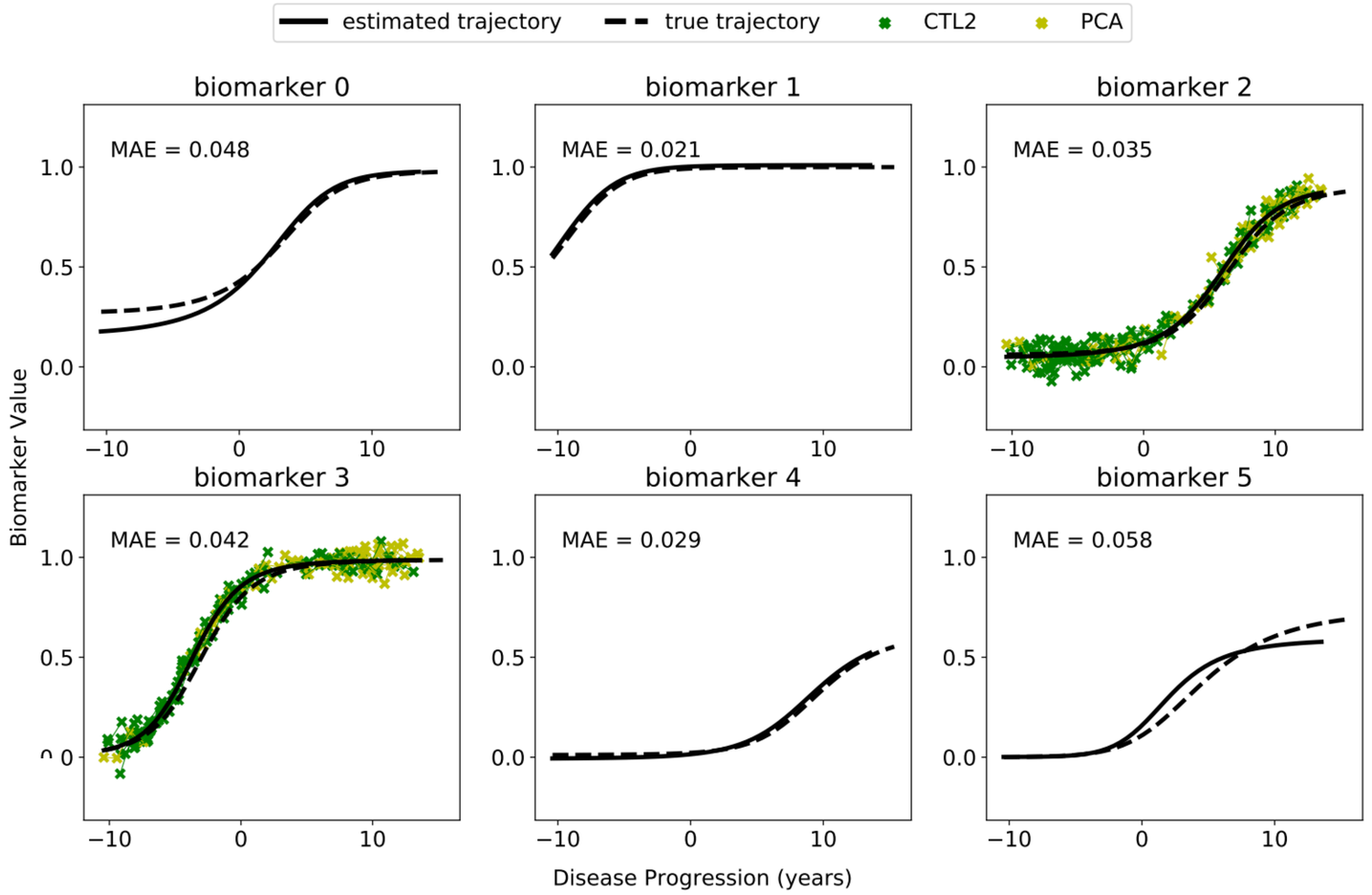


Marinescu et al., MICCAI, 2019

DKT works well on simulated data

- Simulated 100 subjects with two diseases: synAD & synPCA
- To simulate lack of multimodal data in synPCA, we discarded 4/6 biomarkers
- DKT was able to:
  - Reliably fit the data
  - Infer the missing biomarkers in synPCA

