A Machine Learning Analysis of the Progression of Alzheimer's Disease Using the Alzheimer's Disease Neuroimaging Initiative (ADNI) Cohort

JSM 2014,
August 6th, 2014,

bruno.jedynak@jhu.edu

Engineering team: Andrew Lang, Bo Liu, Jerry Prince, Brian Caffo, Murat bilgel, Runze Tan, Chun-Guang, and Zhou Ye;

Medical team: Kostas Lyketsos, Susan Resnik, Sterling Johnson, and Pierre Jedynak

Computational Neurology of Neurodegenerative Diseases

- Longitudinal databases of hundreds of subjects observed during several years with tens of validated biomarkers are becoming available, allowing the **creative** use of computational methods in neurology:
- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Parkinson Progression Marker Initiative (PPMI)
- Predict-HD, TrackOn-HD for Huntington disease
- .. Much more
- Would it be possible that a discovery in neurology come from an innovative analysis of the data ?

Potential contributions

1. Insight about the disease process
    1. Is there a single mechanism under a given disease name ?
    2. Validate a biomarker of disease progression
2. Provide instruments for more efficient drug discovery process
    1. Entry criteria: Define subjects which are likely to benefit from a given drug
    2. Measure precisely the disease progression
3. Help the neurologist
    1. Which biomarkers are the most informative at a given stage of a disease
    2. What can be expected if the patient is untreated

Note Title 6/28/2013

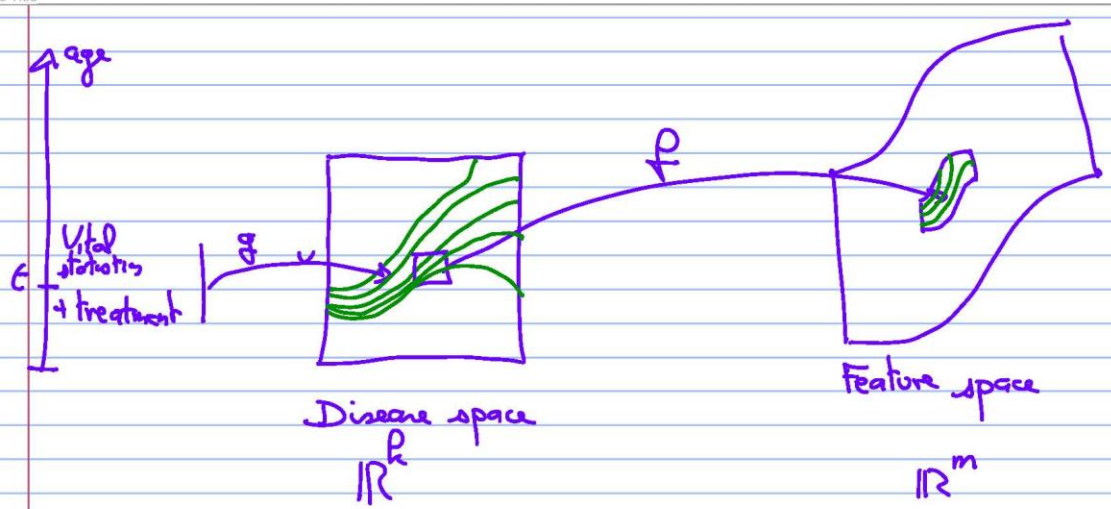
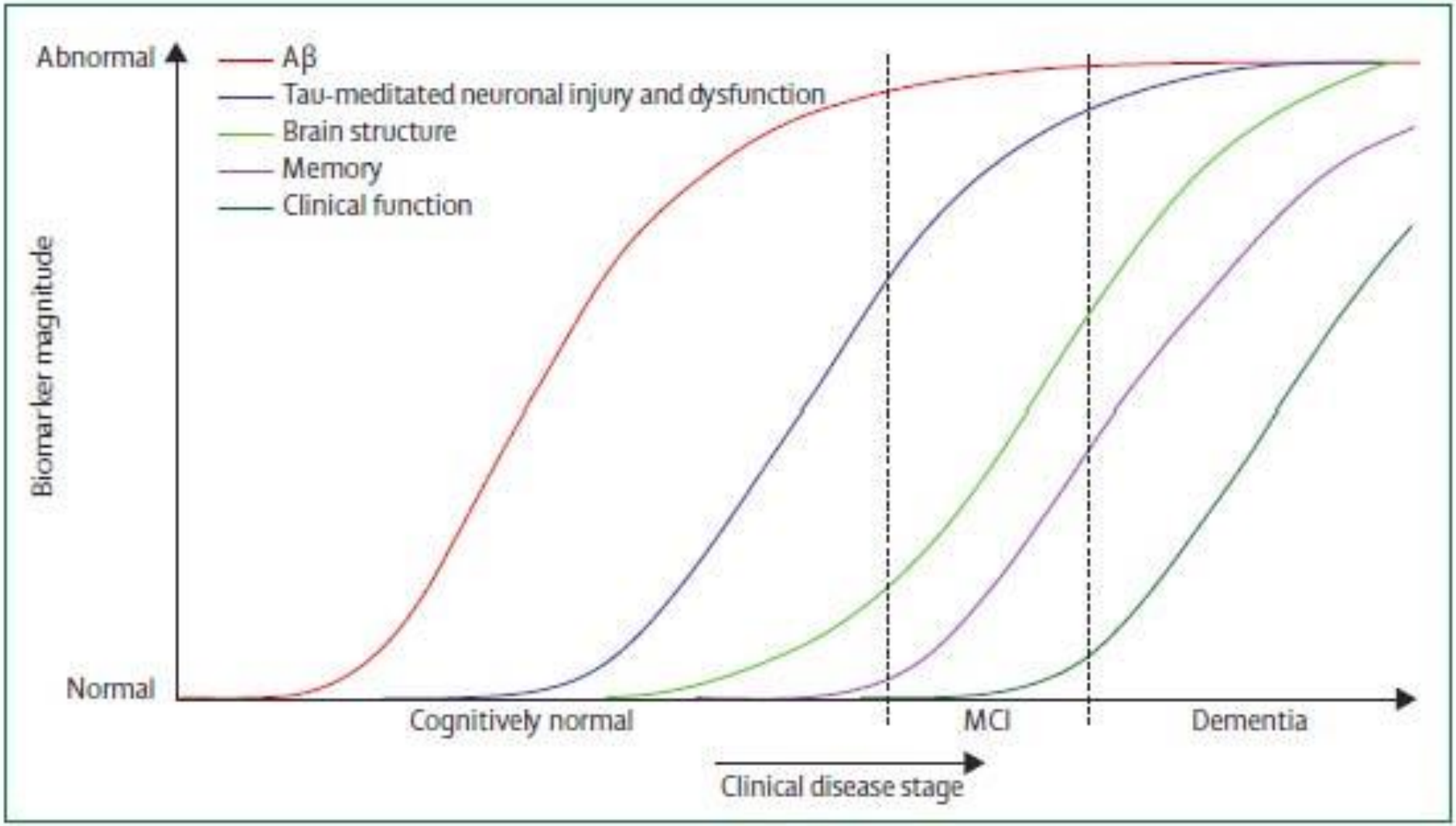


