Maladies neurodégénératives et IA: vers une approche intégrative

Marco Lorenzi
Université Côte d'Azur, Inria Sophia Antipolis
Epione Research Group

William Utermolhen (1933-2007)

Self-portraits

1967

1996

1997

1998

1999

2000

1995: diagnosis of Alzheimer’s disease

Alzheimer's disease and Neurodegenerative disorders

Language problems

Memory loss

Functionality loss

Apraxia

Memory impairment

Mood alterations

Human and social costs

Most costly disease in Europe and USA

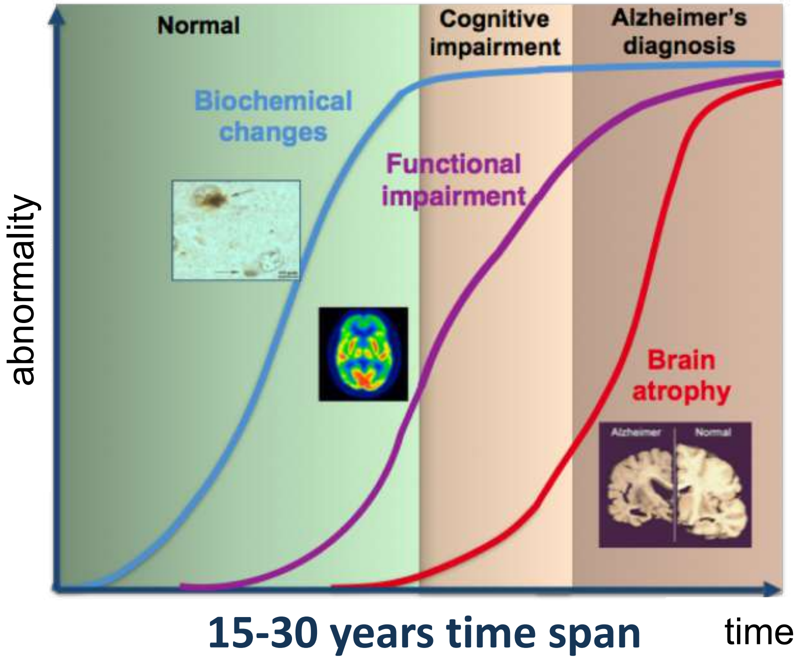
Long-term care
Median life expectancy of 9 years
[Castro et al., Dement Neuropsychol. 2010]

Impact on caregivers
66-75% of average family income
[Castro et al., Dement Neuropsychol. 201]

Global Health care
50 million people affected in 2018
1 trillion $ cost worldwide in 2018
Source World Alzheimer Report 2018 (alz.co.uk)

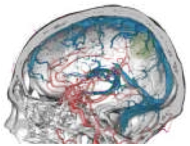
Dynamics of neurodegeneration

Jack et al, Lancet Neurol 2010;
Frisoni et al, Nature Rev Neurol 2010



15-30 years time span

Genetics



Vascularity

?

Sociodemographic

Microbiome

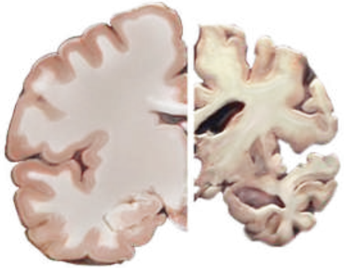
...

Data Science meets biomedical research

Statistical learning

+

Neurology



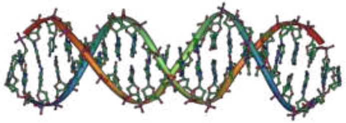
Statistical learning
- formalizes hypothesis into models
- verifies models by means of data

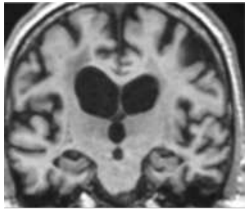
A discipline to answer challenging questions in medicine

Data Science meets biomedical research

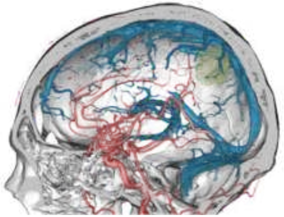
- Better understanding of the pathology

How to integrate heterogenous biomedical measures?





baseline

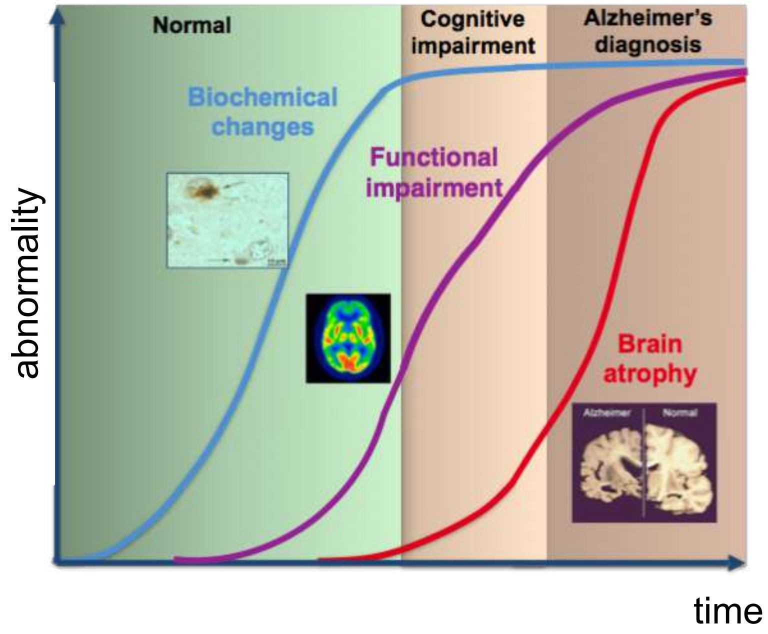


…

Data Science meets biomedical research

- Early diagnosis for better cure

How to predict and interpret a pathological evolution?



Challenges

- How to integrate heterogeneous biomedical measures?

- How to integrate the temporal dimension?

- How to unveil the biological mechanisms of the pathology?

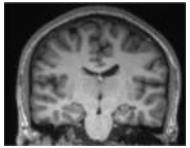
Challenges

- How to integrate heterogeneous biomedical measures?

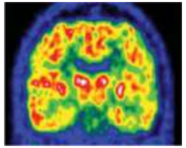
- How to integrate the temporal dimension?

- How to unveil the biological mechanisms of the pathology?

Generative representation of the disease



X_1

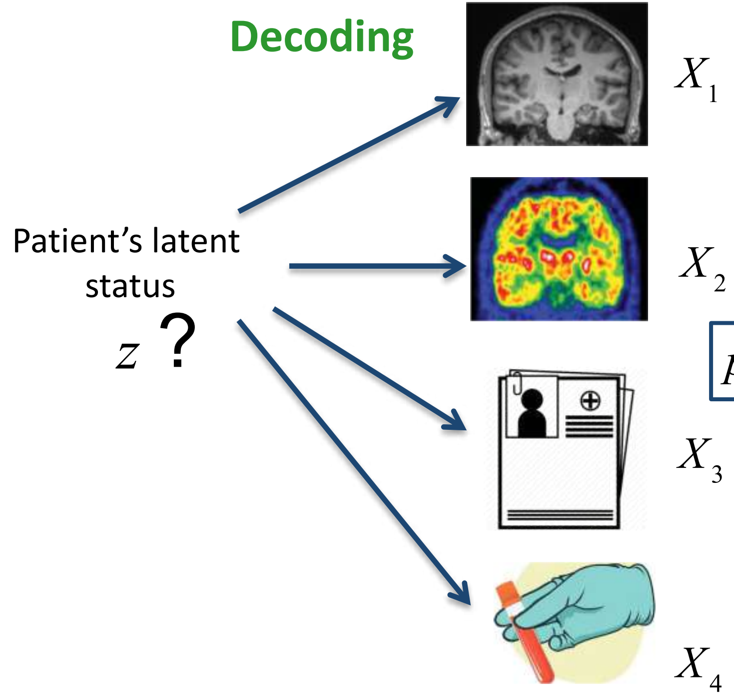


X_2

X_3

X_4

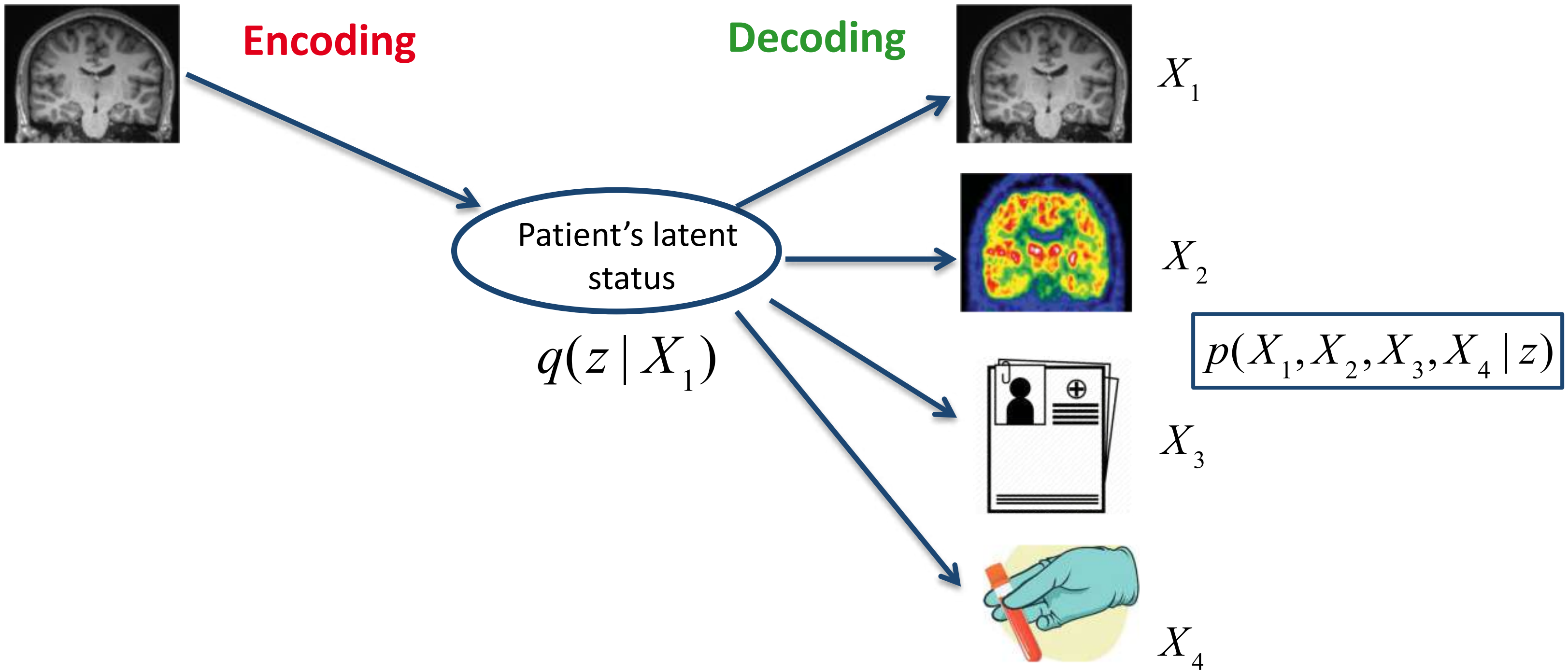
Generative representation of the disease



$$\boxed{p(X_1, X_2, X_3, X_4 \mid z)}$$

Decoding: data reconstruction from the latent representation

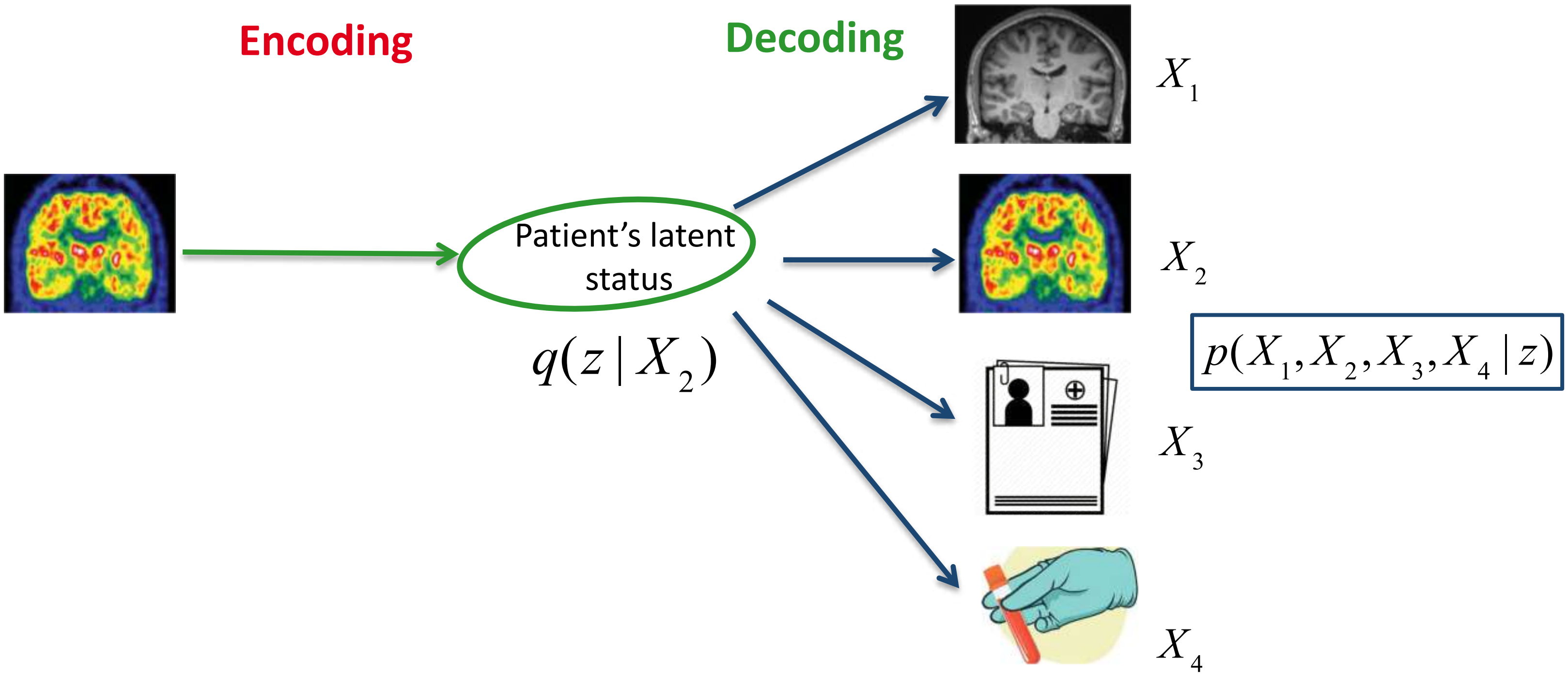
Generative representation of the disease



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

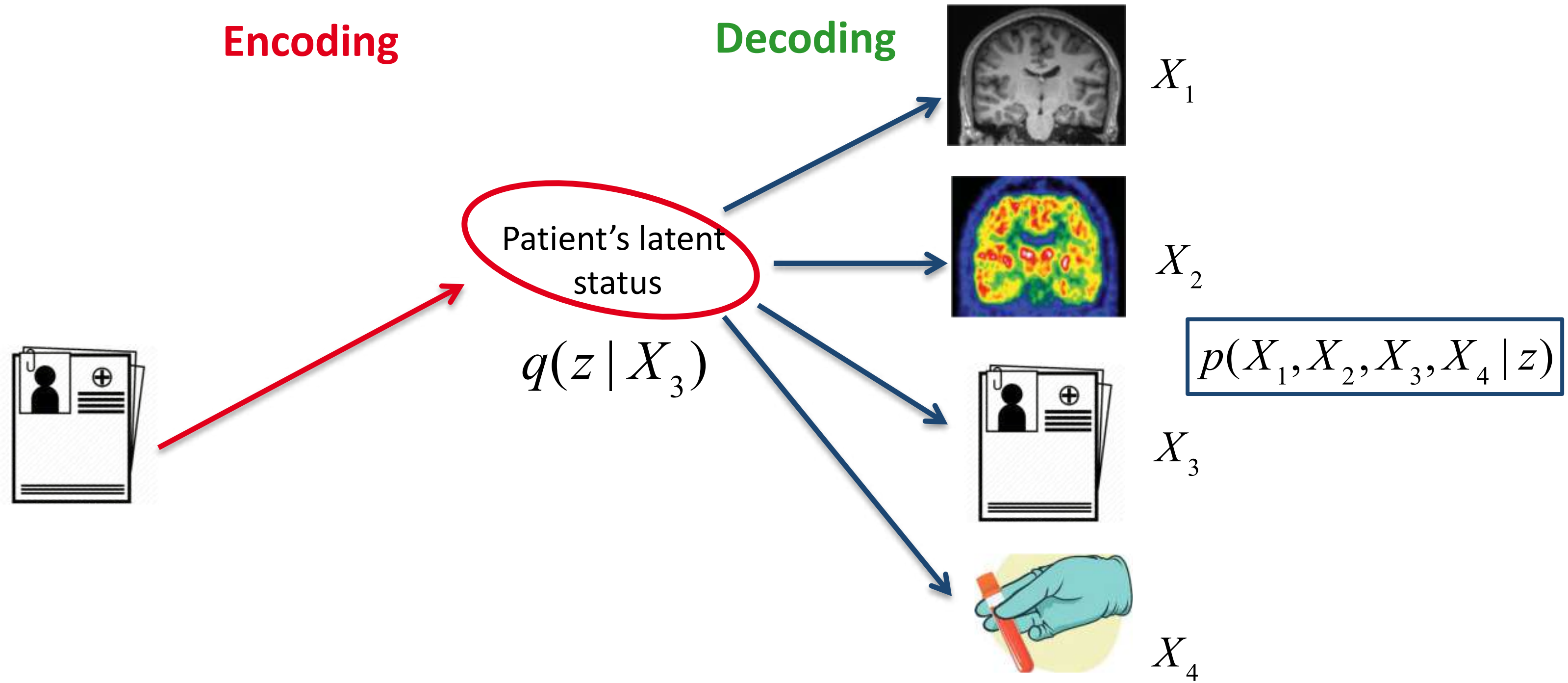
Generative representation of the disease