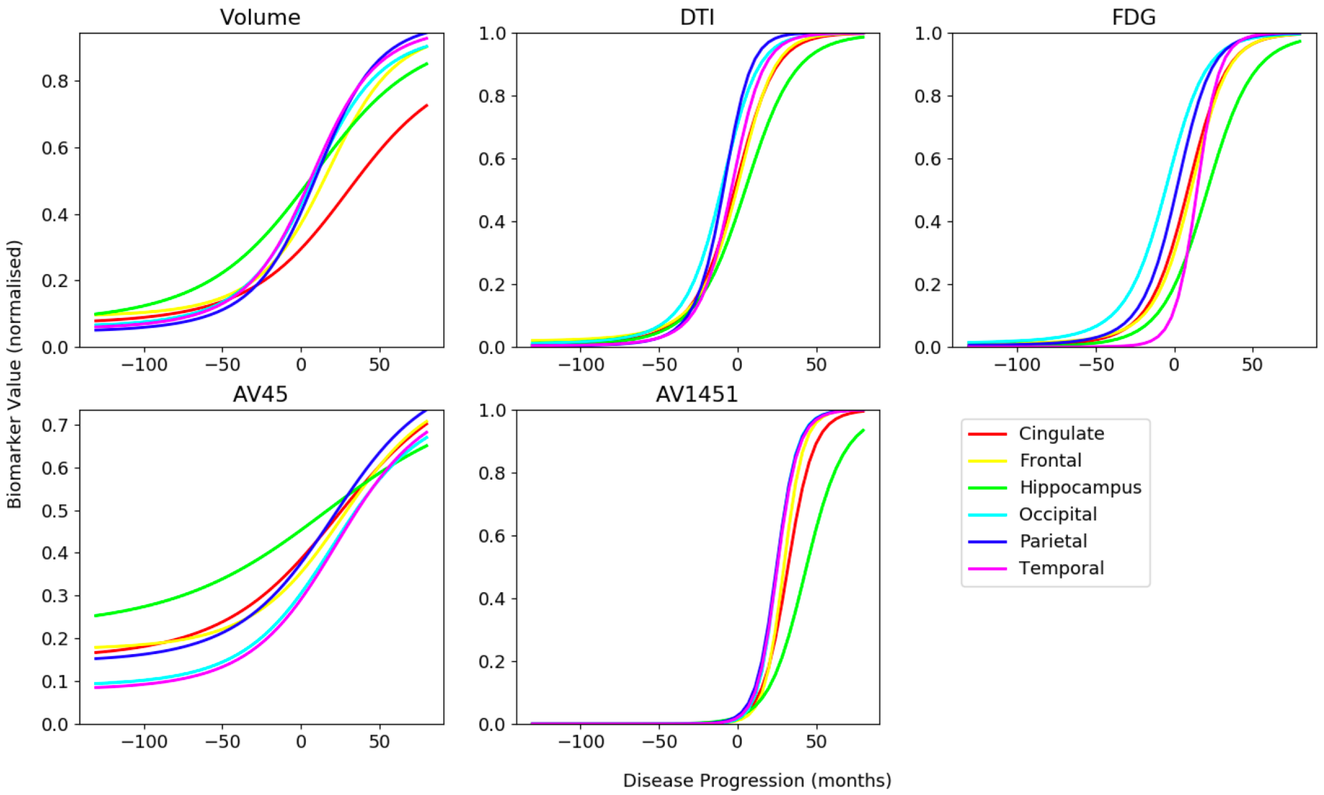




Marinescu et al., MICCAI, 2019

On real data, DKT can estimate *multimodal* trajectories of Posterior Cortical Atrophy only using structural MRI