Table of content

Reprinted from The Lancet Neurology, Vol. 9, Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, John Q Trojanowski, Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade, pages 119-128, 2010, with permission from Elsevier



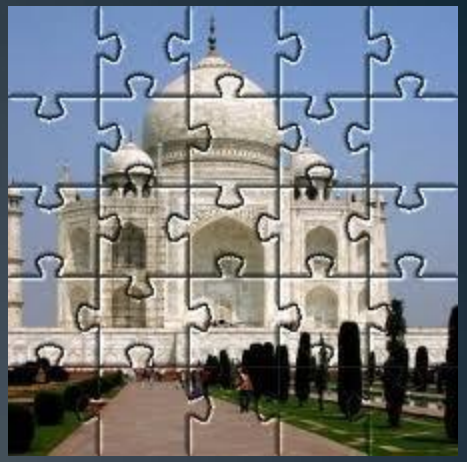
Principles for building an Alzheimer's Disease Progression Scale (ADPS) for each subject in the ADNI

1. Qualitatively, there is a single disease progression for late onset AD.
2. Along the disease progression, each biomarker change continuously and monotonically from Normal status to Abnormal status, following a sigmoid curve.
3. Each subject in the ADNI is progressing at constant speed relative to its age during the time it is observed.

One dimensional model

- $y_{ijk} = f_k(\alpha_i t_{ij} + \beta_i, \theta_k) + \sigma_k \varepsilon_{ijk}$ with
- $f_k(s, \theta = (a, b, c, d)) = \frac{a}{1+e^{-b(s-c)}}$+d
- t_{ij} is the age of subject $i$at visit j
- $s_{ij} = \alpha_i t_{ij} + \beta_i$ is the ADPS for subject i at visit j
- k is the index of a biomarker
- ε_{ijk} are independent, standard, normal
- Ref: self-modeling (K=1)

Puzzle analogy



Think of a puzzle which is finished. It provides

1. A picture, a scene, which was invisible when the pieces were scrambled
1. A localization for each piece

Similarly, computing the ADPS provides

1. A visualization of the biomarker values along the time-line of AD
2. A score for each time-point of each subject

Note:

The ADPS requires calibration in translation and scale. This calibration is performed with the Normal subjects in ADNI