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

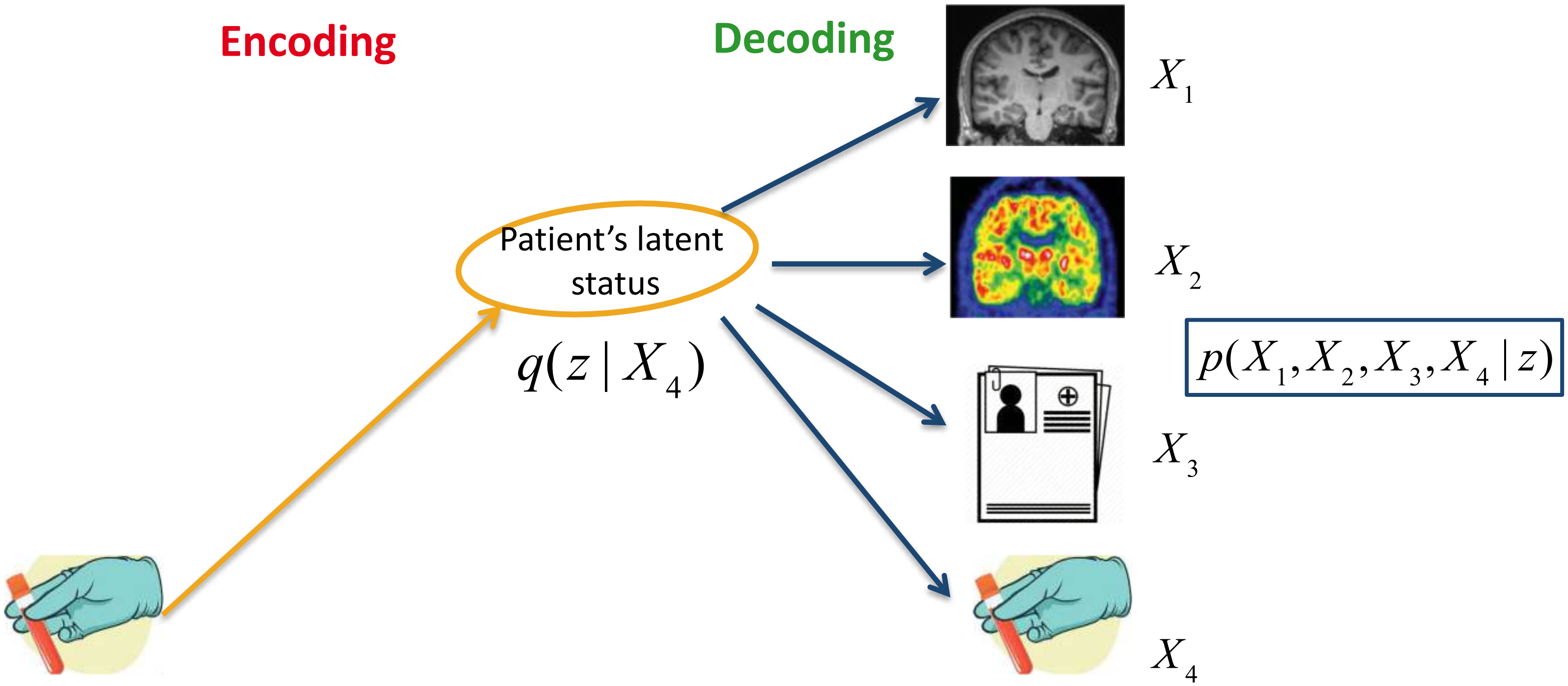
Generative representation of the disease



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

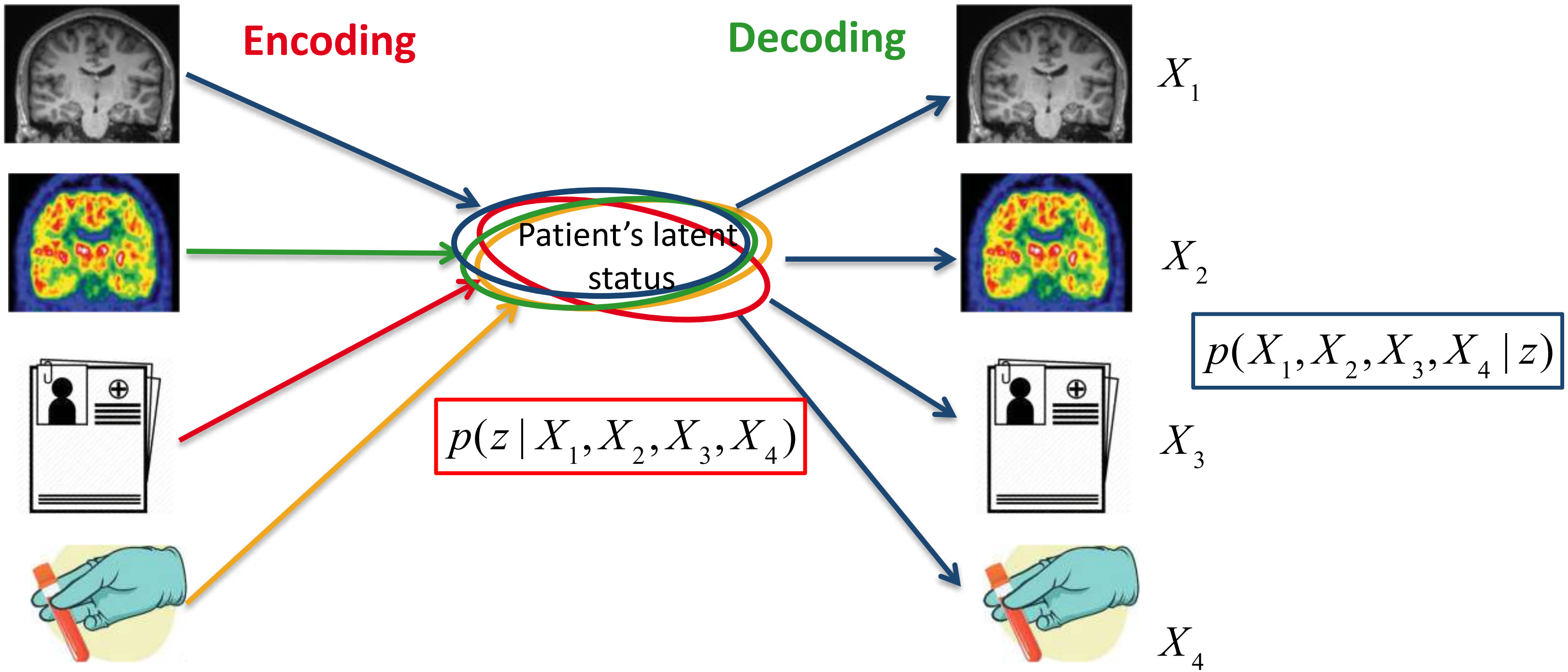
Generative representation of the disease



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

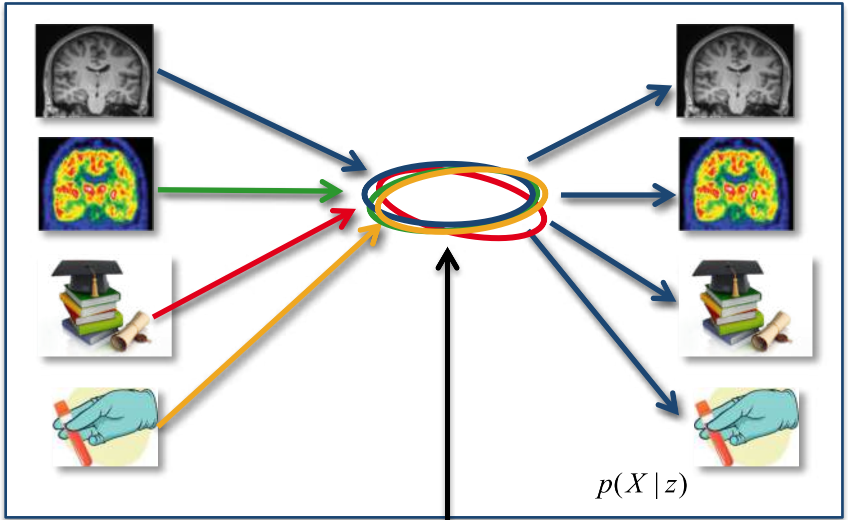
Generative representation of the disease



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

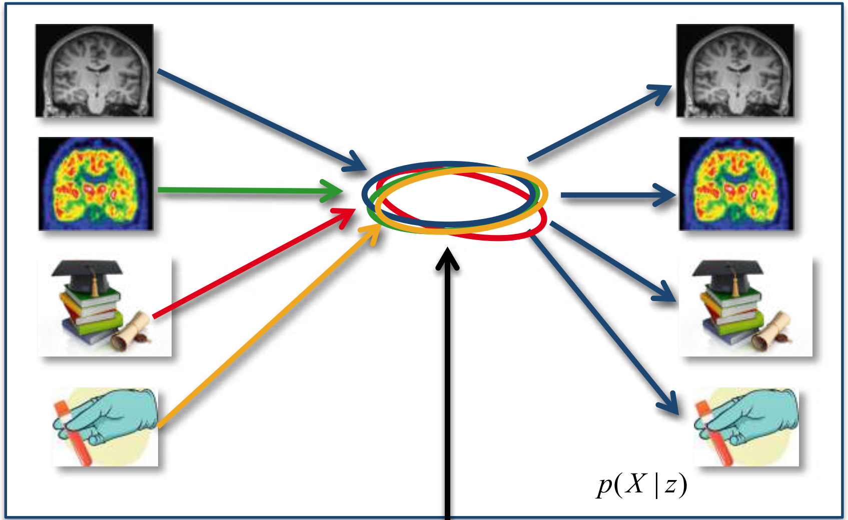
Generative representation of the disease



minimize

$$dist\Big(q(z \mid X_c), p(z \mid X_1, X_2, \ldots, X_c)\Big)$$

Generative representation of the disease



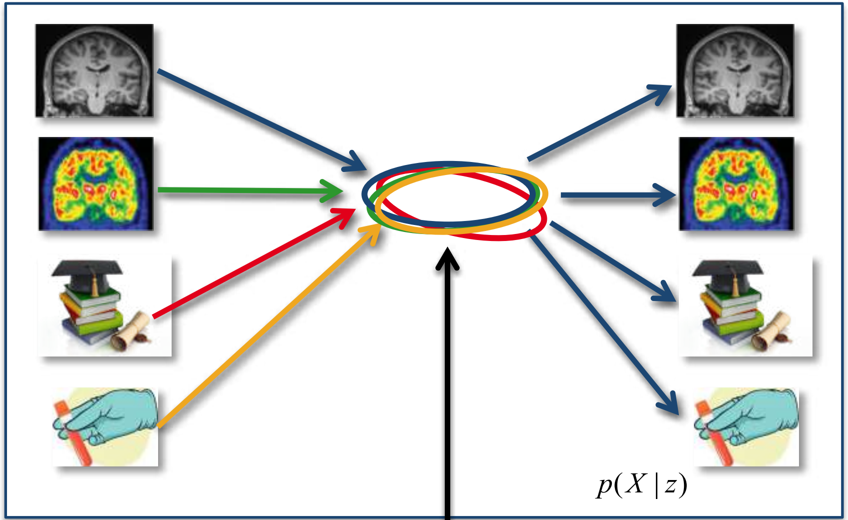
minimize

$$dist\left(q(z \mid X_c), p(z \mid X_1, X_2, \ldots, X_c)\right)$$

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C} E_{q(z|X_c)}[\sum_i \ln p(X_i \mid z)] - DKL\left(q(z \mid X_c) \,||\, p(z)\right)$$

Generative representation of the disease



minimize

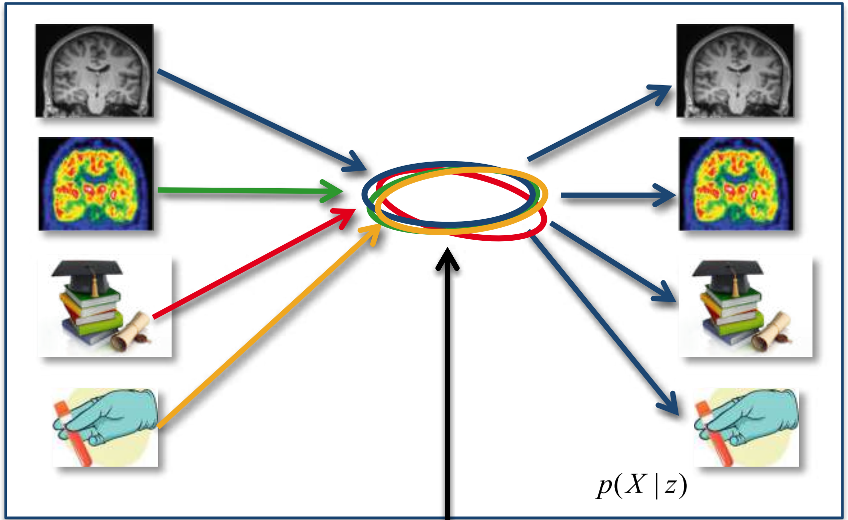
$$dist\left(q(z \mid X_c), p(z \mid X_1, X_2, \ldots, X_c)\right)$$

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C} E_{q(z|X_c)}[\sum_i \ln p(X_i \mid z)] - DKL\left(q(z \mid X_c) \,\|\, p(z)\right)$$

Encoding for given channel

Generative representation of the disease



minimize

$$dist\left(q(z \mid X_c), p(z \mid X_1, X_2, \ldots, X_c)\right)$$

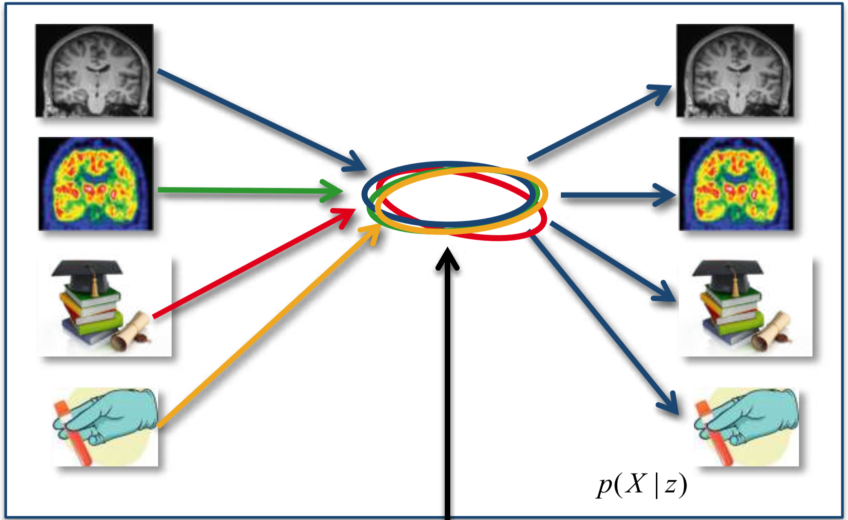
Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C} E_{q(z|X_c)}\left[\sum_i \ln p(X_i \mid z)\right] - DKL\left(q(z \mid X_c) \,\|\, p(z)\right)$$

Encoding for given channel

Reconstruction of all channels

Generative representation of the disease



minimize

$$dist\left(q(z \mid X_c), p(z \mid X_1, X_2, \ldots, X_c)\right)$$

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C} E_{q(z|X_c)}\left[\sum_i \ln p(X_i \mid z)\right] - DKL\left(q(z \mid X_c) \,\|\, p(z)\right)$$