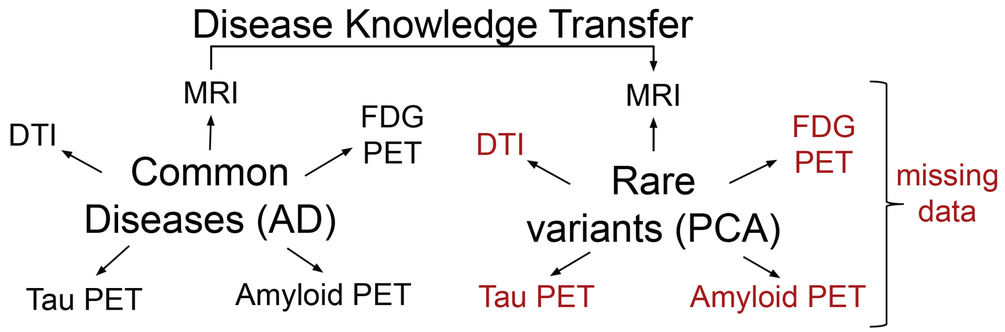
- Ran on 76 PCA subjects from the Dementia Research Center UK
- Given structural MRI, DKT was able to infer missing DTI, FDG, Tau PET and Amyloid PET in PCA, in lack of such data.
  - We subsequently validate the DTI trajectories
- The first such longitudinal trajectories of **multimodal** biomarkers in Posterior Cortical Atrophy

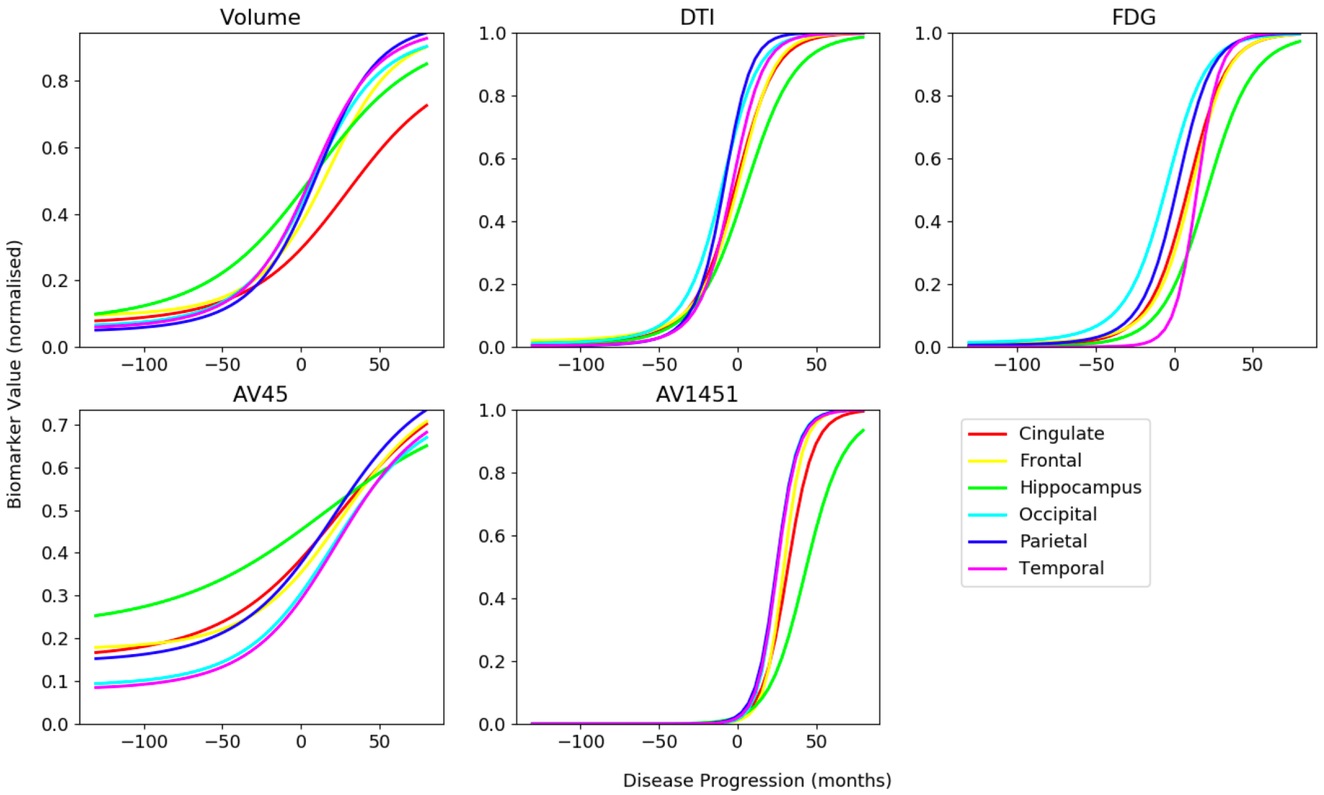


Marinescu et al., MICCAI, 2019

On real data, DKT can estimate *multimodal* trajectories of Posterior Cortical Atrophy only using structural MRI