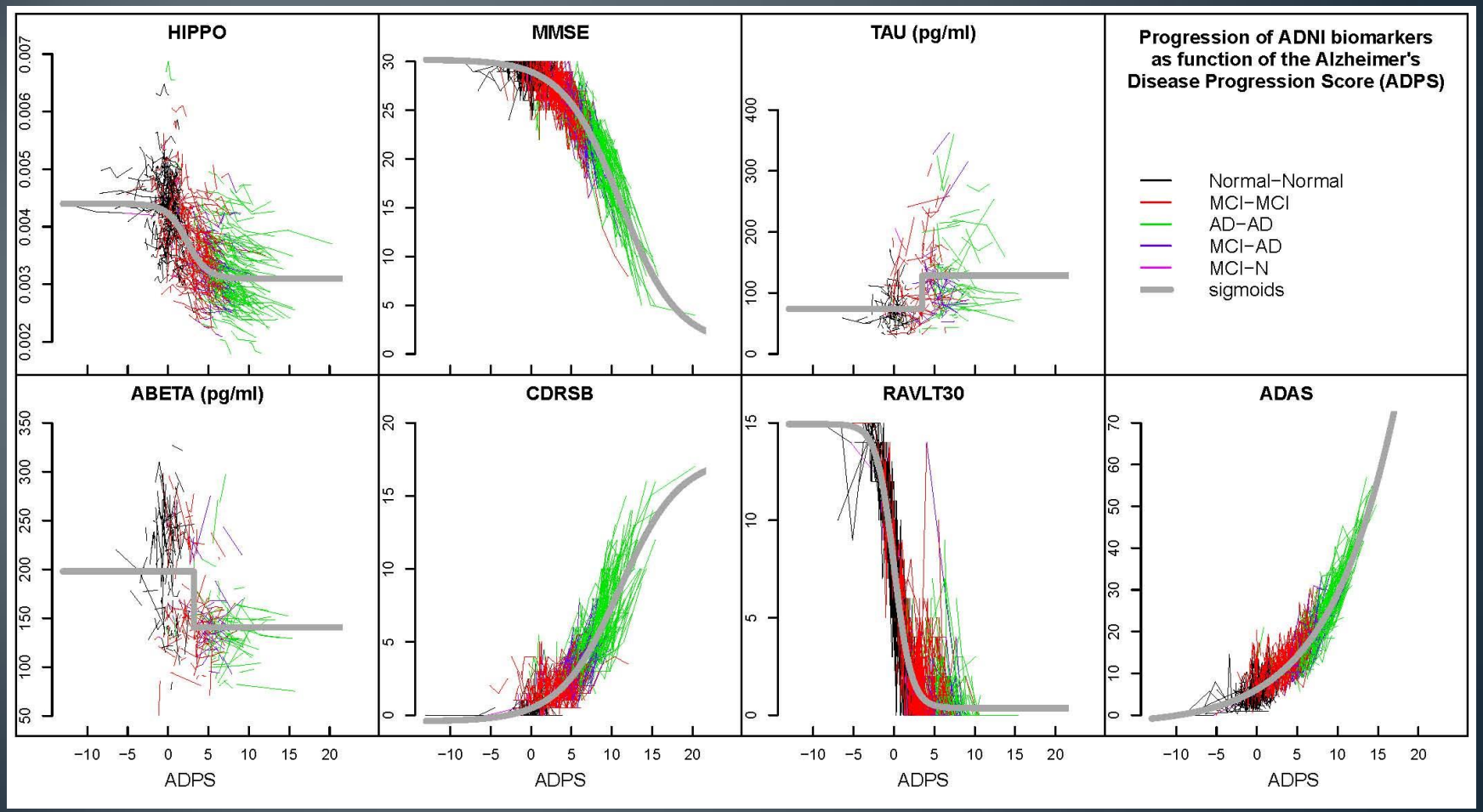
Algorithm for building an Alzheimer's Disease Progression Score (ADPS) for each subject in the ADNI

1. Correct each biomarker for the age effect. Initialize the ADPS of each subject with its age.
2. Repeat
   A. Fit a sigmoid (=4 parameters) to each biomarker, fixing the ADPS for each subject.
   B. Fit the ADPS of each subject (=2 parameters) fixing the sigmoid for each biomarker.
   C. Fit the variance of the noise (=1 parameter) for each biomarker.
3. Standardize the ADPS of all subjects, such that the median of the normal subjects is 0 and the median absolute deviation (mad) of the normal subjects is 1