Encoding for given channel

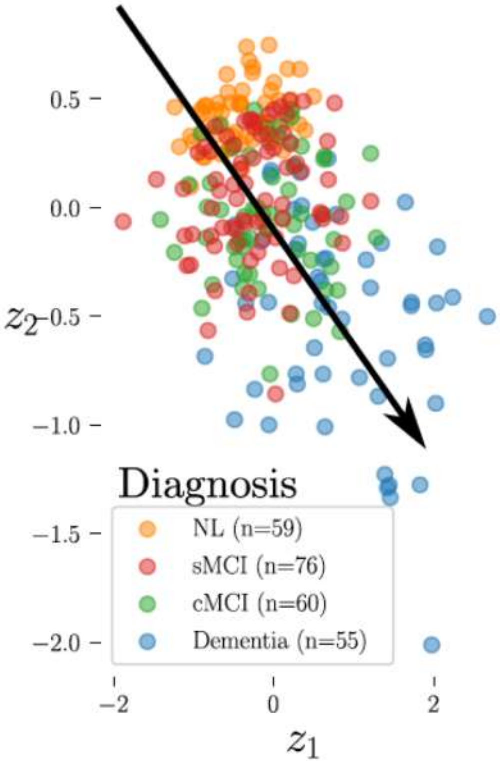
Reconstruction of all channels

Regularization: sparsity inducing prior

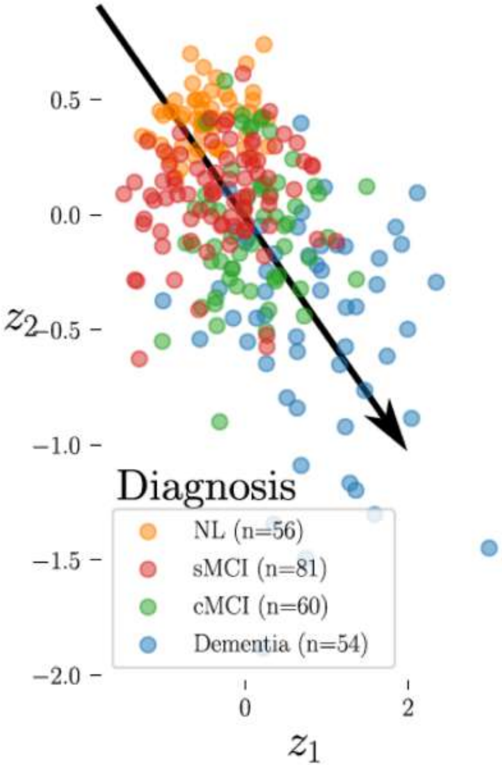
[Kingma et al, NIPS, 2015; Molchanov et al, ICML 2017]

Prediction from latent space

training



testing

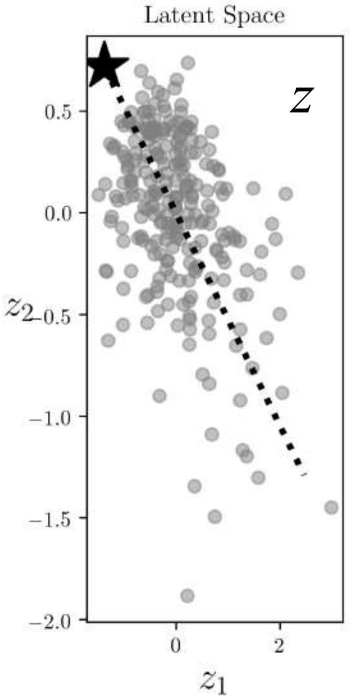


Joint modeling of

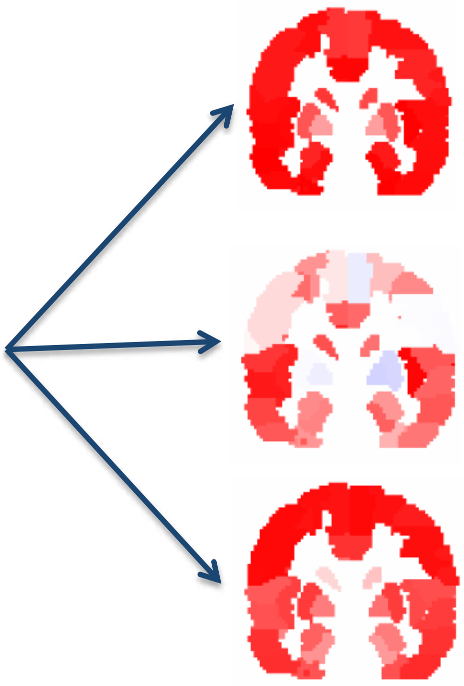
- Brain imaging:
  - Structural (T1 MRI)
  - Molecular (FDG-PET + Amy-PET)
- Socio-demographic factors
- Clinical scores

	Accuracy (SD)
Cognitively Healthy	**0.89** (0.03)
Stable Mild Cognitive Impairment (sMCI)	**0.75** (0.02)
MCI converting to Dementia (cMCI)	**0.70** (0.05)
Dementia	**0.94** (0.05)

Generation from latent space



Patient's latent status



Atrophy abnormality (MRI)

Metabolism abnormality (FDG)

Amyloid abnormality (AV45)

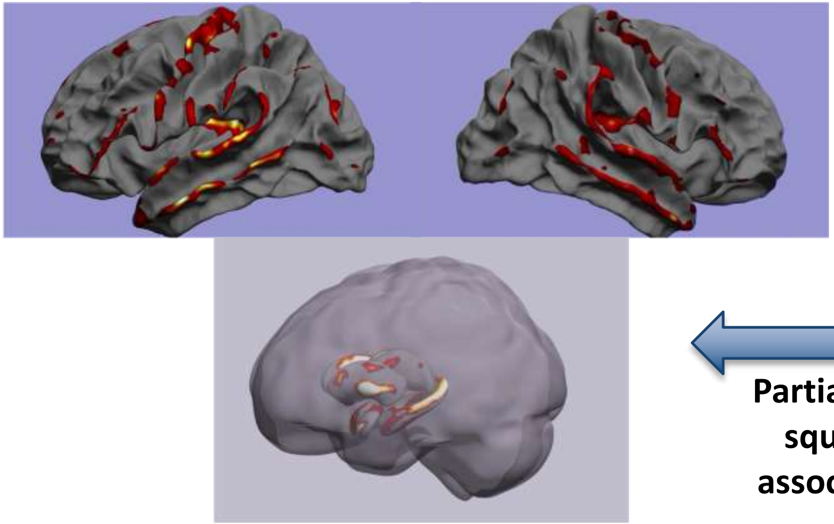
healthy

pathological

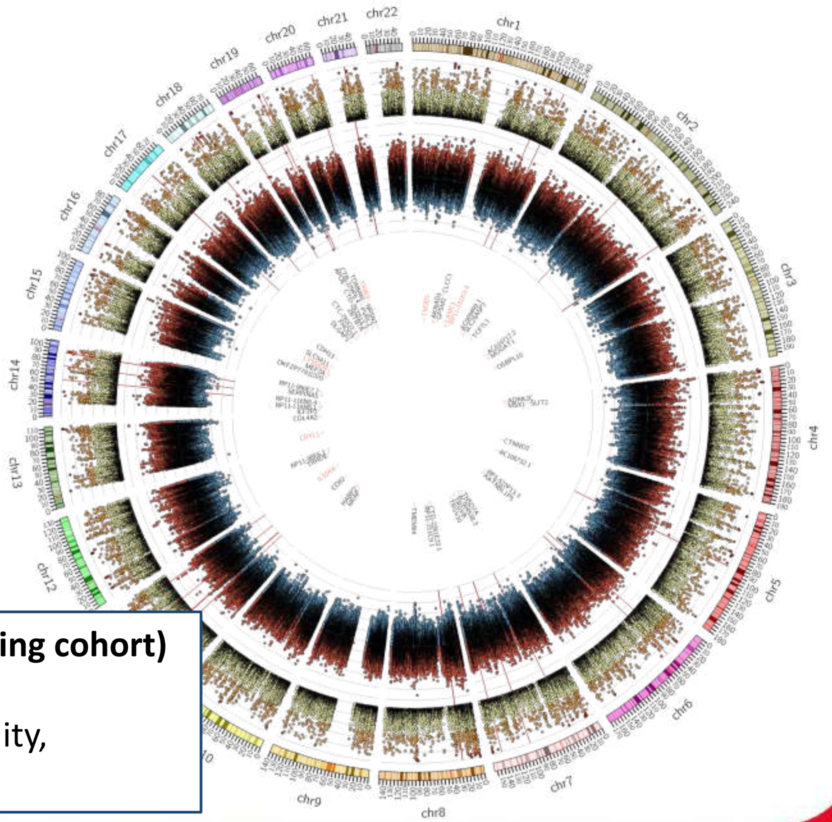
- Improved interpretability
- Multi-channel: working with missing data/data imputation
- Simulations for clinical trials

Another illustration: multivariate Imaging-genetics

Atrophy profile from brain imaging



Genome



Partial least squares association

639 individuals
401 healthy
238 Alzheimer's

TRIB3 gene (significant on testing cohort)
- neuronal cell death,
- modulation of PSEN1 stability,
- interaction with APP.

Integrating heterogeneous biomedical measures

- Latent variable models provide powerful discovery tools

- Specific properties needed: scalability, flexibility, interpretability

- Curse of dimensionality and generalization:
  Synergy between data scientists, biologists and clinicians

Challenges

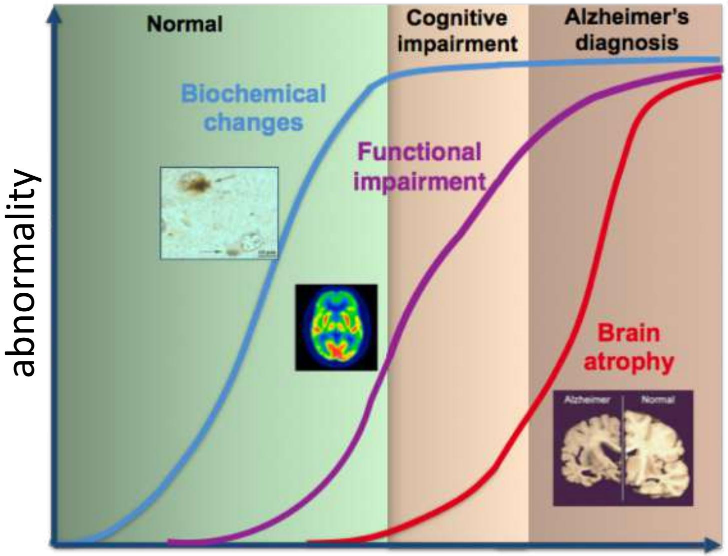
- How to integrate heterogeneous biomedical measures?

- How to integrate the temporal dimension?

- How to unveil the biological mechanisms of the pathology?

Modeling the natural history of neurodegeneration

Hypothetical model



Disease time (15-30 years)

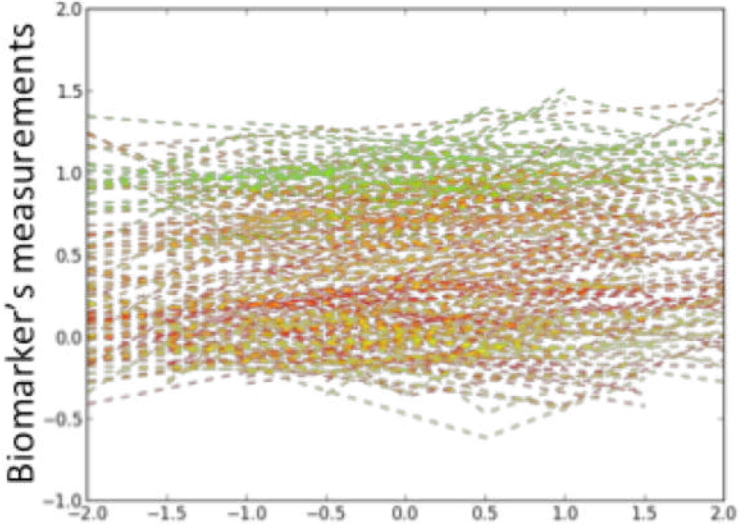
Large time span

?

Assessing

- Patient severity
- Drugs effect

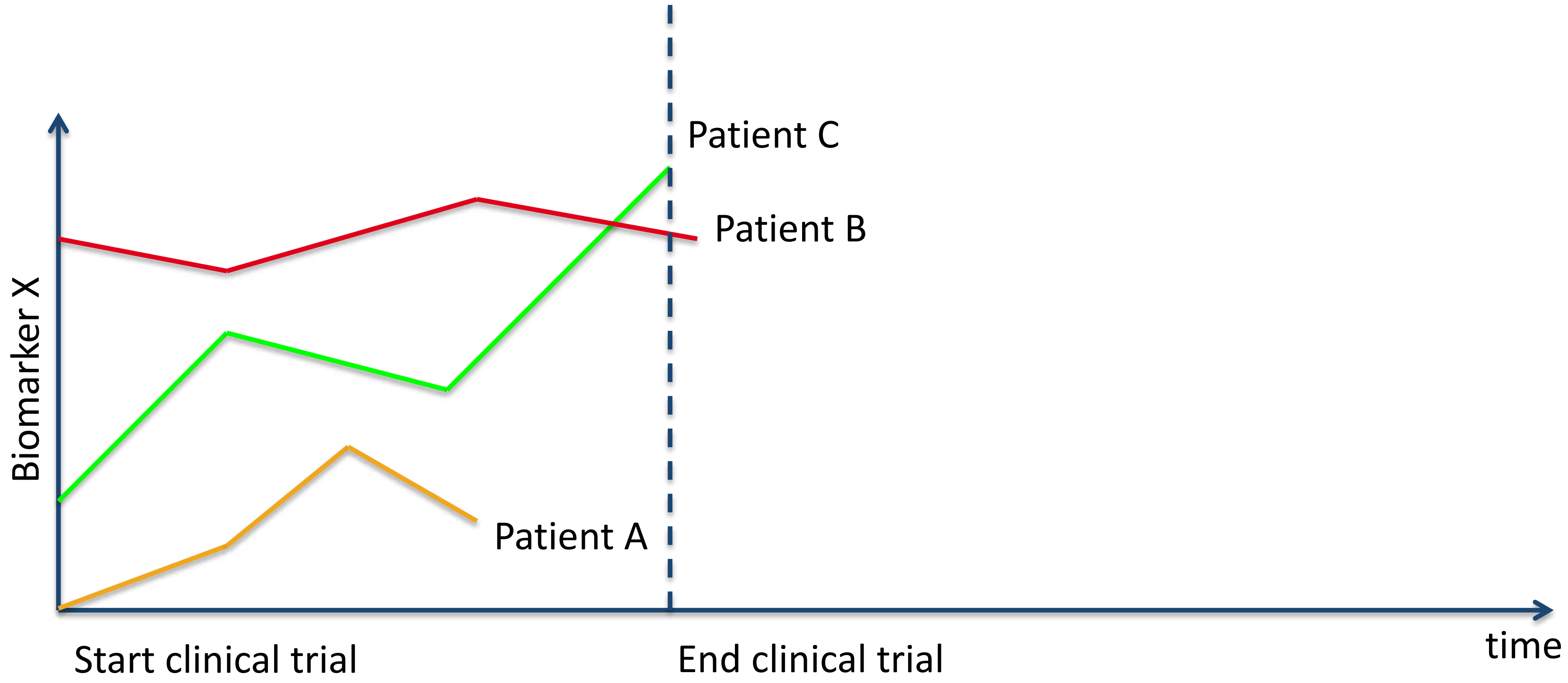
The reality



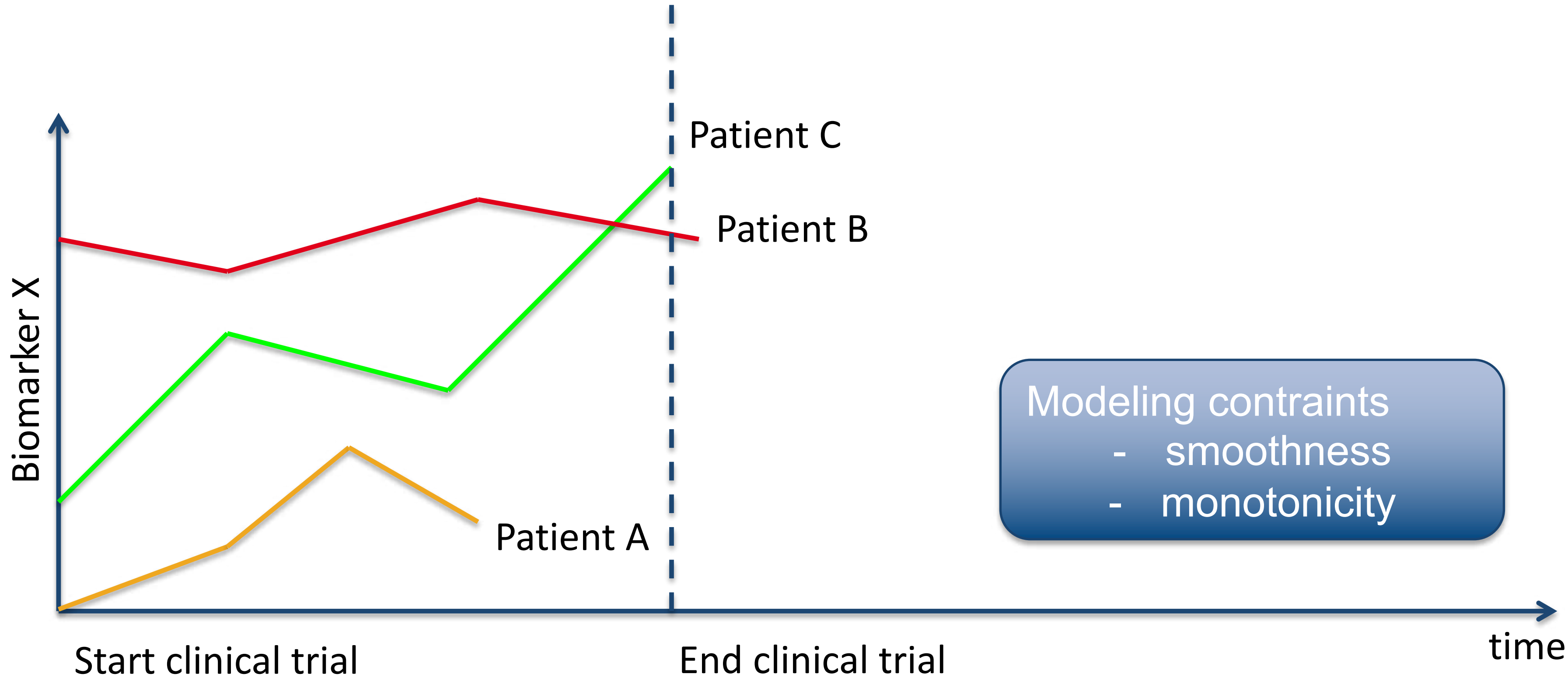
Clinical trial time (few years)