- Ran on 76 PCA subjects from the Dementia Research Center UK
- Given structural MRI, DKT was able to infer missing DTI, FDG, Tau PET and Amyloid PET in PCA, in lack of such data.
  - We subsequently validate the DTI trajectories
- The first such longitudinal trajectories of **multimodal** biomarkers in Posterior Cortical Atrophy

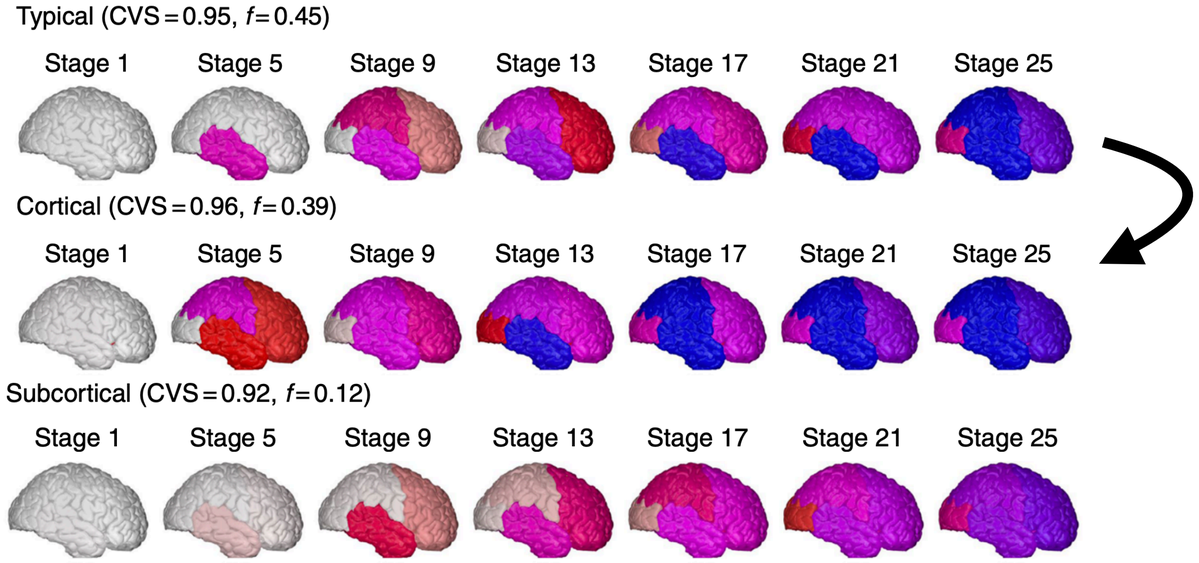




Marinescu et al., MICCAI, 2019

Quantitative evaluation & validation through transfer learning

- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)