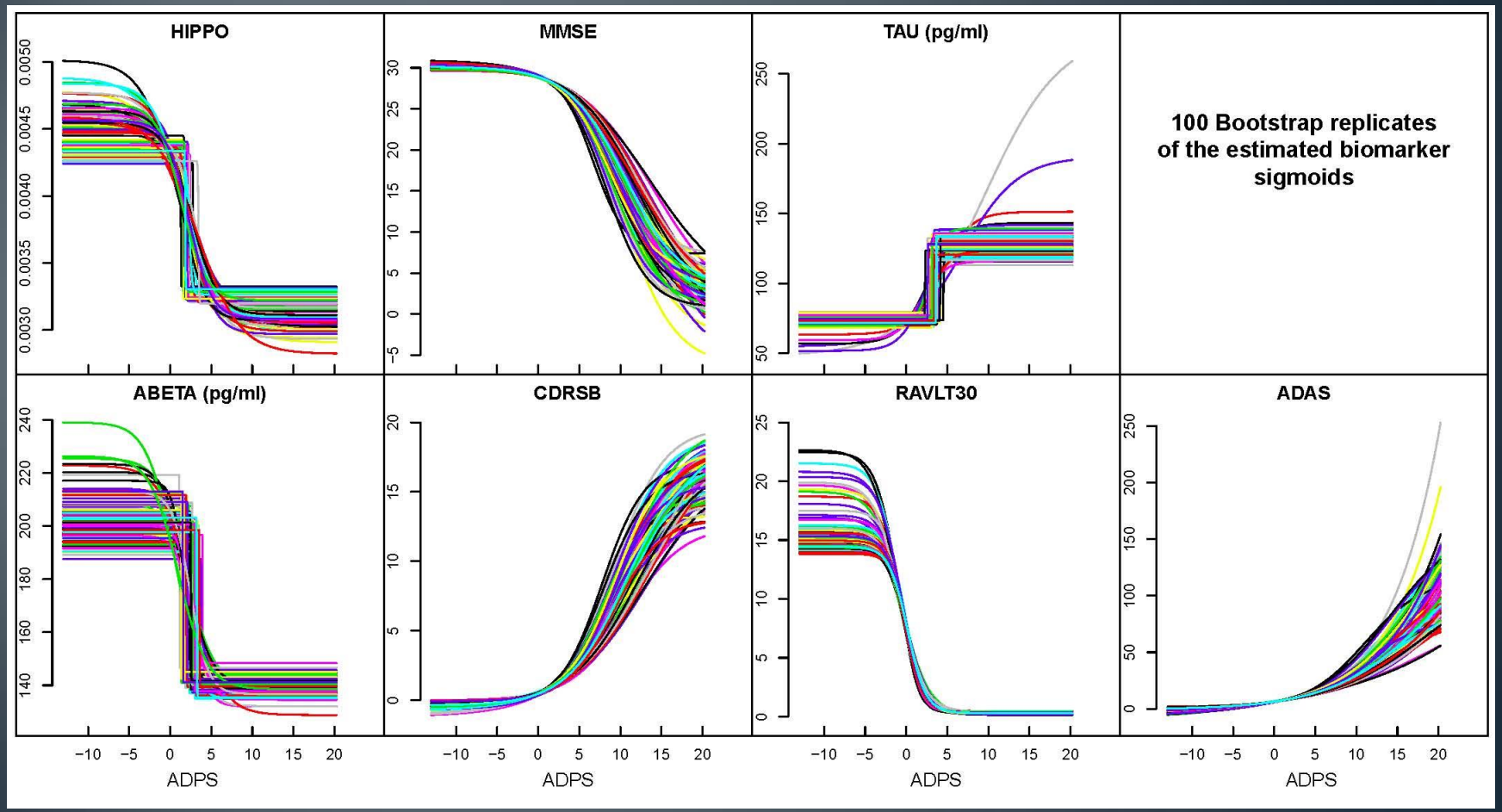
ADNI I dataset

1. 687 Subjects have MRI volumetric data and 2 to 6 visits (5 in average)
2. Expert selected biomarkers:
   A. Dementia ratings: Alzheimer's Disease Assessment Scale (ADAS), Mini Mental State Examination (MMSE), Clinical Dementia Rating Sum of Boxes (CDRSB)
   B. CSF measurements: proteins: Aβ_{42}, tau
   C. MRI measurement: Hippocampus volume over intra-cranial volume (Hippo)
   D. Memory rating: Rey Auditory Verbal Learning Test, 30 min (RAVLT_30min)

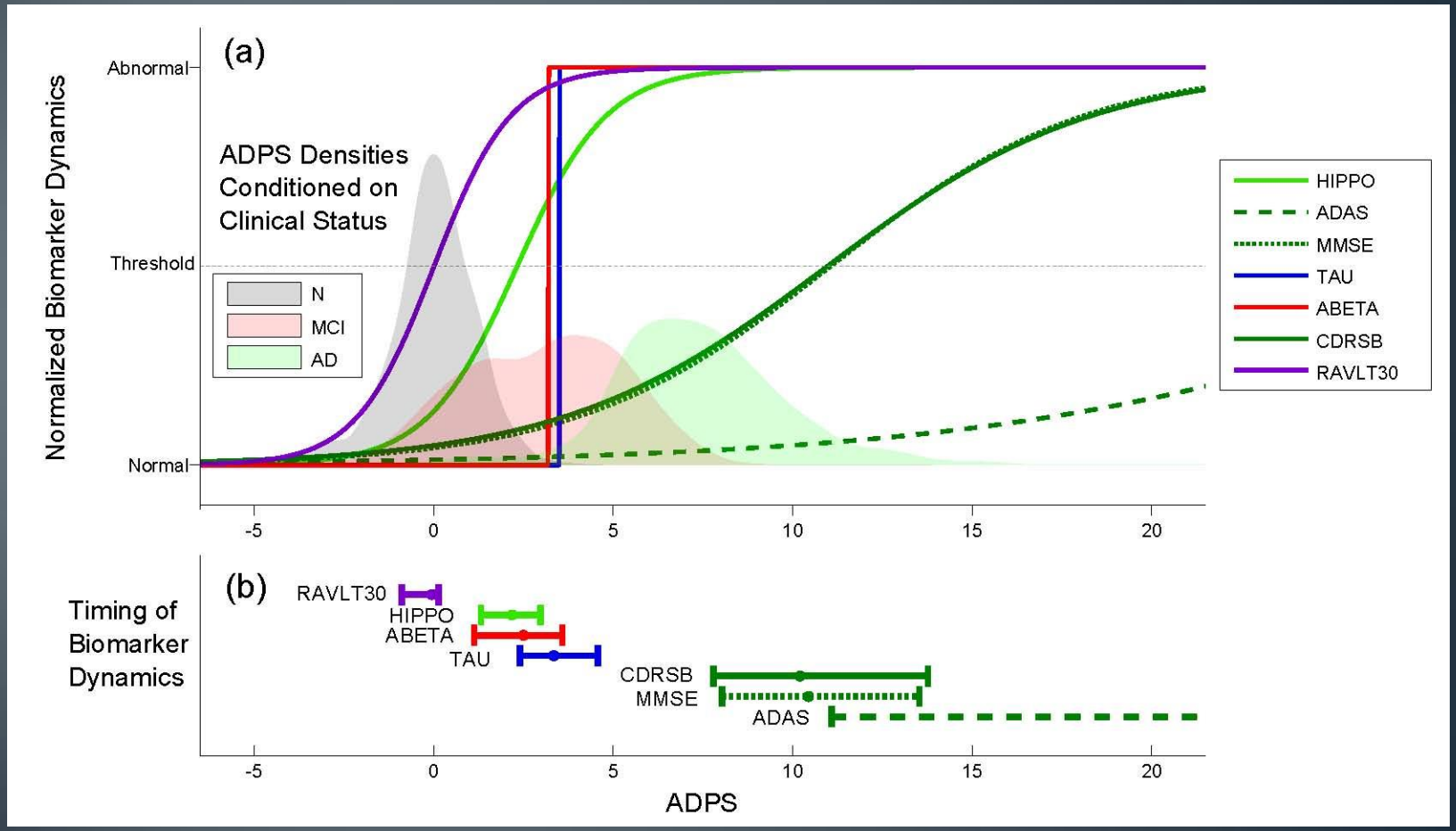
Biomarkers as function of the Alzheimer's disease Progression Score