Short time span

Statistical disease progression model

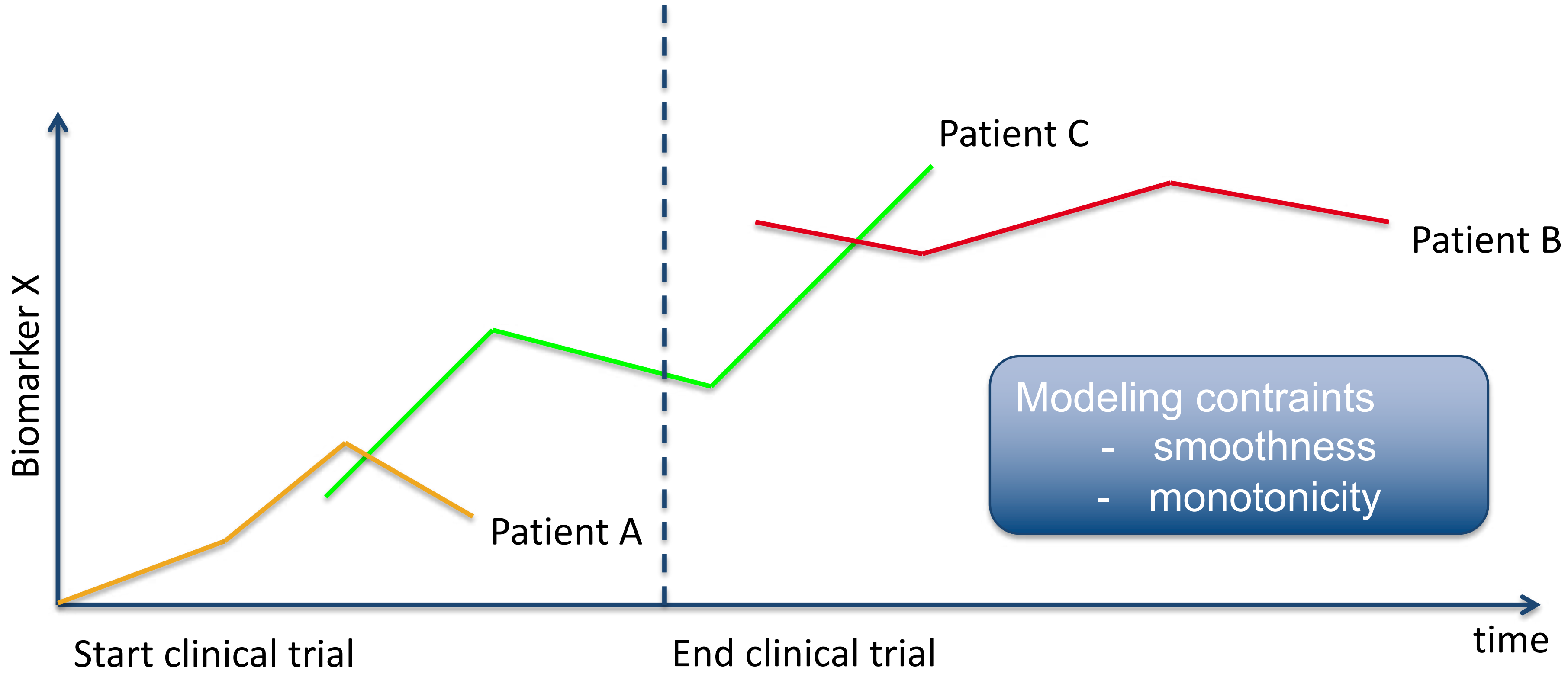


Statistical disease progression model

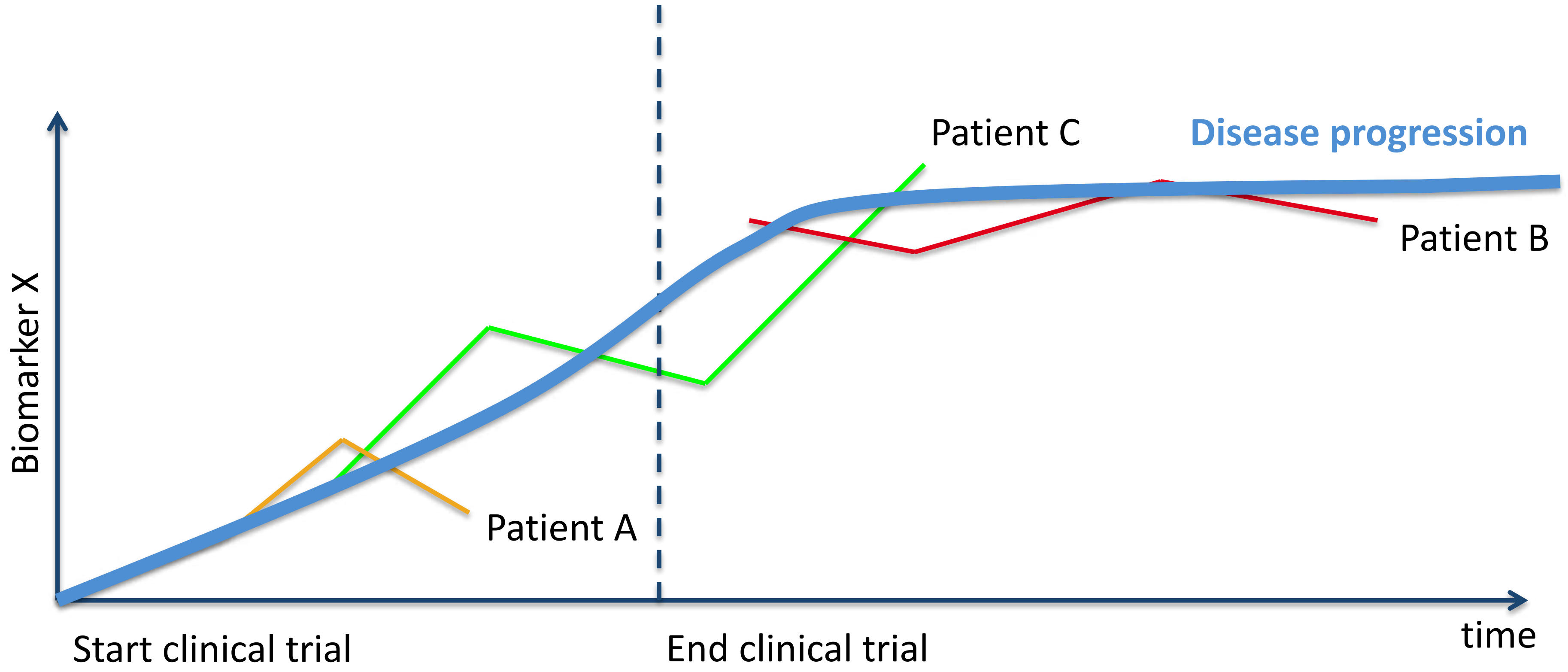


Modeling contraints
- smoothness
- monotonicity

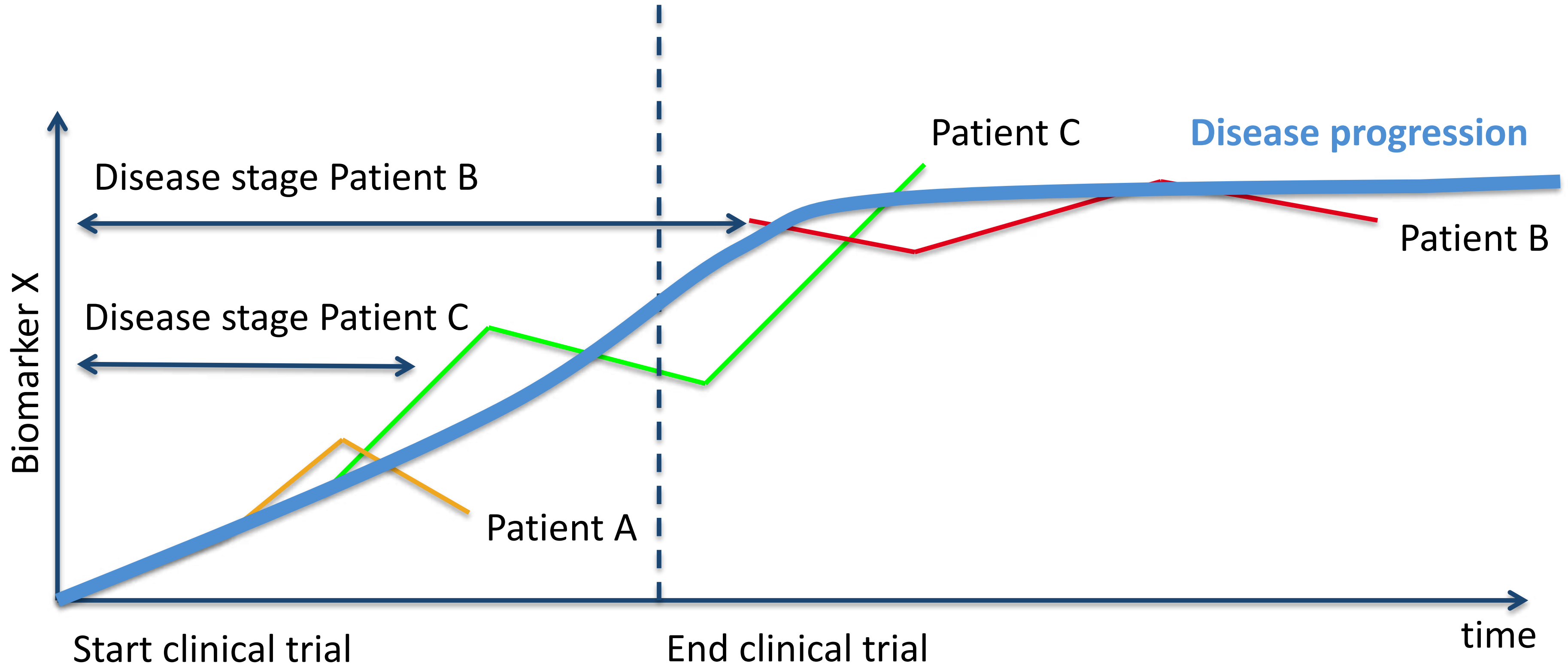
Statistical disease progression model



Statistical disease progression model

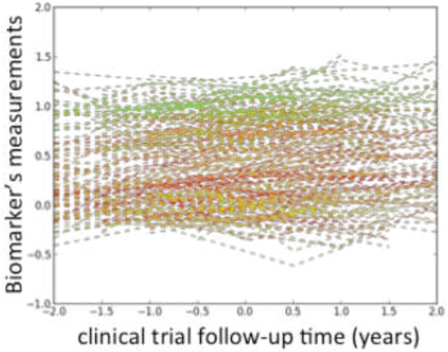


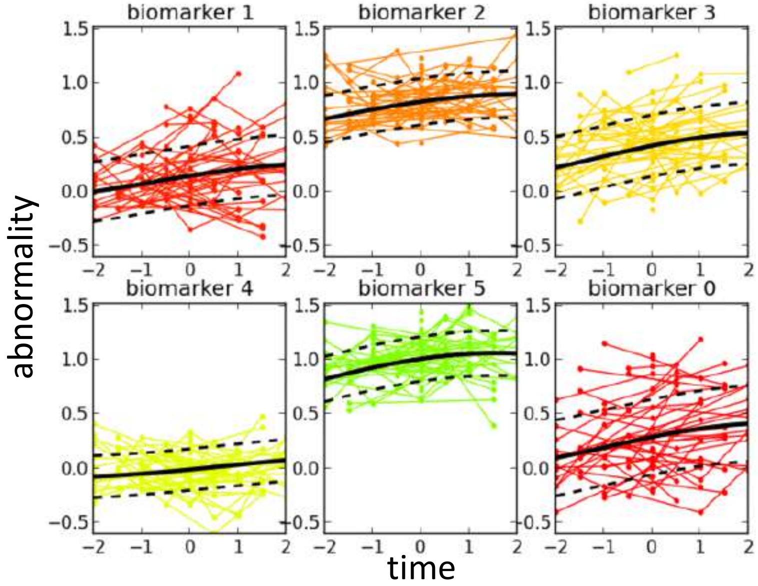
Statistical disease progression model



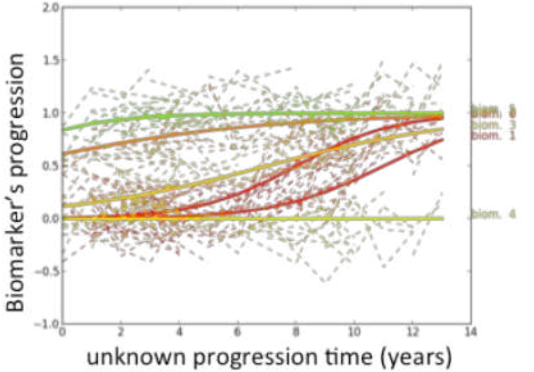
Statistical disease progression model via monotonic Gaussian Processes (GP)

Short term data



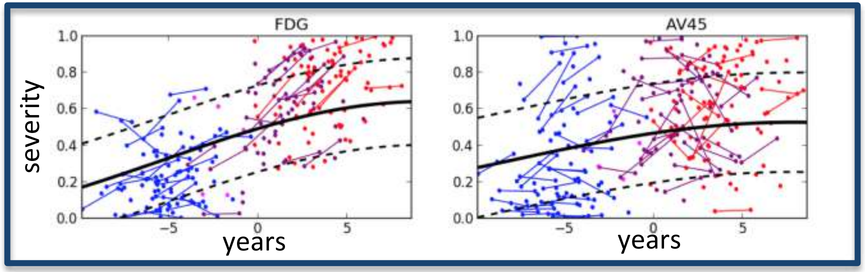


Estimated long term progressions



- Multivariate non-parametric random effects modeling
- Monotonic GP [Riihimäki & Vehtari, PMLR, 2010; Lorenzi & Filippone, ICML, 2018]
- Time reparameterization [Jedynak et al, NeuroImage 2012; Durrleman et al, IJCV, 2013; Schiratti et al, NIPS 2015]

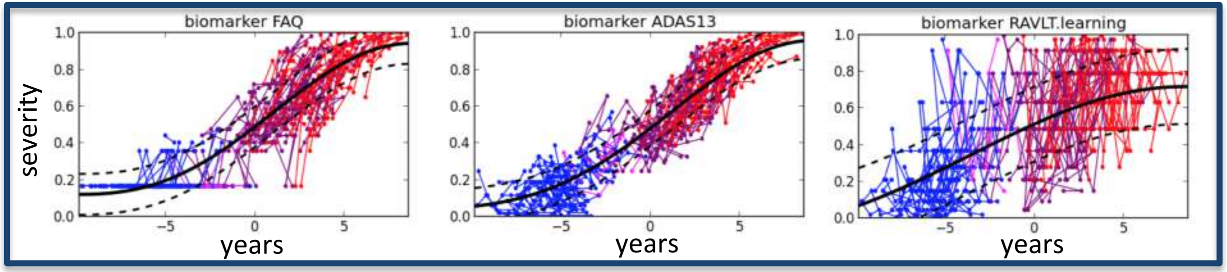
Highlighting **dynamics and relationship** between biomarkers



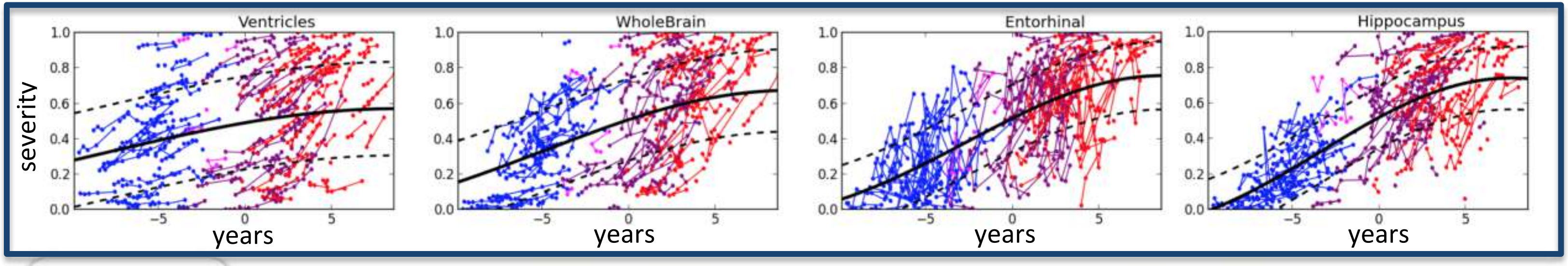
Metabolism + Amyloid

200 training subjects
- 67 healthy
- 75 AD
- 53 MCI converted
- 5 healthy converted