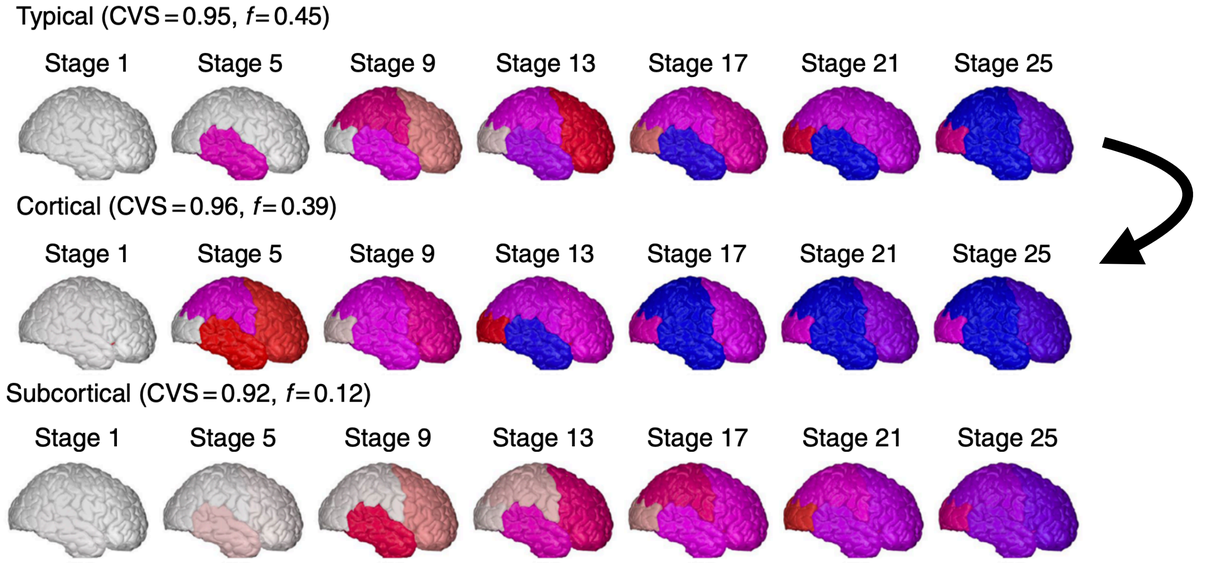


Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*

Quantitative evaluation & validation through transfer learning

- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)

- Transferred information from Cortical to Hippocampal subgroups

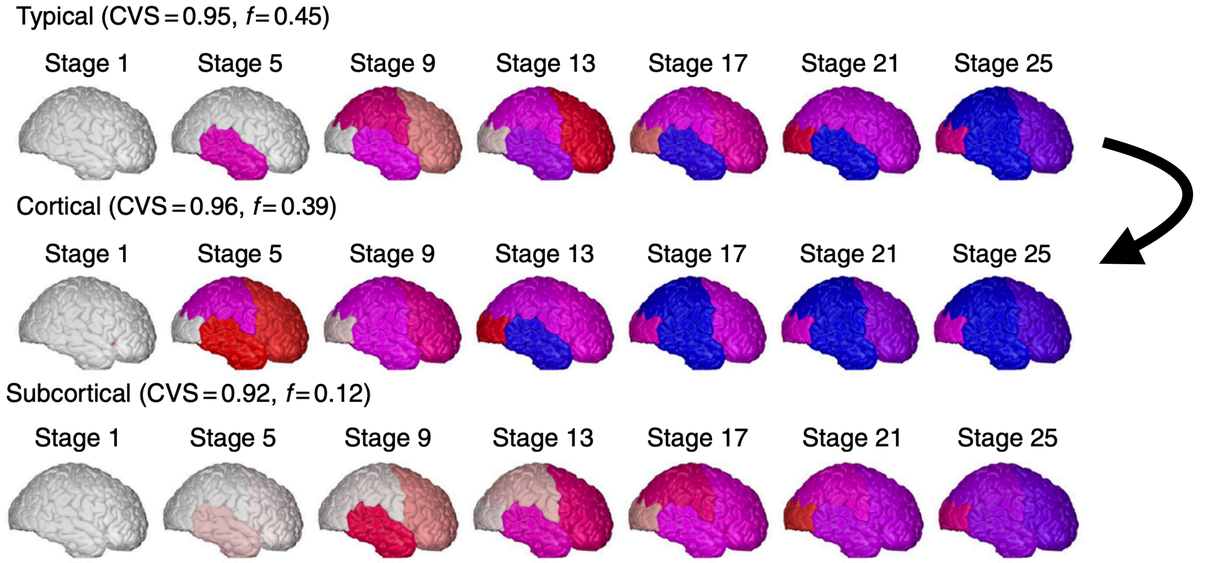


Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*

Quantitative evaluation & validation through transfer learning

- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)