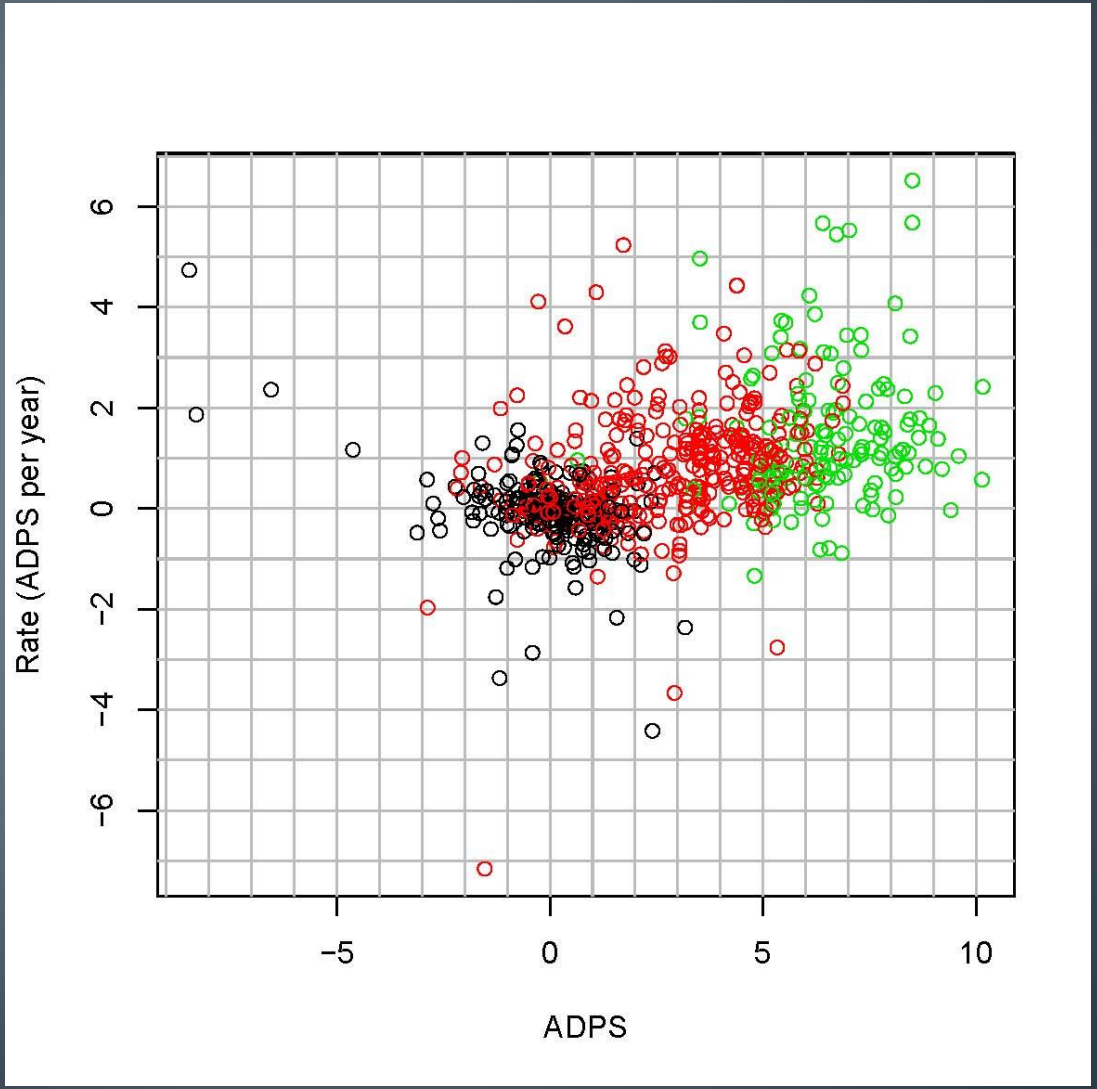
Bootstrap analysis of the variability of the sigmoids