5 years observational time



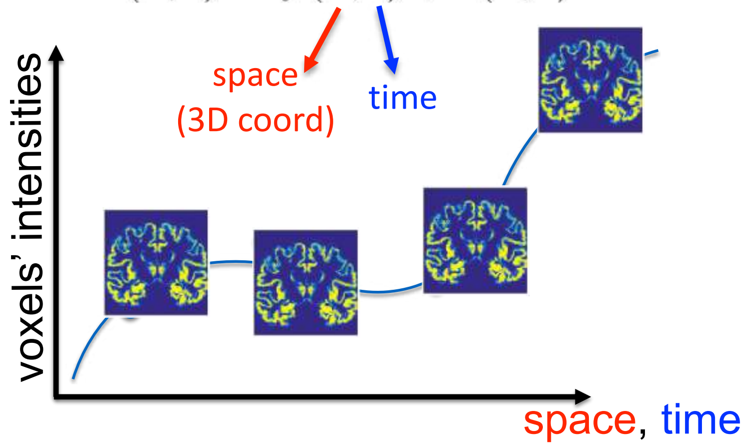
Cognition



Brain Volumes

In-silico model of brain pathology

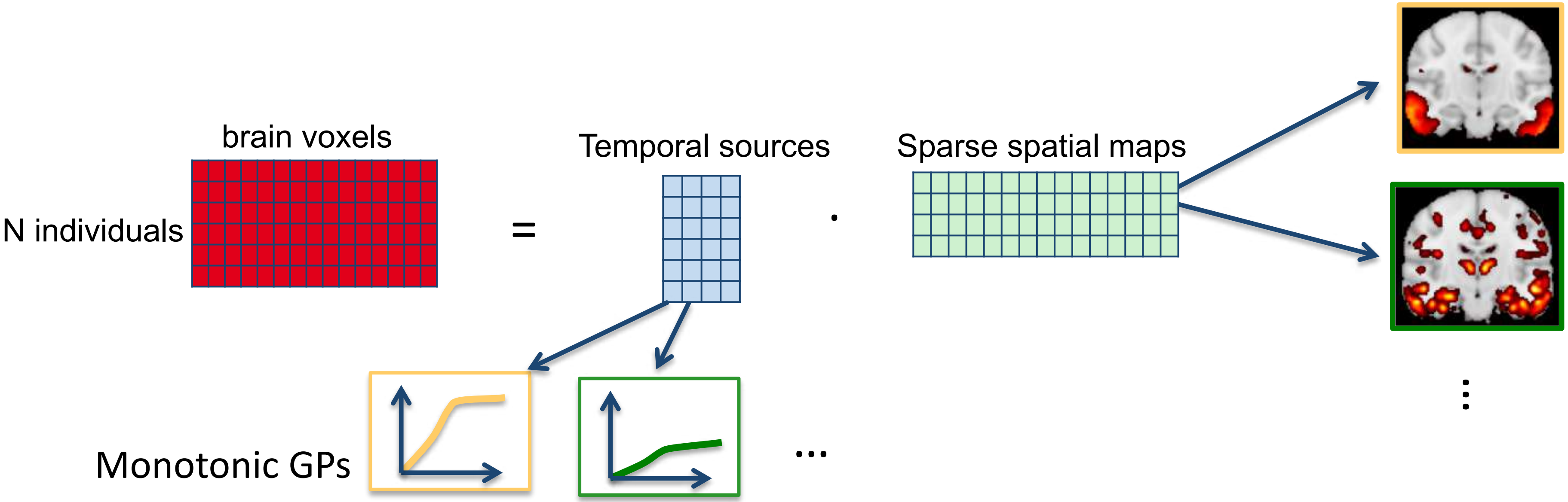
$$I(u,t) = f(u,t) + \epsilon(u,t)$$



Representing the full disease process

- Structure
  MRI
- Hypometabolism
  FDG PET
- Clinical status
  ADAS 11, MMSE, ...

In-silico model of brain pathology

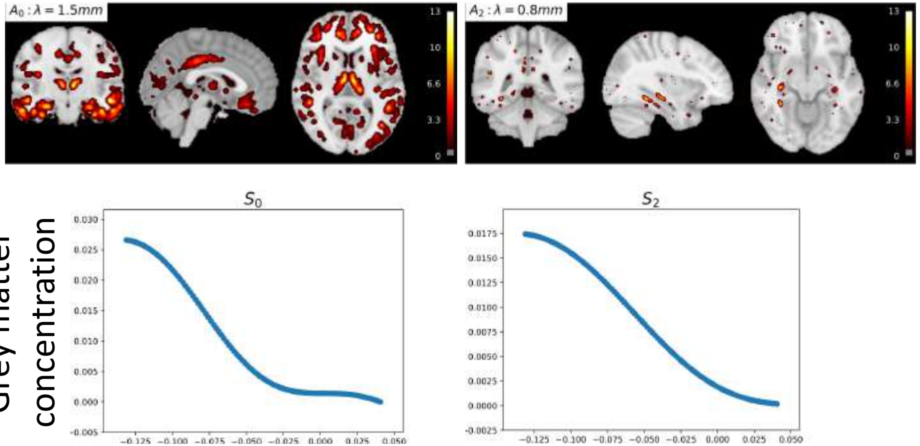


Efficient formulation through stochastic variational inference:

- Random feature expansion for GP regression [Cutajar et al, ICML 2017]
- Monotonic constraint in deep GP [Lorenzi & Filippone, ICML 2018]
- Variational dropout for sparsity and model selection [Kingma et al, NIPS, 2015; Molchanov et al, ICML 2017]

In-silico model of brain pathology

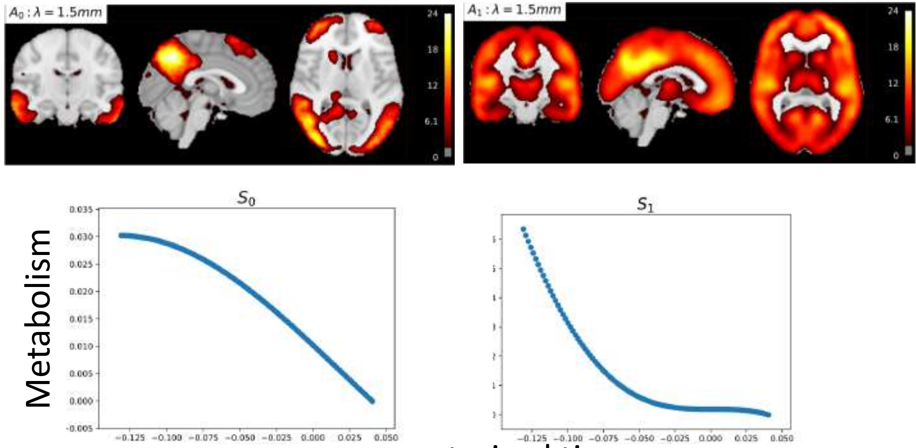
MRI (atrophy)



Grey matter concentration

reparameterized time

FDG-PET (hypometabolism)



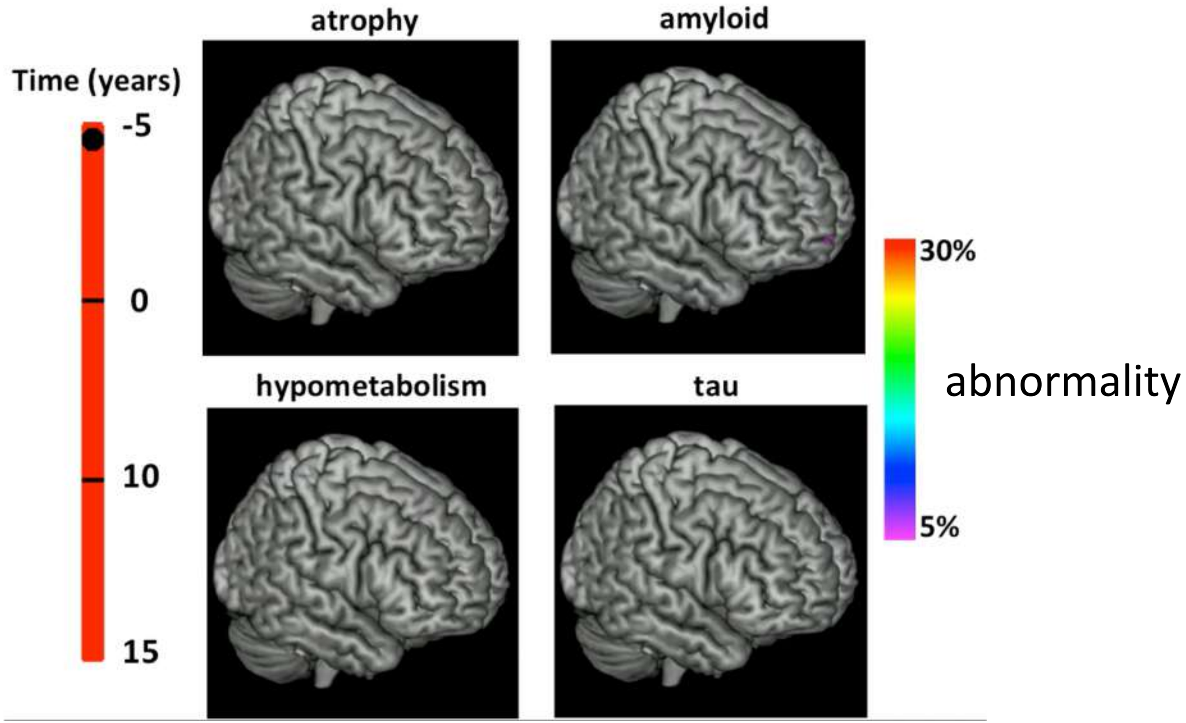
Metabolism

reparameterized time

In-silico model of brain pathology

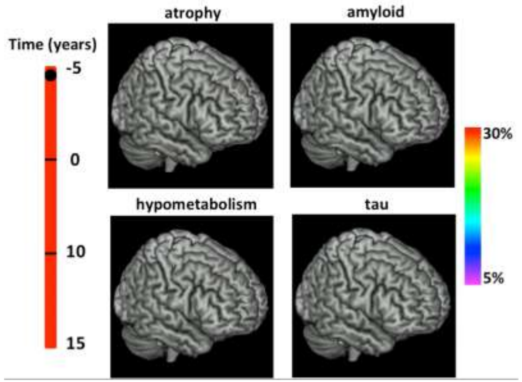
572 individuals
131 healthy
320 impaired
121 Alzheimer's

5 years max follow-up



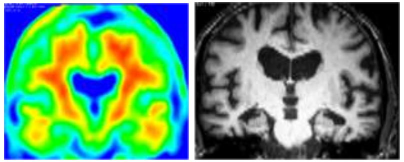
Neurodegeneration model as reference for independent studies

Geneva Memory Clinic cohort (91 individuals)



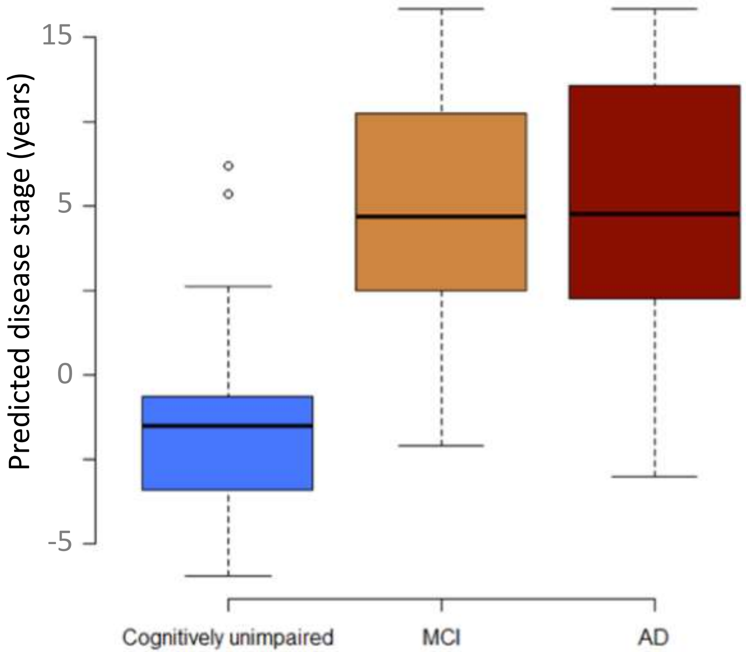
Most likely time point?

test subject

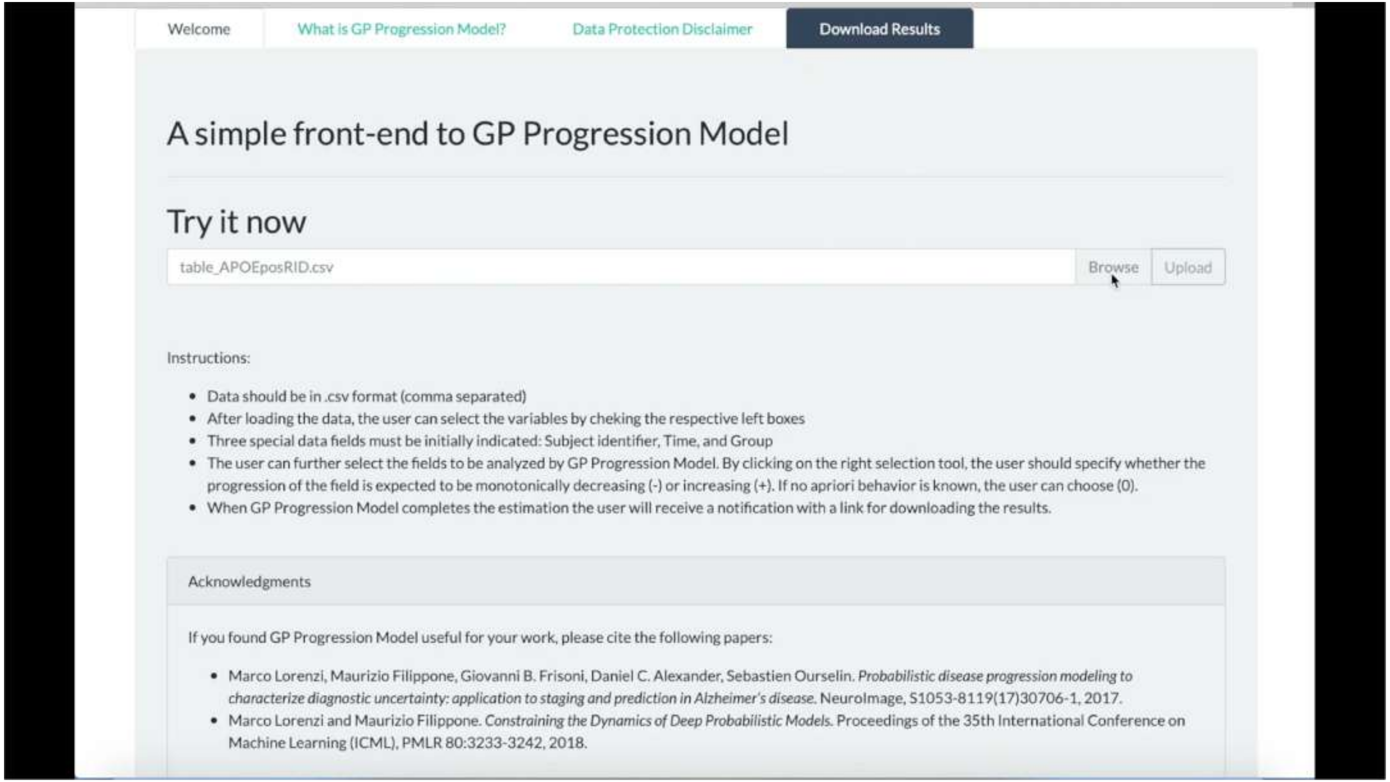


scan X

scan Y



gpprogressionmodel.inria.fr



Welcome | What is GP Progression Model? | Data Protection Disclaimer | Download Results

A simple front-end to GP Progression Model

Try it now

table_APOEposRID.csv Browse Upload

Instructions:

- Data should be in .csv format (comma separated)
- After loading the data, the user can select the variables by cheking the respective left boxes
- Three special data fields must be initially indicated: Subject identifier, Time, and Group
- The user can further select the fields to be analyzed by GP Progression Model. By clicking on the right selection tool, the user should specify whether the progression of the field is expected to be monotonically decreasing (-) or increasing (+). If no apriori behavior is known, the user can choose (0).
- When GP Progression Model completes the estimation the user will receive a notification with a link for downloading the results.

Acknowledgments

If you found GP Progression Model useful for your work, please cite the following papers:

- Marco Lorenzi, Maurizio Filippone, Giovanni B. Frisoni, Daniel C. Alexander, Sebastien Ourselin. *Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer's disease.* NeuroImage, S1053-8119(17)30706-1, 2017.
- Marco Lorenzi and Maurizio Filippone. *Constraining the Dynamics of Deep Probabilistic Models.* Proceedings of the 35th International Conference on Machine Learning (ICML), PMLR 80:3233-3242, 2018.

Thanks to Inria SED Team

Julia Elizabeth Luna,

Thibaud Kloczko,

David Rey

Challenges

- How to integrate heterogeneous biomedical measures?

- How to integrate the temporal dimension?