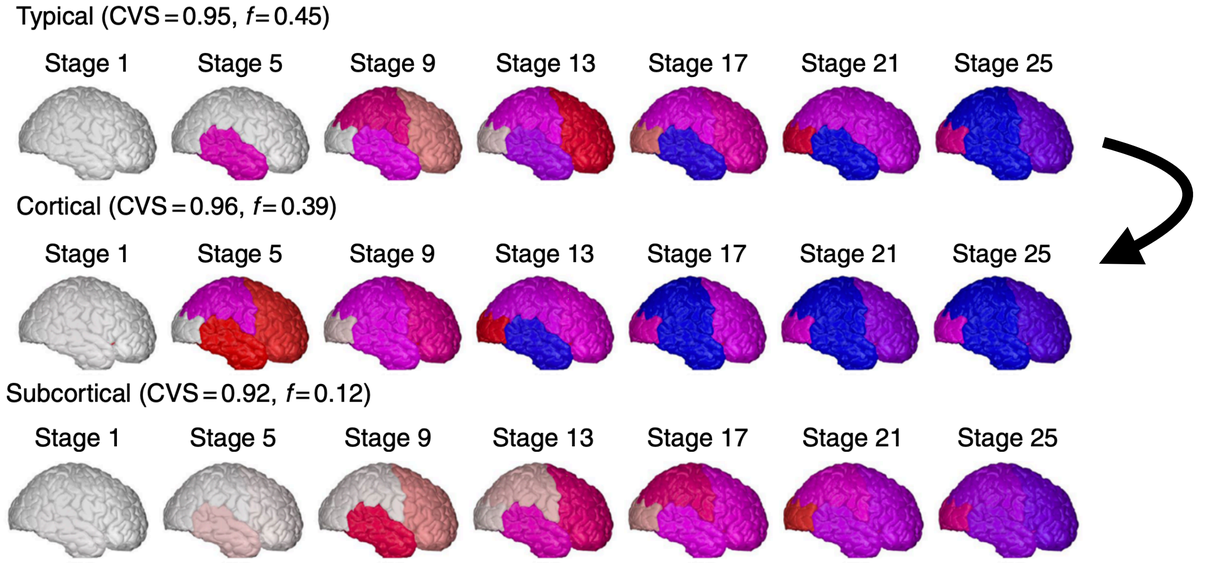
- Transferred information from Cortical to Hippocampal subgroups



Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*
	typical Alzheimer's to Posterior Cortical Atrophy					
DKT (ours)	0.77 ± 0.11	0.39 ± 0.26	0.75 ± 0.09	0.60 ± 0.14	**0.55 ± 0.24**	**0.35 ± 0.22**
Latent stage	**0.80 ± 0.09**	**0.53 ± 0.17**	**0.80 ± 0.12**	0.56 ± 0.18	0.50 ± 0.21	0.32 ± 0.24
Multivariate	0.73 ± 0.09	0.45 ± 0.22	0.71 ± 0.08	-0.28 ± 0.21*	0.53 ± 0.22	0.25 ± 0.23*
Spline	0.52 ± 0.20*	-0.03 ± 0.35*	0.66 ± 0.11*	0.09 ± 0.25*	0.53 ± 0.20	0.30 ± 0.21*
Linear	0.52 ± 0.20*	0.34 ± 0.27	0.66 ± 0.11*	**0.64 ± 0.17**	0.54 ± 0.22	0.30 ± 0.21*

Quantitative evaluation & validation through transfer learning

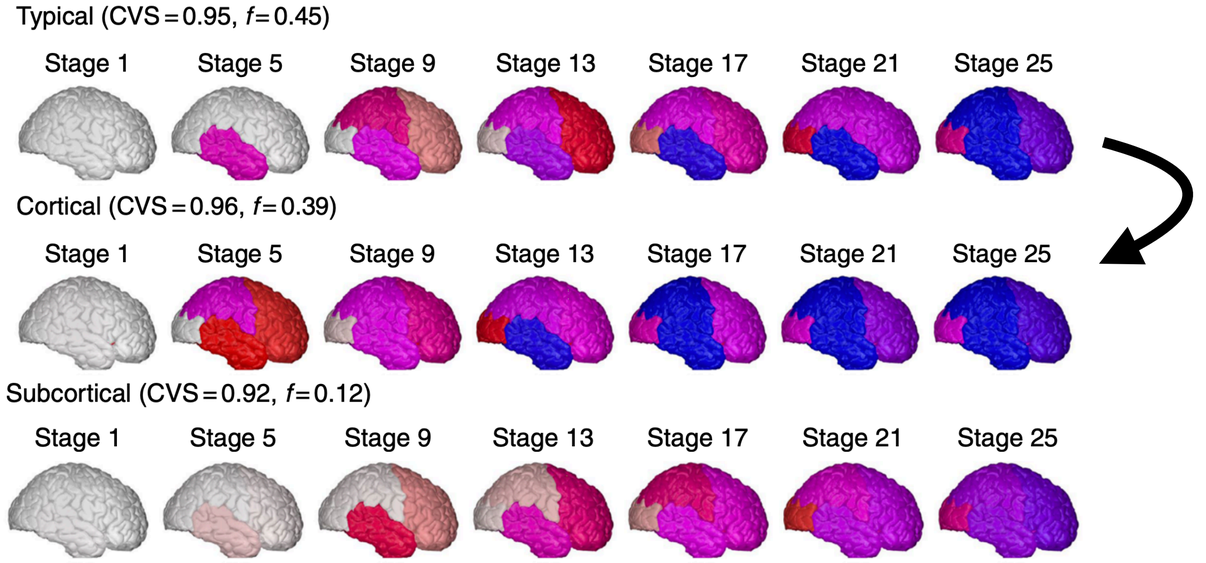
- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)
- Transferred information from Cortical to Hippocampal subgroups
- From typical AD to Posterior Cortical Atrophy



Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*
	typical Alzheimer's to Posterior Cortical Atrophy					
DKT (ours)	0.77 ± 0.11	0.39 ± 0.26	0.75 ± 0.09	0.60 ± 0.14	**0.55 ± 0.24**	**0.35 ± 0.22**
Latent stage	**0.80 ± 0.09**	**0.53 ± 0.17**	**0.80 ± 0.12**	0.56 ± 0.18	0.50 ± 0.21	0.32 ± 0.24
Multivariate	0.73 ± 0.09	0.45 ± 0.22	0.71 ± 0.08	-0.28 ± 0.21*	0.53 ± 0.22	0.25 ± 0.23*
Spline	0.52 ± 0.20*	-0.03 ± 0.35*	0.66 ± 0.11*	0.09 ± 0.25*	0.53 ± 0.20	0.30 ± 0.21*
Linear	0.52 ± 0.20*	0.34 ± 0.27	0.66 ± 0.11*	**0.64 ± 0.17**	0.54 ± 0.22	0.30 ± 0.21*

Quantitative evaluation & validation through transfer learning

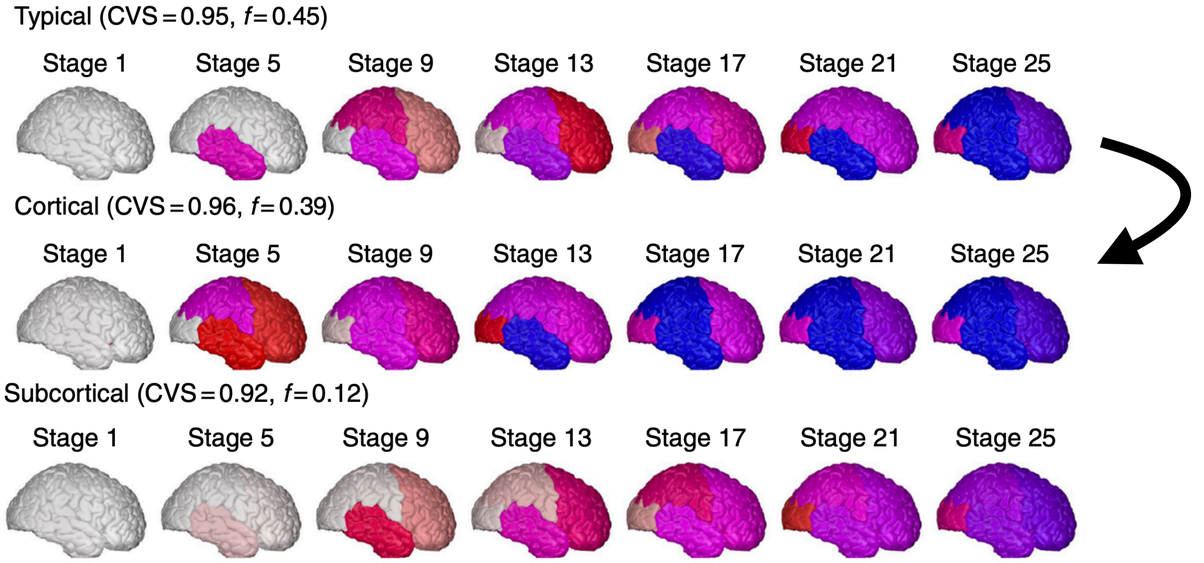
- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)
- Transferred information from Cortical to Hippocampal subgroups
- From typical AD to Posterior Cortical Atrophy
- Validated on 20 left-out diffusion scans on PCA



Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*
	typical Alzheimer's to Posterior Cortical Atrophy					
DKT (ours)	0.77 ± 0.11	0.39 ± 0.26	0.75 ± 0.09	0.60 ± 0.14	**0.55 ± 0.24**	**0.35 ± 0.22**
Latent stage	**0.80 ± 0.09**	**0.53 ± 0.17**	**0.80 ± 0.12**	0.56 ± 0.18	0.50 ± 0.21	0.32 ± 0.24
Multivariate	0.73 ± 0.09	0.45 ± 0.22	0.71 ± 0.08	-0.28 ± 0.21*	0.53 ± 0.22	0.25 ± 0.23*
Spline	0.52 ± 0.20*	-0.03 ± 0.35*	0.66 ± 0.11*	0.09 ± 0.25*	0.53 ± 0.20	0.30 ± 0.21*
Linear	0.52 ± 0.20*	0.34 ± 0.27	0.66 ± 0.11*	**0.64 ± 0.17**	0.54 ± 0.22	0.30 ± 0.21*

Quantitative evaluation & validation through transfer learning

- Split ADNI into three different subgroups with different disease progressions (using SuStaIn)
- Transferred information from Cortical to Hippocampal subgroups
- From typical AD to Posterior Cortical Atrophy
- Validated on 20 left-out diffusion scans on PCA
  - Fractional anisotropy maps



Model	Cingulate	Frontal	Hippocam.	Occipital	Parietal	Temporal
	TADPOLE: Hippocampal subgroup to Cortical subgroup					
DKT (ours)	0.56 ± 0.23	**0.35 ± 0.17**	**0.58 ± 0.14**	-0.10 ± 0.29	**0.71 ± 0.11**	**0.34 ± 0.26**
Latent stage	0.44 ± 0.25	0.34 ± 0.21	0.34 ± 0.24*	**-0.07 ± 0.22**	0.64 ± 0.16	0.08 ± 0.24*
Multivariate	**0.60 ± 0.18**	0.11 ± 0.22*	0.12 ± 0.29*	-0.22 ± 0.22	-0.44 ± 0.14*	-0.32 ± 0.29*
Spline	-0.24 ± 0.25*	-0.06 ± 0.27*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.10 ± 0.25*
Linear	-0.24 ± 0.25*	0.20 ± 0.25*	0.58 ± 0.17	-0.16 ± 0.27	0.23 ± 0.25*	0.13 ± 0.23*
	typical Alzheimer's to Posterior Cortical Atrophy					
DKT (ours)	0.77 ± 0.11	0.39 ± 0.26	0.75 ± 0.09	0.60 ± 0.14	**0.55 ± 0.24**	**0.35 ± 0.22**
Latent stage	**0.80 ± 0.09**	**0.53 ± 0.17**	**0.80 ± 0.12**	0.56 ± 0.18	0.50 ± 0.21	0.32 ± 0.24
Multivariate	0.73 ± 0.09	0.45 ± 0.22	0.71 ± 0.08	-0.28 ± 0.21*	0.53 ± 0.22	0.25 ± 0.23*
Spline	0.52 ± 0.20*	-0.03 ± 0.35*	0.66 ± 0.11*	0.09 ± 0.25*	0.53 ± 0.20	0.30 ± 0.21*
Linear	0.52 ± 0.20*	0.34 ± 0.27	0.66 ± 0.11*	**0.64 ± 0.17**	0.54 ± 0.22	0.30 ± 0.21*

Summary and future work

- Proposed a model to perform transfer learning across different neurodegenerative diseases
- Transfer learning is done through sharing the underpinning disease mechanisms
- Model evaluated and validated in simulations as well as real data (ADNI & Dementia Research Center UK) on the largest PCA cohort to date
- Future work: transfer learning using deep-learning approaches, by synthesizing PET/DTI/CT scans for rarer neurodegenerative diseases where such data is very limited
- Such synthesis will enable characterizing their progression, which can help identify novel drug targets, stratify cohorts for clinical trials and identify suitable endpoints.