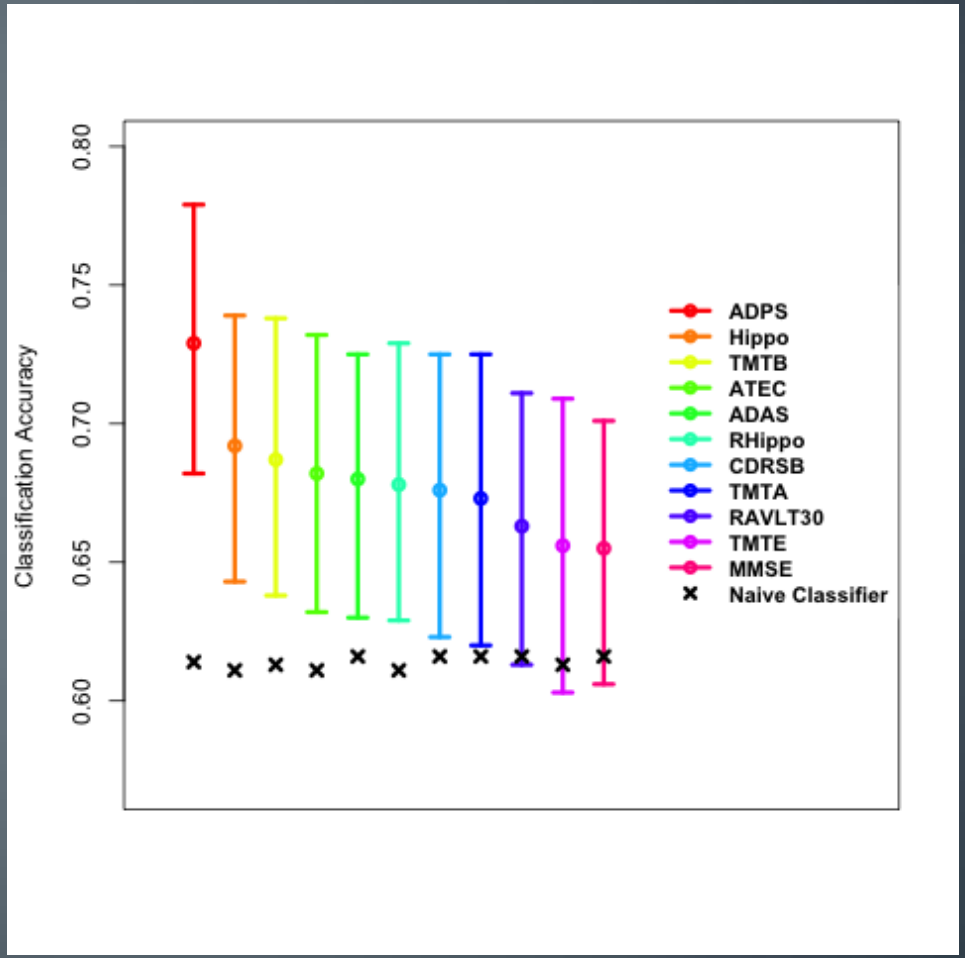
Standardized biomarkers



Rate of change of the ADPS as function of the ADPS



Prediction of conversion from MCI to AD in 2 years



Cognitive tests used in WRAP analyses

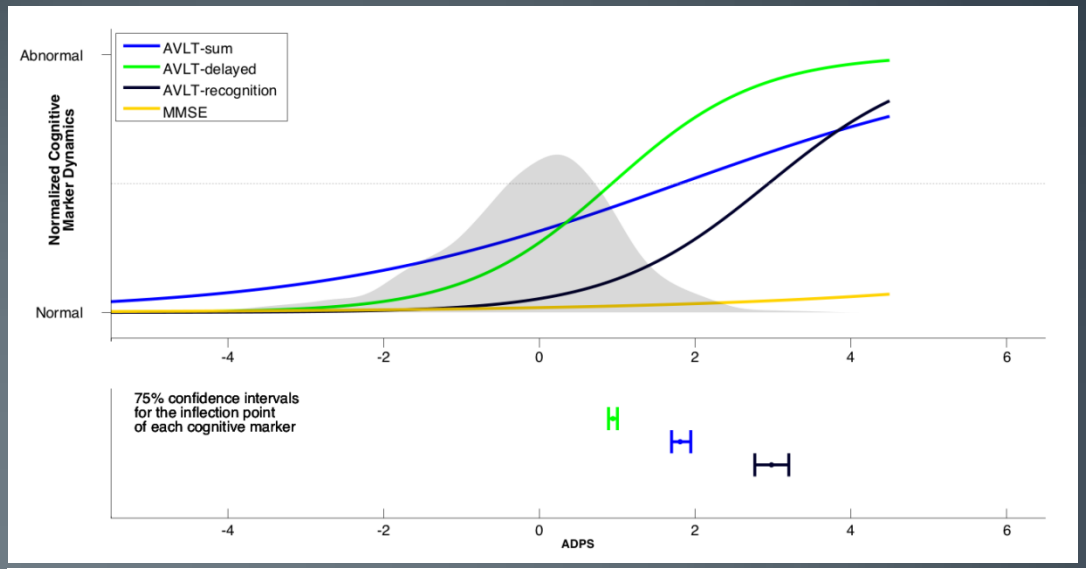
WRAP

- AVLT-sum
  - Auditory verbal learning test
  - Immediate recall/total learning score summed across 5 trials
  - Range 0-75
- AVLT-delayed
  - Delayed free recall score (~20 mins delay)
  - Range 0-15
- AVLT-recognition
  - Recognition of list words in a paragraph
  - Range 0-15
- MMSE
  - Mini-mental state exam
  - Range 0-30