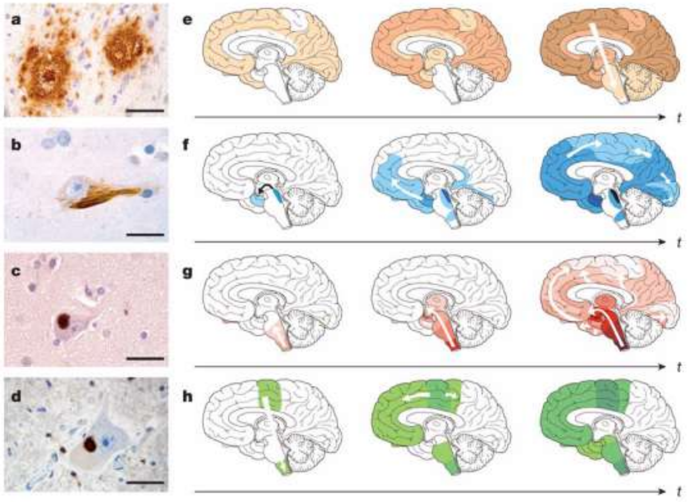
- How to unveil the biological mechanisms of the pathology?

Review Article | Published: 04 September 2013

Self-propagation of pathogenic protein aggregates in neurodegenerative diseases

Mathias Jucker & Lary C. Walker

Nature **501**, 45–51 (05 September 2013) | Download Citation



Alzheimer's (Amyloid-β)

Alzheimer's (Tau)

Parkinson's (α-Synuclein)

ALS (TDP-43)

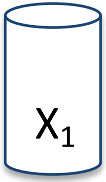
From Jucker and Walker, Nature, 2013

Neurodegeneration as prion-like disease across brain architectures

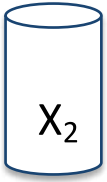
Mechanisms still unclear!

Modeling and inference of disease dynamics?

Data-driven inference of propagation dynamics

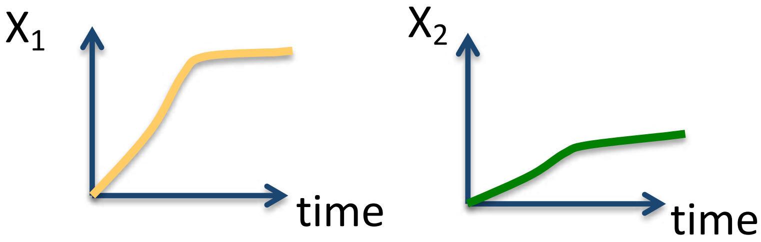


Brain area 1

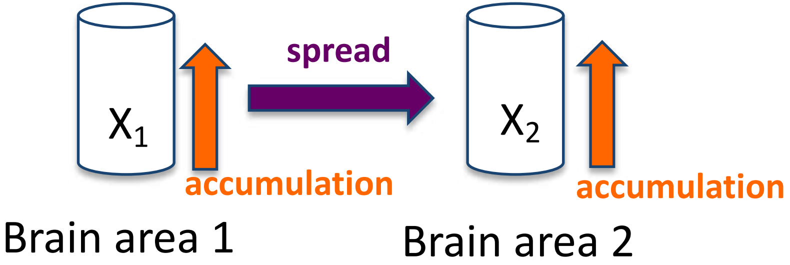


Brain area 2

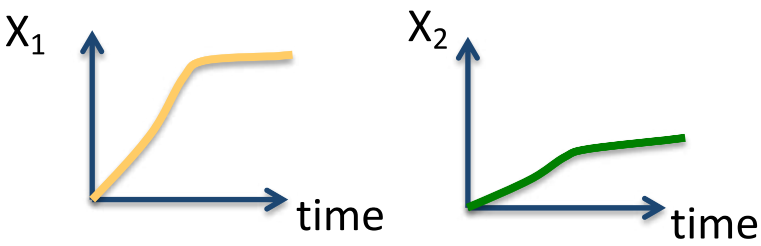
Gaussian processes



Data-driven inference of propagation dynamics

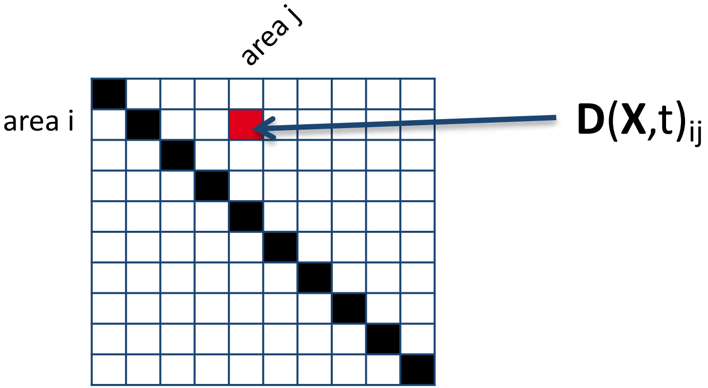


Gaussian processes

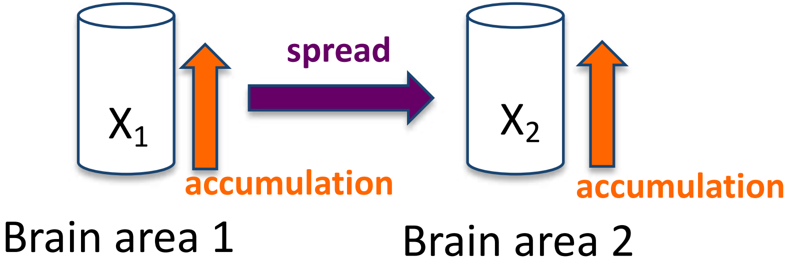


Dynamical System Modeling

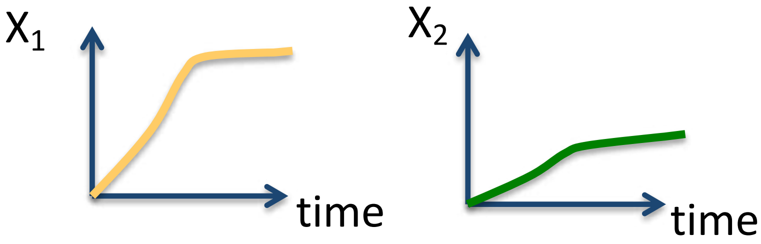
$$d\mathbf{X}/dt = \mathbf{D}(\mathbf{X},t)\ \mathbf{X}$$



Data-driven inference of propagation dynamics



Gaussian processes



Dynamical System Modeling

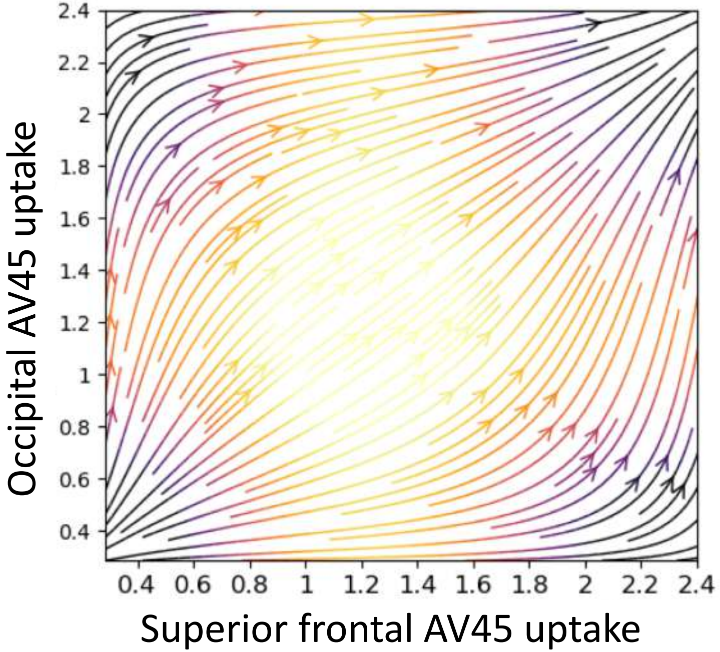
$$d\mathbf{X}/dt = \mathbf{D}(\mathbf{X},t)\ \mathbf{X}$$

- **GP for dynamical systems modeling** [Lorenzi & Filippone, ICML 2018]
- **Modeling AV45-PET imaging data**
- **Time reparameterization**

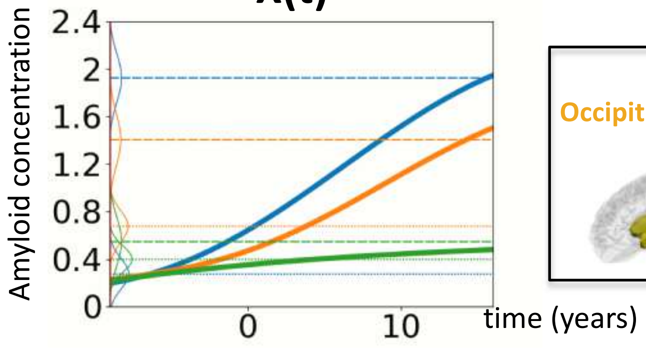
1090 individuals
369 healthy
526 impaired
195 Alzheimer's

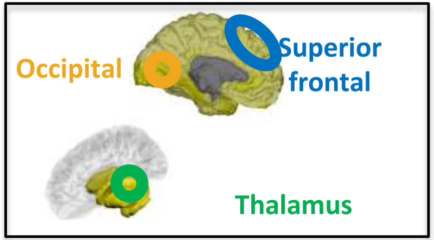
Learned propagation dynamics

X(t)

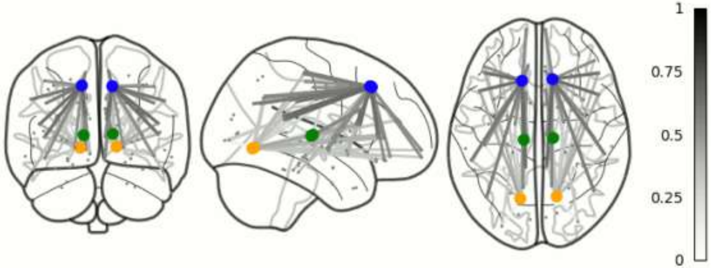


X(t)

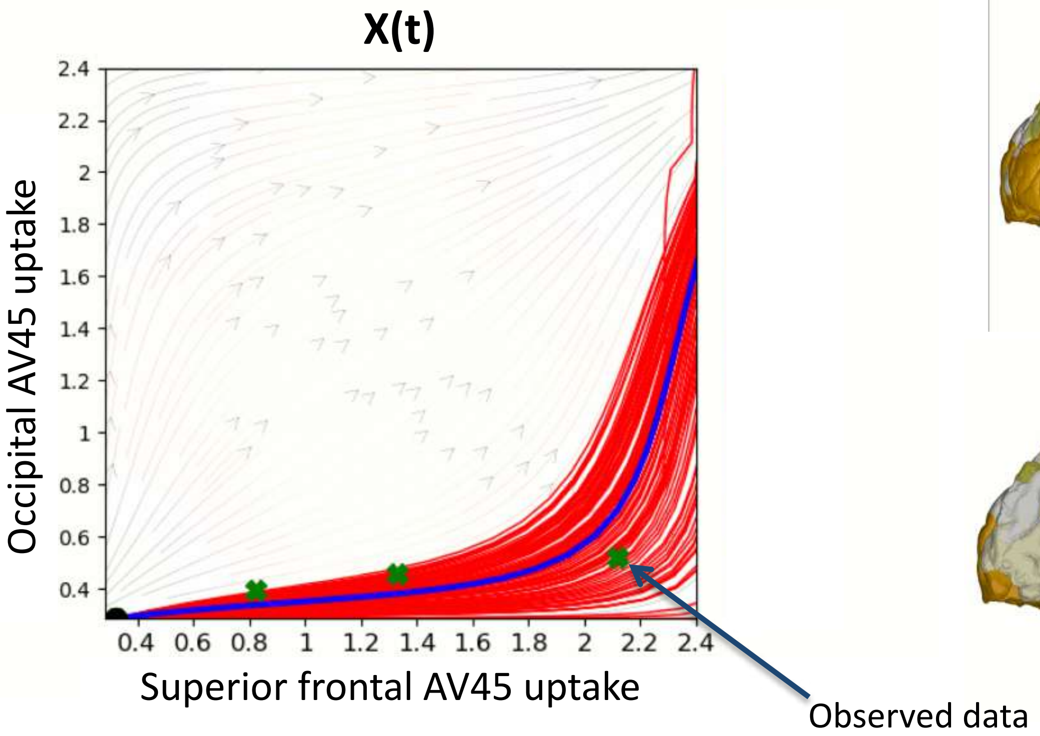


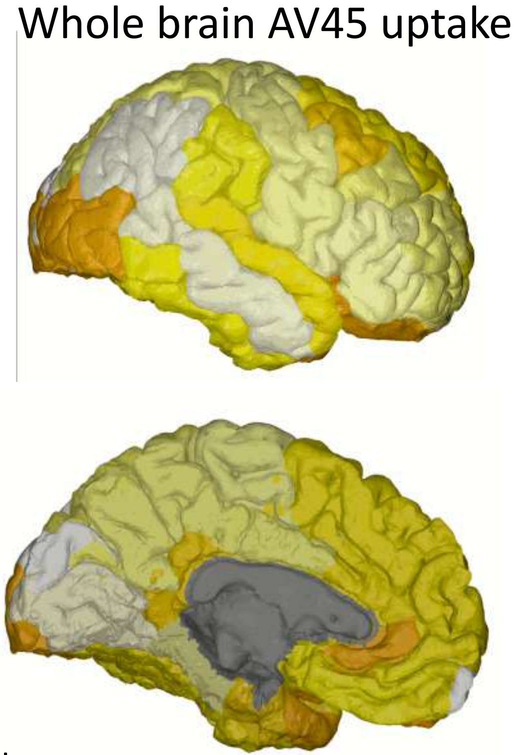


Propagation strength: D(X,t)$_{ij}$



Personalization





Prediction and interpretation of pathological evolutions

- A long-term progression model can be estimated from short-term clinical data

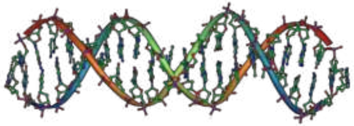
cfr. L'imagerie médicale et apprentissage automatique: vers une intelligence artificielle?

Collège de France, 2 may 2018

- Biological/Clinical statistical constraints: improved plausibility and reliability

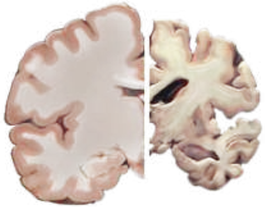
- Valuable quantitative tool: diagnosis and clinical trials

Data science



+

+



Bridging domains through statistical learning

Informatics, Mathematics, Biology, Medicine

Potential for great discoveries through Big Data

Inria

UNIVERSITÉ CÔTE D'AZUR

Epione
e-patient / e-medicine

mNC³


N. Ayache


P. Robert


V. Manera


C. Abi Nader


L. Antelmi

UCA-Ville de Nice Young Researcher award


S. Garbarino

UCA MDLab


M. Milanesio

Meta-ImaGen


S.S. Silva

Brain-Heart


J. Banus


M. Sermesant

EURECOM Sophia Antipolis


M. Filippone

UCL


A. Altmann

ILLINOIS INSTITUTE OF TECHNOLOGY


B. Gutman

HUG Hôpitaux Universitaires Genève

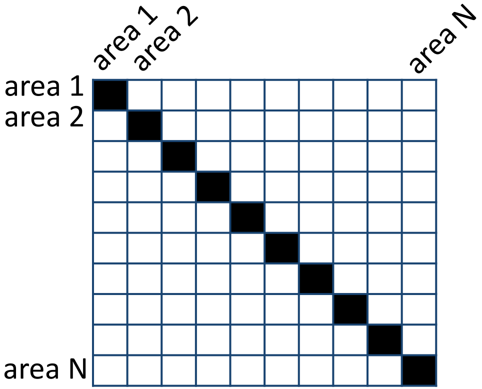

G.B. Frisoni


F. Ribaldi

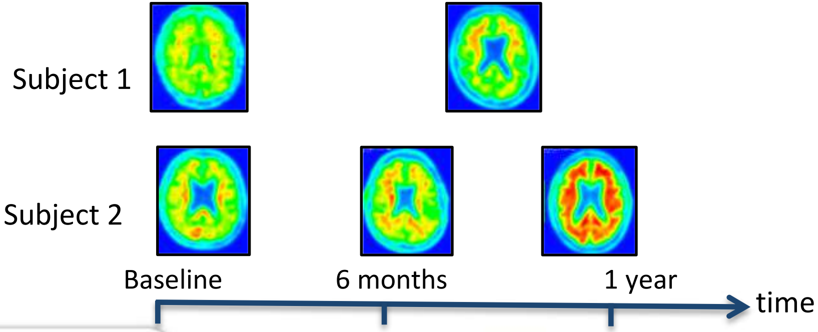
Thank you!