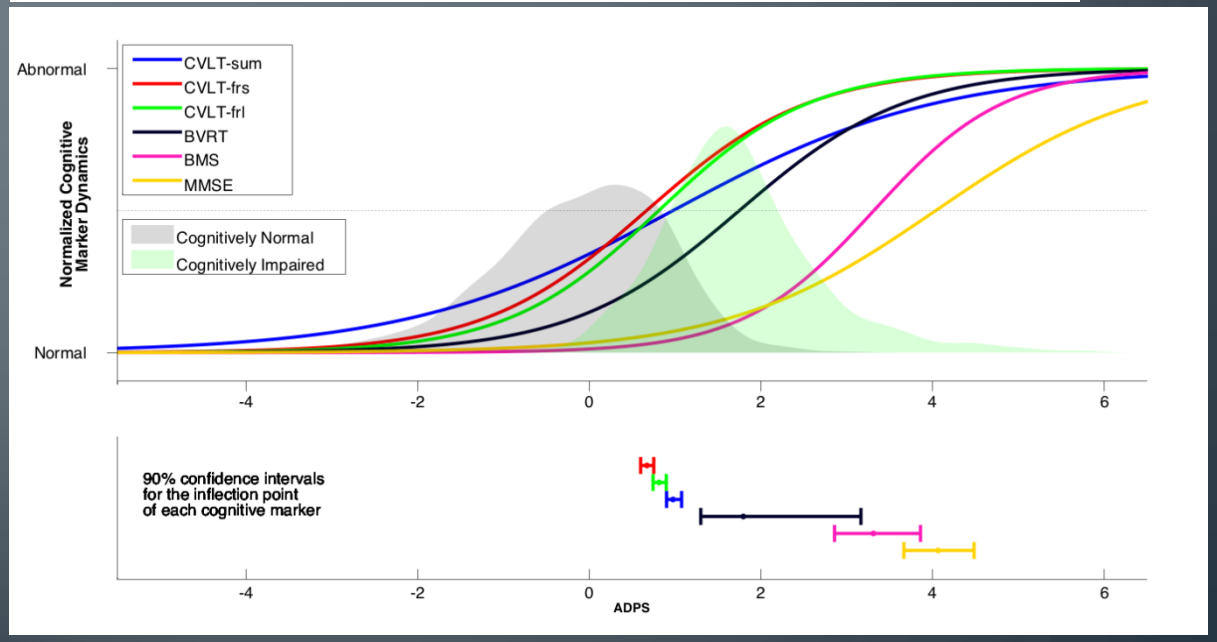
BLSA

- CVLT-sum
  - California verbal learning test
  - Immediate recall/total learning score summed across 5 trials
  - Range 0-80
- CVLT-frs and frl
  - Delayed free recall scores with short delay (after List B) and long delay (~20 mins later)
  - Range 0-16
- Benton visual retention
  - Scored for errors in drawing replication
- BMS
  - Blessed information memory concentration score
  - Scored for errors
- MMSE
  - Mini-mental state exam
  - Range 0-30

WRAP


Abnormal
Normal
Normalized Cognitive Marker Dynamics
AVLT-sum
AVLT-delayed
AVLT-recognition
MMSE
-4
-2
0
2
4
6
75% confidence intervals for the inflection point of each cognitive marker
ADPS

BLSA


Abnormal
Normal
Normalized Cognitive Marker Dynamics
CVLT-sum
CVLT-frs
CVLT-frl
BVRT
BMS
MMSE
Cognitively Normal
Cognitively Impaired
-4
-2
0
2
4
6
90% confidence intervals for the inflection point of each cognitive marker
ADPS

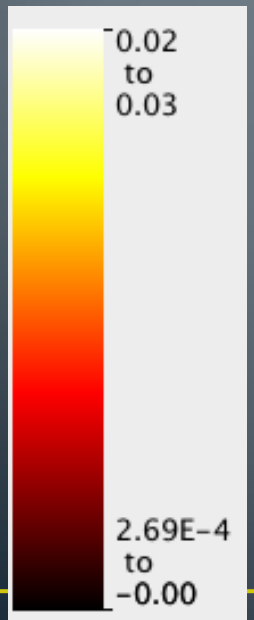
18 subjects
48 visits total

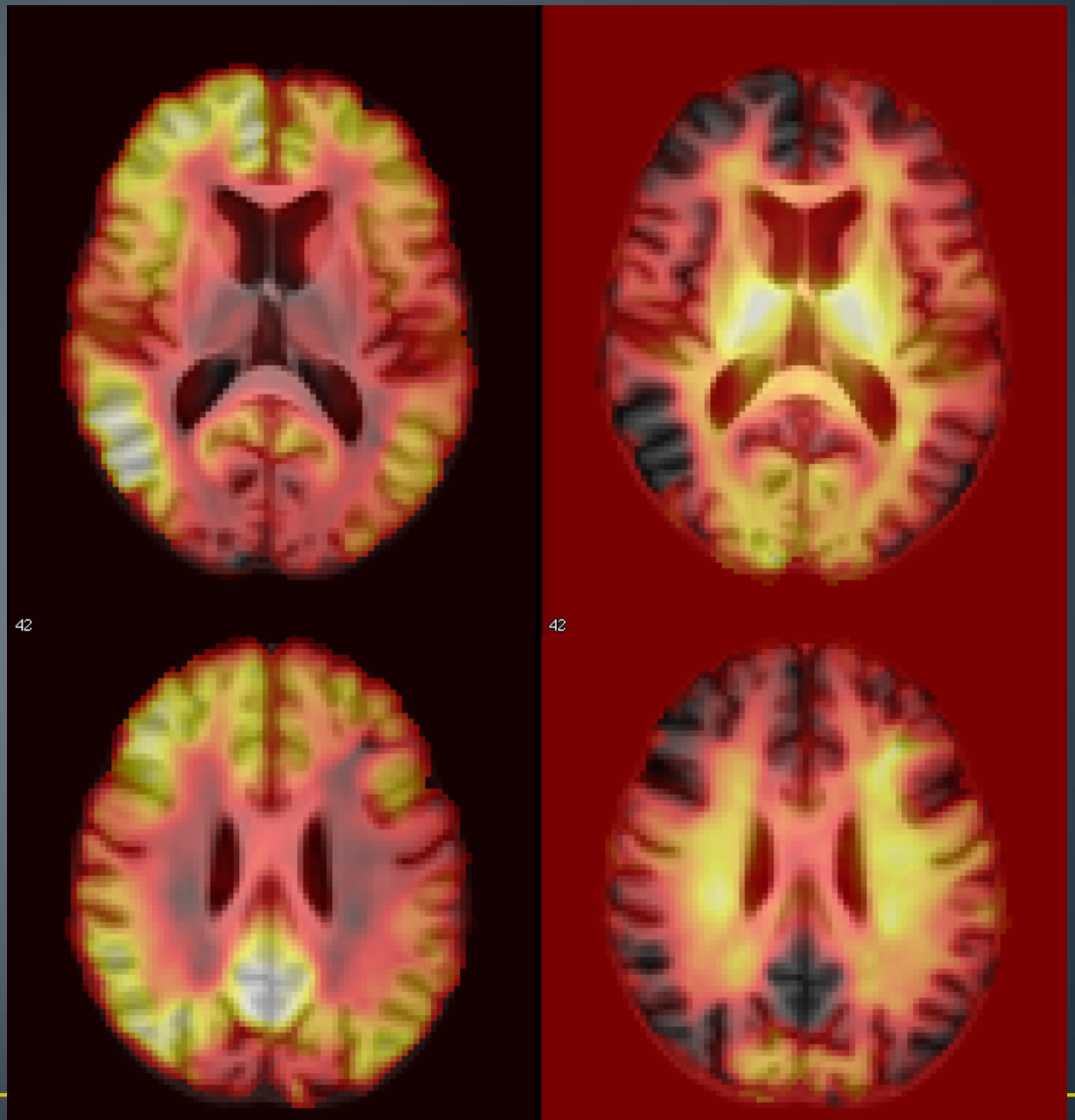
2 x 2 x 2 mm PiB-DVR images (amyloid)