Brain architecture: structural connectivity across anatomical areas

+



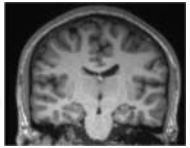
A proxy for protein deposition: data collections of longitudinal AV45-PET amyloid scans



- No consistent definition of time axis
- No clear model dynamics

Latent variable modeling of multivariate data

Patient's latent status



X

Imaging

~ 10^5 voxels

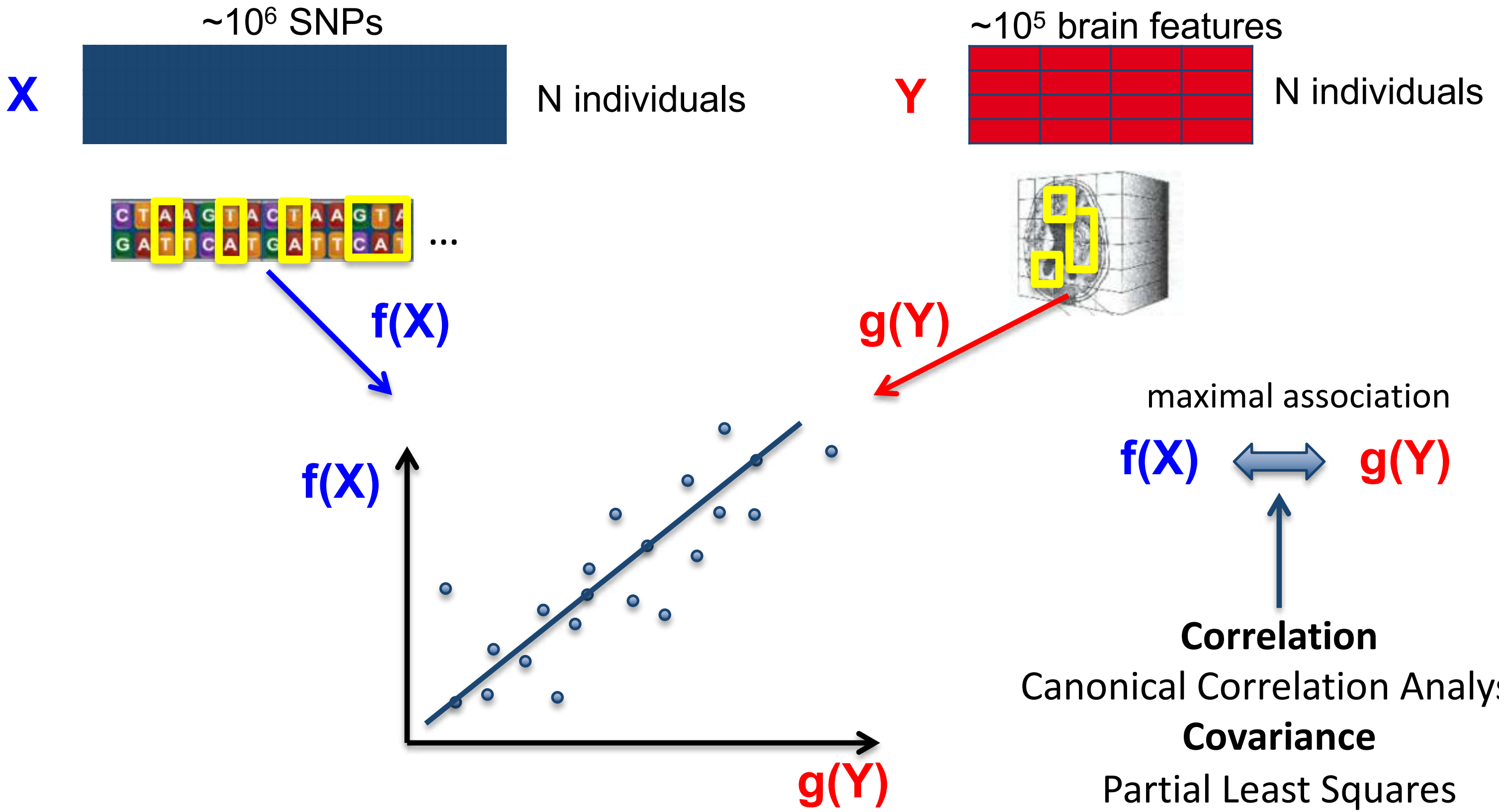
?

Y

Genetics

~ 10^6 single nucleotide plymorphisms (SNPS)

Latent variable modeling of multivariate data



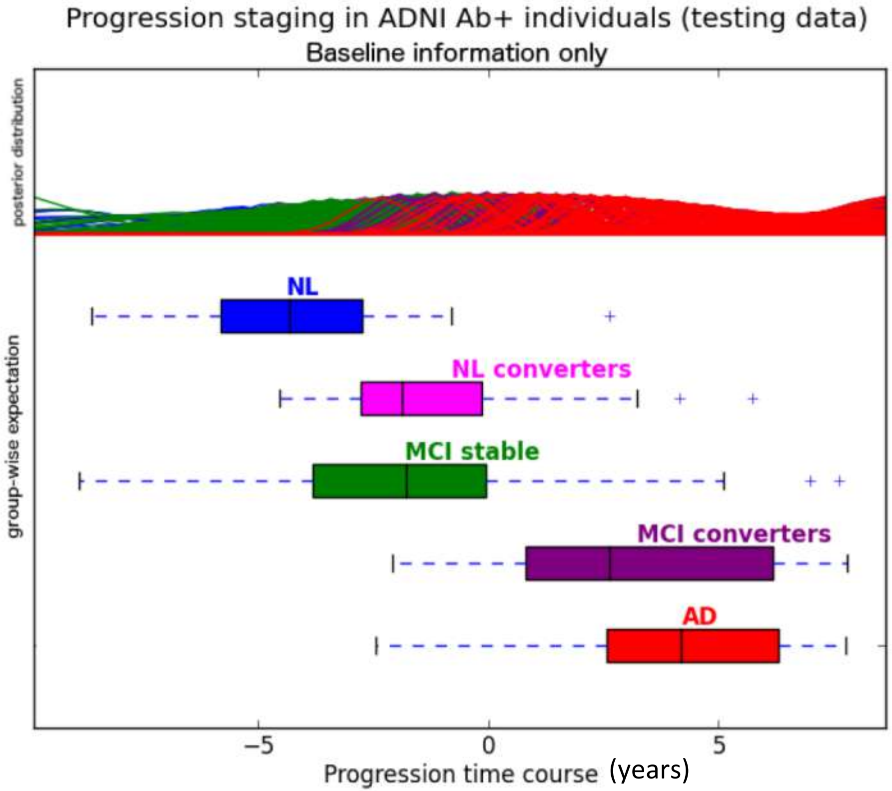
Correlation

Canonical Correlation Analysis

Covariance

Partial Least Squares

Predictions in testing data



585 testing subjects

- 74 healthy
- 145 AD
- 106 MCI converted
- 17 healthy converted
- 243 MCI stable

Accuracy

- 89% AD vs NL
- 82% MCIc vs MCIs
- 83% NL vs NLconv

Challenges

- How to integrate heterogenous biomedical measures

- How to analyse private healthcare data collected worldwide?

- How to predict pathological evolution for a given individual?

Big Data in medicine

Single hospital: 100s – 1’000s patients

Data from many hospitals needed

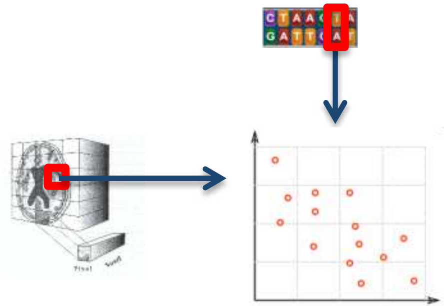
Access to multiple centers data
falls into General Data Protection Regulation (GDPR):
Privacy, confidentiality, security, ...

Data cannot be gathered in a single centre!
Standard learning algorithms cannot be used in multicentric data

Big Data in medicine

Circumventing the problem of data access

Federated-analysis (or meta-analysis)



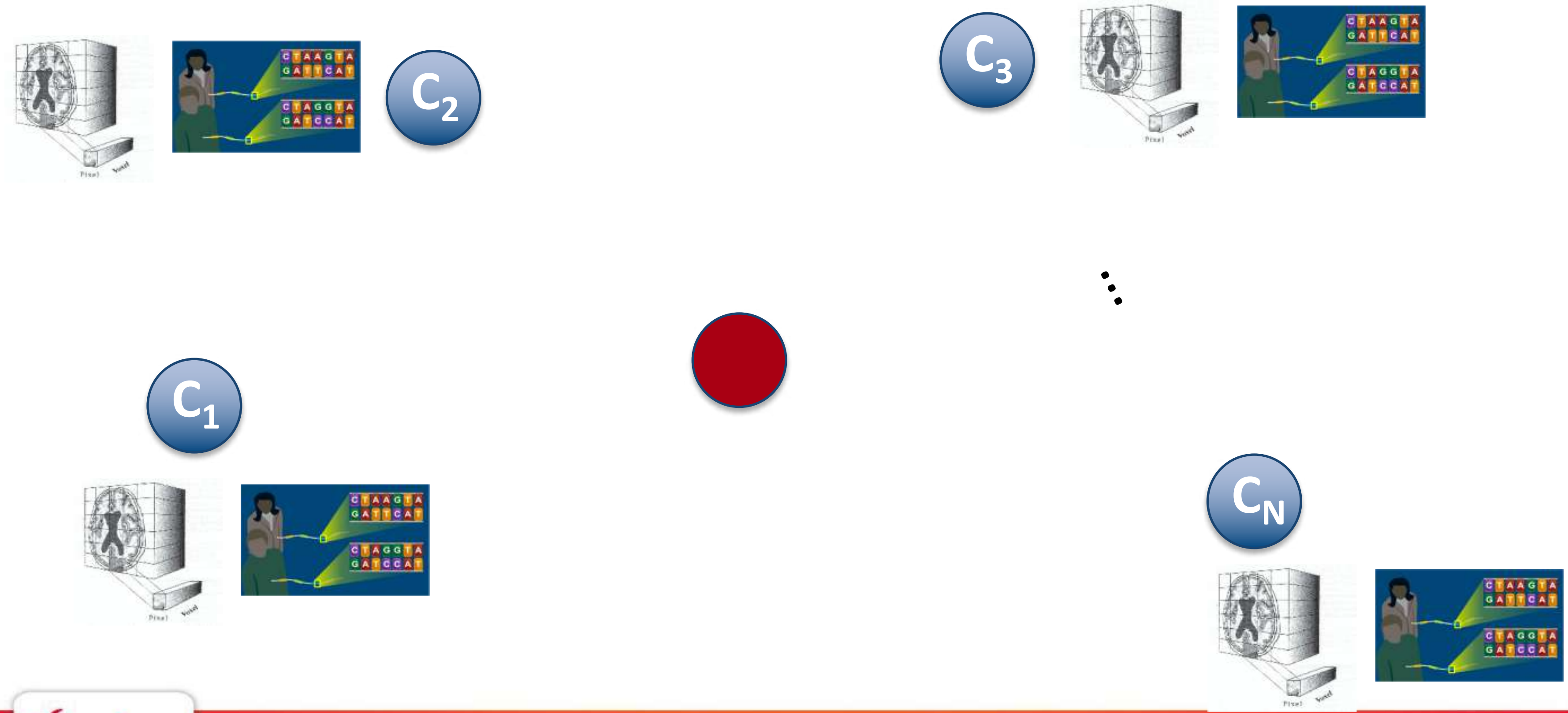
Is the association significant?

Hospital	1	2	3	4	…
Answer	yes	yes	no	yes	…

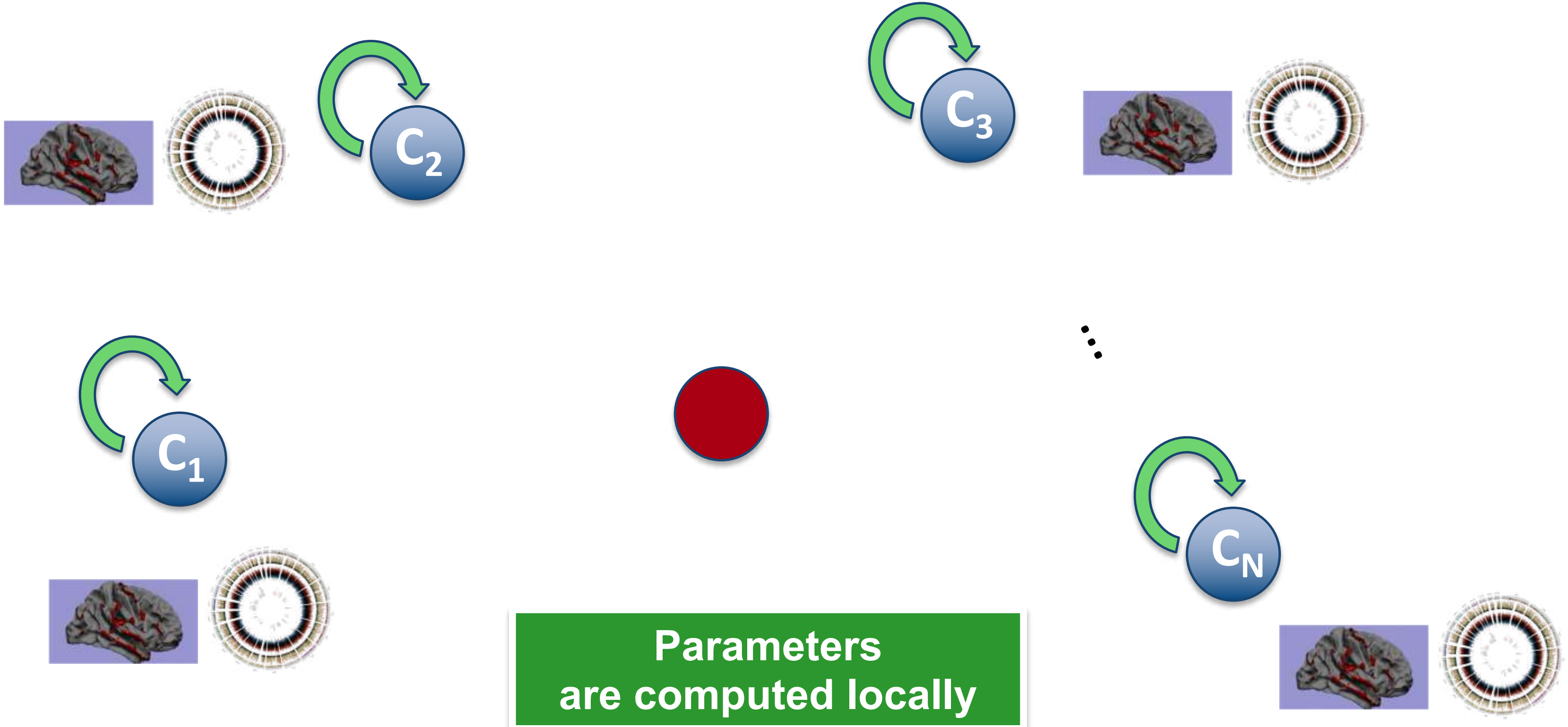
Meta-answer: yes

- No data sharing
- Ok for standard statistical testing (p-values, effect size)
- **No complex modeling possible**

Advanced Data Science through federated learning

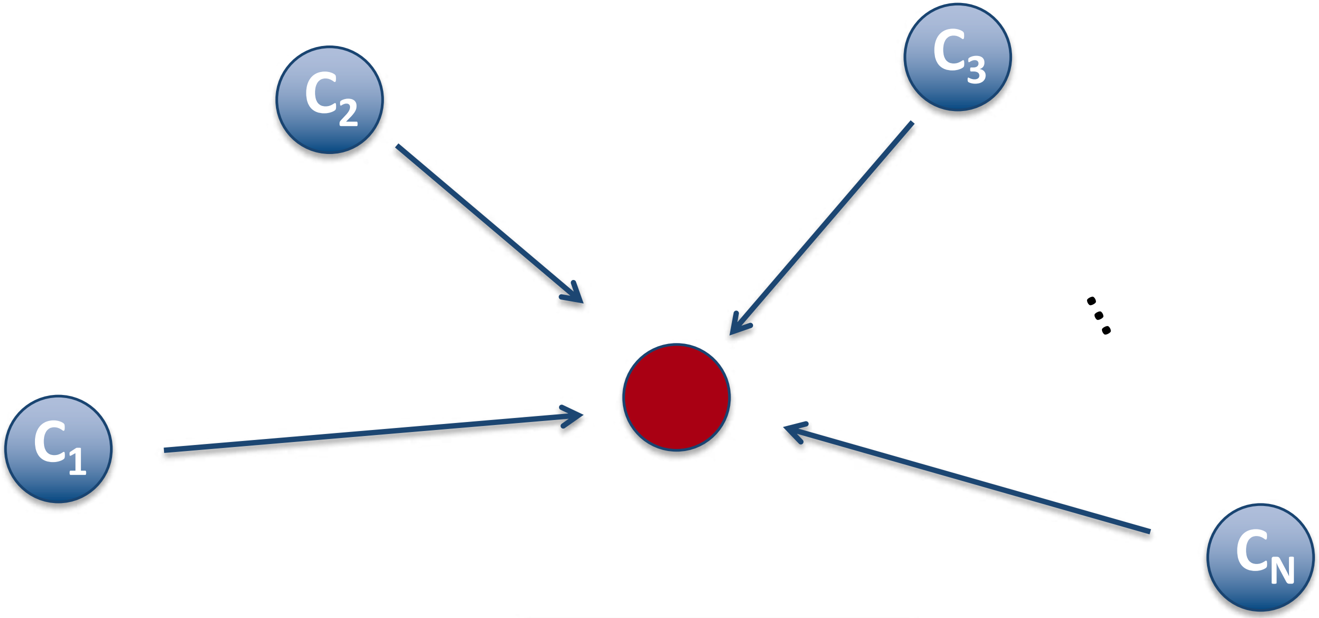


Advanced Data Science through federated learning



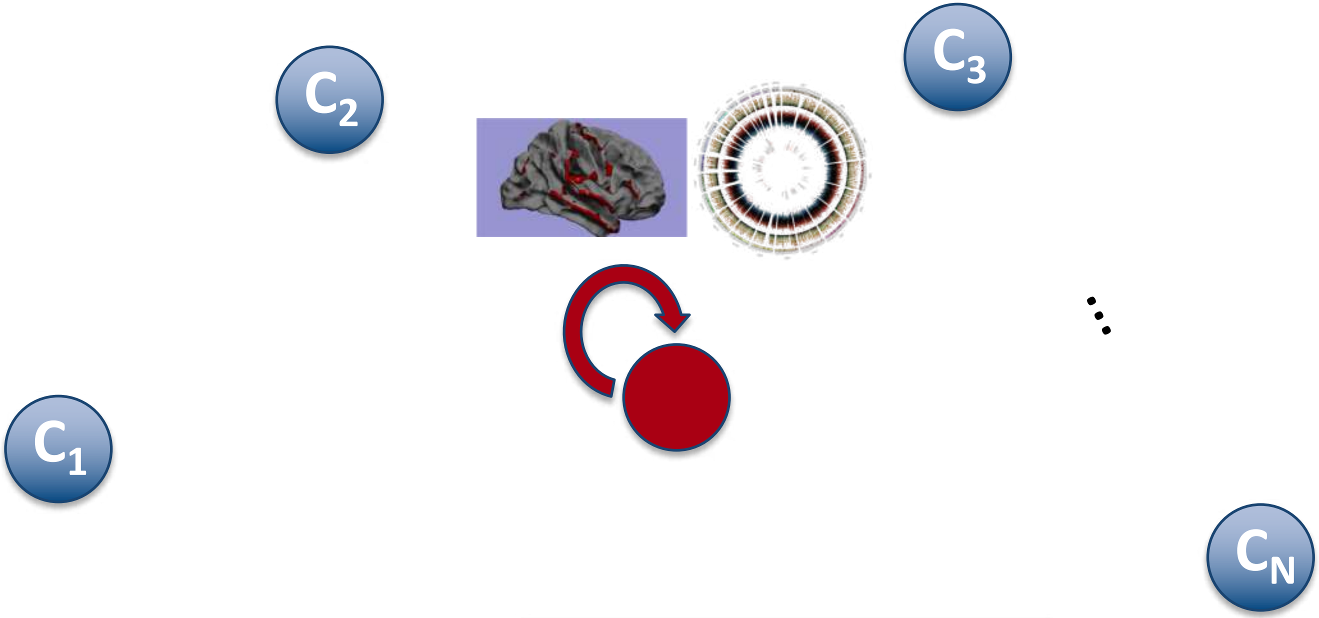
Parameters
are computed locally

Advanced Data Science through federated learning



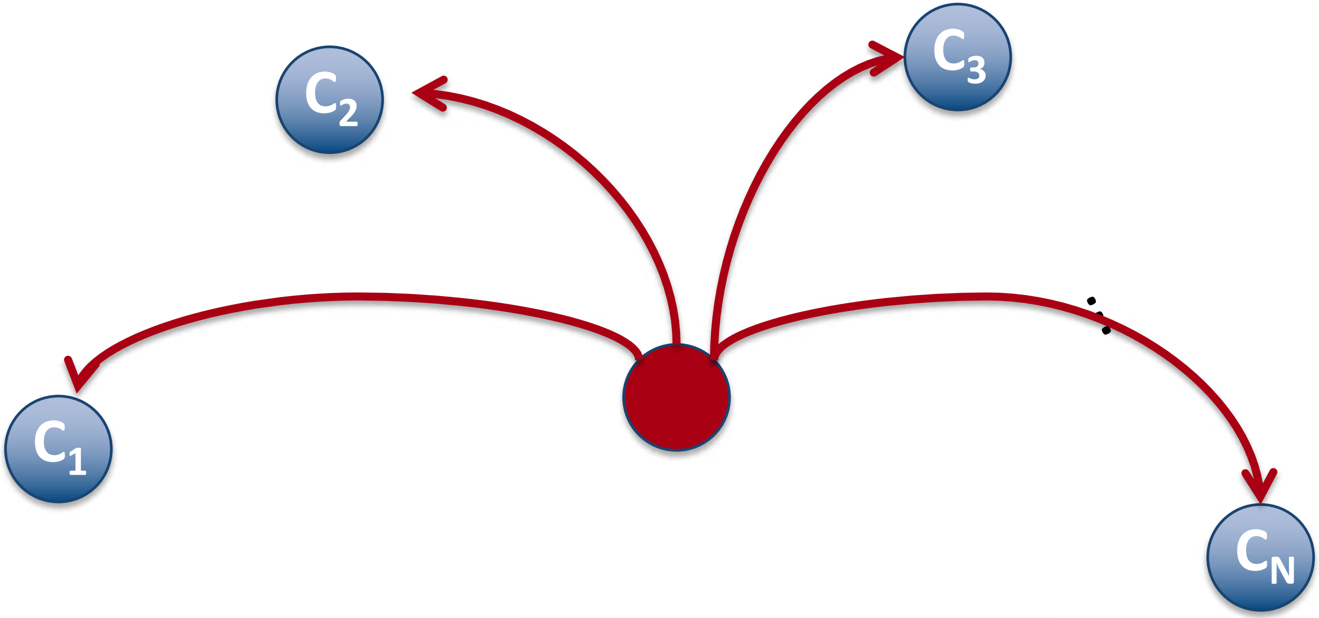
Parameters are shared

Advanced Data Science through federated learning



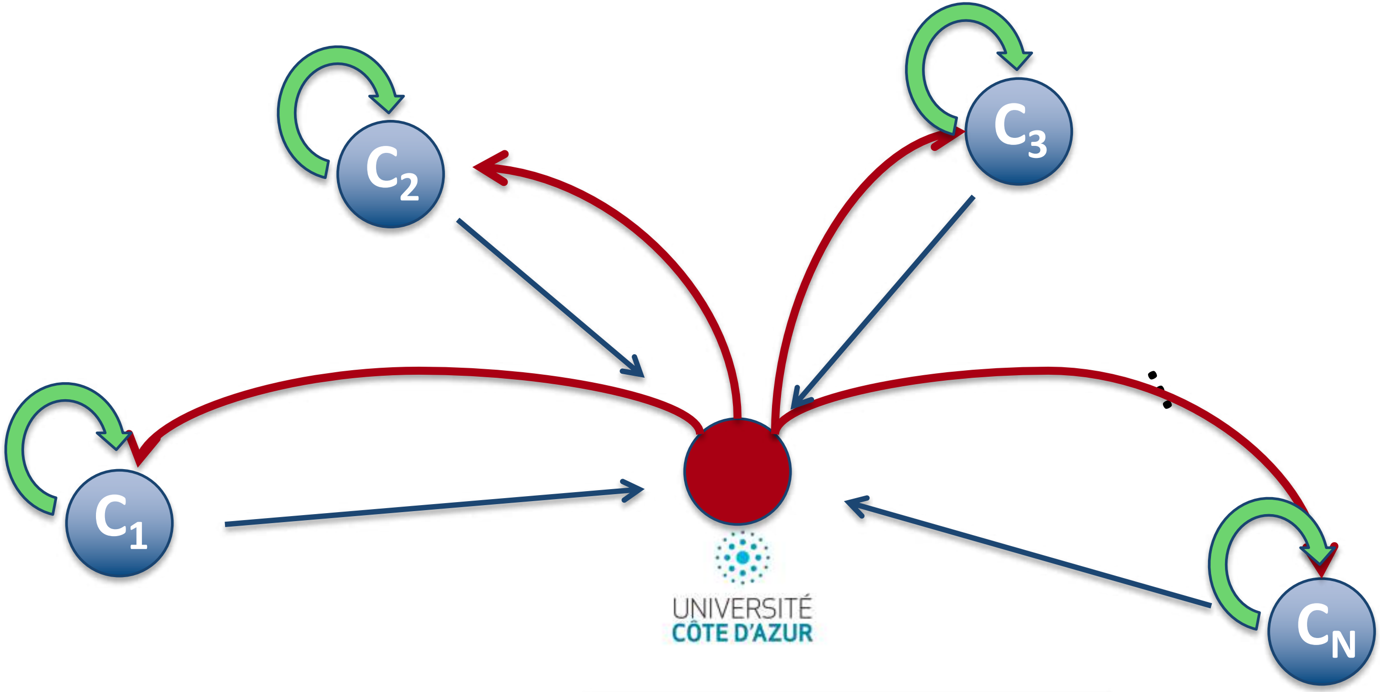
A federated model is learned

Advanced Data Science through federated learning



The federated model is shared

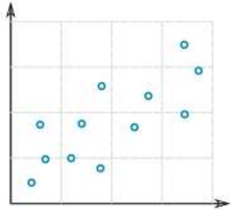
Advanced Data Science through federated learning



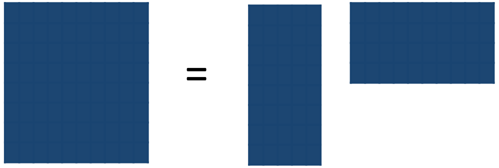
The procedure is iterated

Federated analysis toolkit

A methodology for distributed



Linear modeling



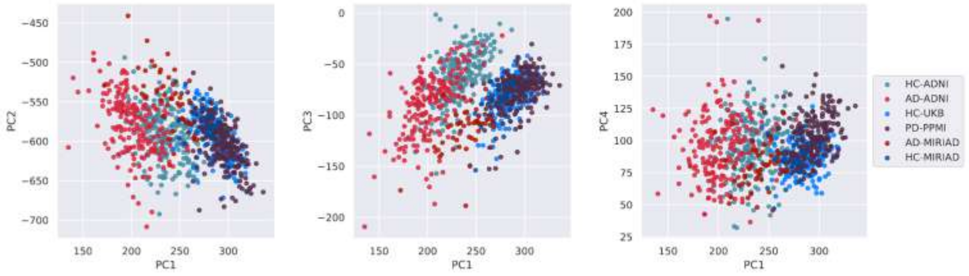
Matrix factorization

Allows a federated framework for several key statistical operations:
Data standardization, accounting for covariates, dimensionality reduction, ...

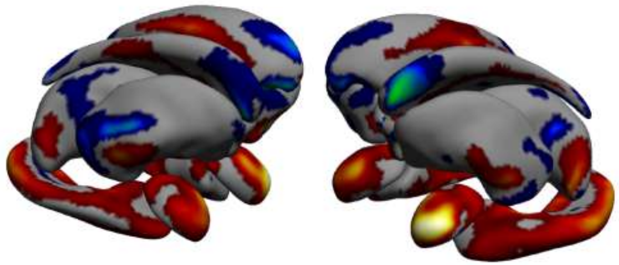
Federated analysis of subcortical brain regions in dementia

ADNI	PPMI	UK Biobank	Miriad
Alzheimer’s	Parkinson’s	Healthy	Alzheimer’s
802	232	208	68

Projection on latent components



Brain subcortical components



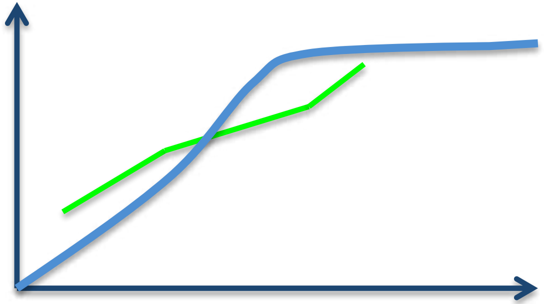
Challenges

How to analyse private healthcare data collected worldwide?

Answer from Statistical Learning:

- Advanced statistical modeling can be compatible with Data privacy and anonymity
- Federated learning applies for simple linear models as well as for more complex ones (e.g. neural networks or Gaussian processes)
- Future research needed for improving data security, data transfer bottlenecks, modeling flexibility

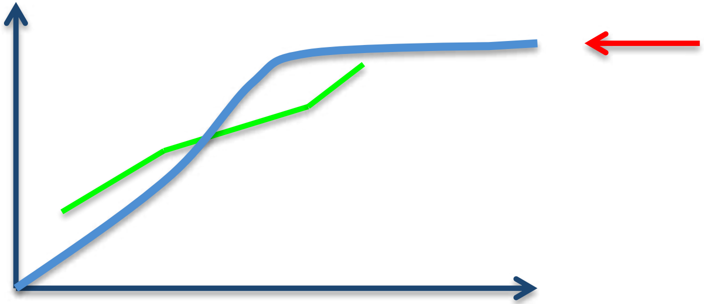
Random effects modeling



Individuals
biomarkers' observations

$$\mathbf{y}^j(t) = \left(y^j_{b_1}(t), y^j_{b_2}(t), \ldots, y^j_{b_{N_b}}(t)\right)^\top$$

Random effects modeling



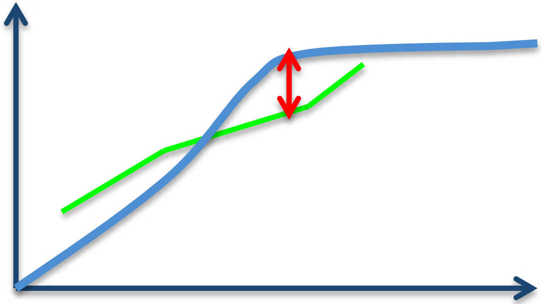
Individuals
biomarkers' observations

$$\mathbf{y}^j(t) = \left(y^j_{b_1}(t), y^j_{b_2}(t), \dots, y^j_{b_{N_b}}(t)\right)^\top$$

$$= \mathbf{f}(\phi^j(\tau))$$

Fixed effect

Random effects modeling



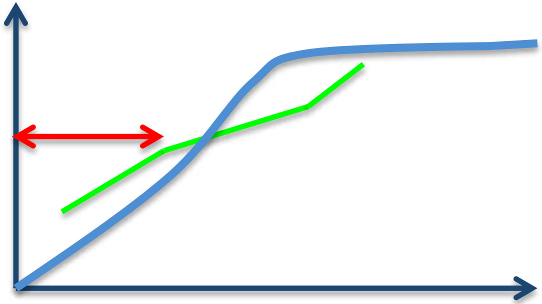
Individuals
biomarkers' observations

$$\mathbf{y}^j(t) = \left(y^j_{b_1}(t), y^j_{b_2}(t), \ldots, y^j_{b_{N_b}}(t)\right)^\top$$

$$= \mathbf{f}(\phi^j(\tau)) + \boldsymbol{\nu}^j(\phi^j(\tau)) + \epsilon$$

Random effect
+ observational noise

Random effects modeling



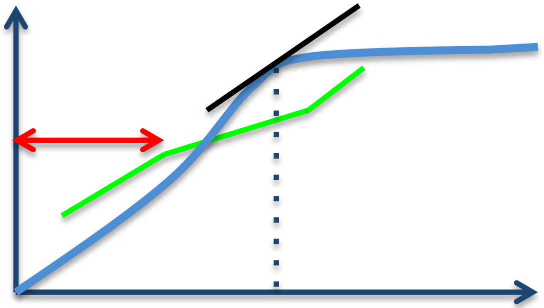
Individuals
biomarkers' observations

$$\mathbf{y}^j(t) = \left(y^j_{b_1}(t), y^j_{b_2}(t), \ldots, y^j_{b_{N_b}}(t)\right)^\top$$

$$= \mathbf{f}(\phi^j(\tau)) + \nu^j(\phi^j(\tau)) + \epsilon$$

Subject-specific time reparameterization

Random effects modeling



Individuals
biomarkers' observations

$$\mathbf{y}^j(t) = \left(y^j_{b_1}(t), y^j_{b_2}(t), \ldots, y^j_{b_{N_b}}(t)\right)^\top$$

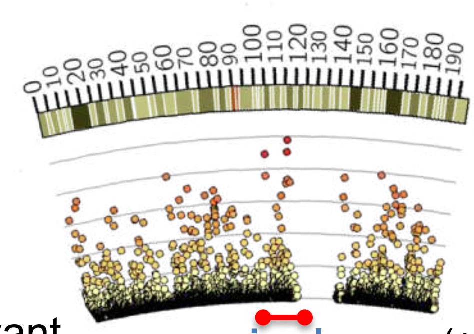
$$= \mathbf{f}(\phi^j(\tau)) + \boldsymbol{\nu}^j(\phi^j(\tau)) + \epsilon$$

Subject-specific time reparameterization

Formulation through **Gaussian process regression with constraint on dynamics**

Multivariate Imaging-genetics modeling

PLS statistical result



p relevant locus

proximal areas (+/- 5kbp)

From 1’000’000 to 148 variants

Meta analysis on gene annotation databases

Significance (p-value)

	training	testing
TM2D1	0.005	0.053
IL10RA	0.107	0.620
TRIB3	0.003	0.003
ZBTB7A	0.036	0.913
LYSMD4	0.000	0.206
CRYL1	0.621	0.118
FAM135B	0.000	0.559
...	...	...

TRIB3

- neuronal cell death,
- modulation of PSEN1 stability,
- interaction with APP.