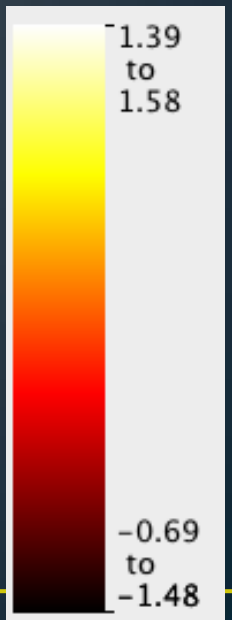
231k brain voxels

Estimates for A_k

Estimates for B_k







Some observations

- $A \approx 0$ in cerebellum.
  - Supports choice of reference tissue
  - Validity of longitudinal DVR estimates
- Sparing of pre- and postcentral gyri.
  - Literature indicates these regions accumulate amyloid in late stages.
- Precuneus, frontal, lateral parietal & temporal.
  - Usual suspects

Estimates for A_k

Conclusion

We have explored a statistical modeling technique aimed at better understanding the progression of AD.

We have experimented with ADNI, BLSA, and WRAP.

Thank you

References

1. Bruno Jedynak, Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T. Wyman, David Raunig, Pierre C. Jedynak, Brian Cao and Jerry Prince for the Alzheimer's Disease Neuroimaging Initiative,"A Computational Neurodegenerative Disease Progression Score: Method and Results with the Alzheimer's Disease Neuroimaging Initiative Cohort", NeuroImage 2012 Nov 15;63(3):1478-86.
2. Bruno M. Jedynak, Bo Liu, Andrew Lang, Yulia Gel and Jerry L. Prince, "A computational method for computing an Alzheimer's Disease Progression Score; experiments and validation with the ADNI dataset", accepted for publication in Neurobiology of Aging.
3. M. Bilgel, Y. An, A. Lang, J. Prince, L. Ferrucci, B. Jedynak, S. M. Resnick, "Trajectories of Alzheimer disease-related cognitive measures in a longitudinal sample", Alzheimer's & Dementia, In press.

Compositional models of disease progression for all subjects, including patients and control

- $y_{ij} = f\big(g(t_{ij}, c_{ij}, v_{ij})\big) + \epsilon_{ij}$
- i: subject
- j: visit index
- y_{ij}: collection of measurements (features, markers, biomarkers) available
- t_{ij}: age of subject i at visit j
- v_{ij}: treatment of subject i at visit j
- $(t, c, v) \mapsto g(t, c, v) \in \mathbb{R}^k$
- k: the intrinsic dimensionality of the "disease space"
- $f: \mathbb{R}^k \mapsto \mathbb{R}^m$: dynamic of the measurements
- $c_{ij} \in \mathbb{R}^l$: vital statistics of subject i. Might include weight, height, intra-cranial volume, …
- ϵ_{ij}: centered noise.

Identifiability

$$\varrho = \{\rho = (a, b, \alpha, \beta, \sigma); I^{-1}\sum_{i=1}^{I}\alpha_i = \alpha_0, I^{-1}\sum_{i=1}^{I}\beta_i = \beta_0, \\ b_k > 0, a_k \neq 0 \text{ for all } k \in \mathcal{I}\}$$

Theorem 1. *The model $\{P_\rho; \rho \in \varrho\}$ is identifiable as long as the following 2 conditions are verified:*

1. *For each biomarker, there is at least 1 subject i with $\alpha_i \neq 0$ and with at least 4 distinct time-points at which this biomarker is available.*
2. *For each subject, there is at least 1 biomarker which is available at 2 time points.*

Hypothetical progression of Parkinson