“Functional Data Analysis: Novel Statistical Methods and Applications in Medical Research”

Sheng Luo

Professor

Department of Biostatistics & Bioinformatics

Duke University

Financial Disclosures for the past 12 months

- Salary: Duke University.
- Grants/Research: National Institutes of Health, CHDI Foundation, International Parkinson and Movement Disorder Society, Parkinson's Foundation.
- Consulting and Advisory Board Membership with honoraria: NIH Study Sections, CHDI Management, Inc., T3D Therapeutics, Kashiv BioSciences, MyMee Inc, GuidePoint.

Outline

- Introduction of Functional Data Analysis (FDA)
- Part I: Novel analytic approaches to investigate minute-level actigraphy and association with physical function
- Part II: Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease
- Conclusion and remarks

Some common statistical regression methods

- Logistic regression: binary outcome
- Cox regression: time-to-event outcome
- Poisson regression: event counts as outcome

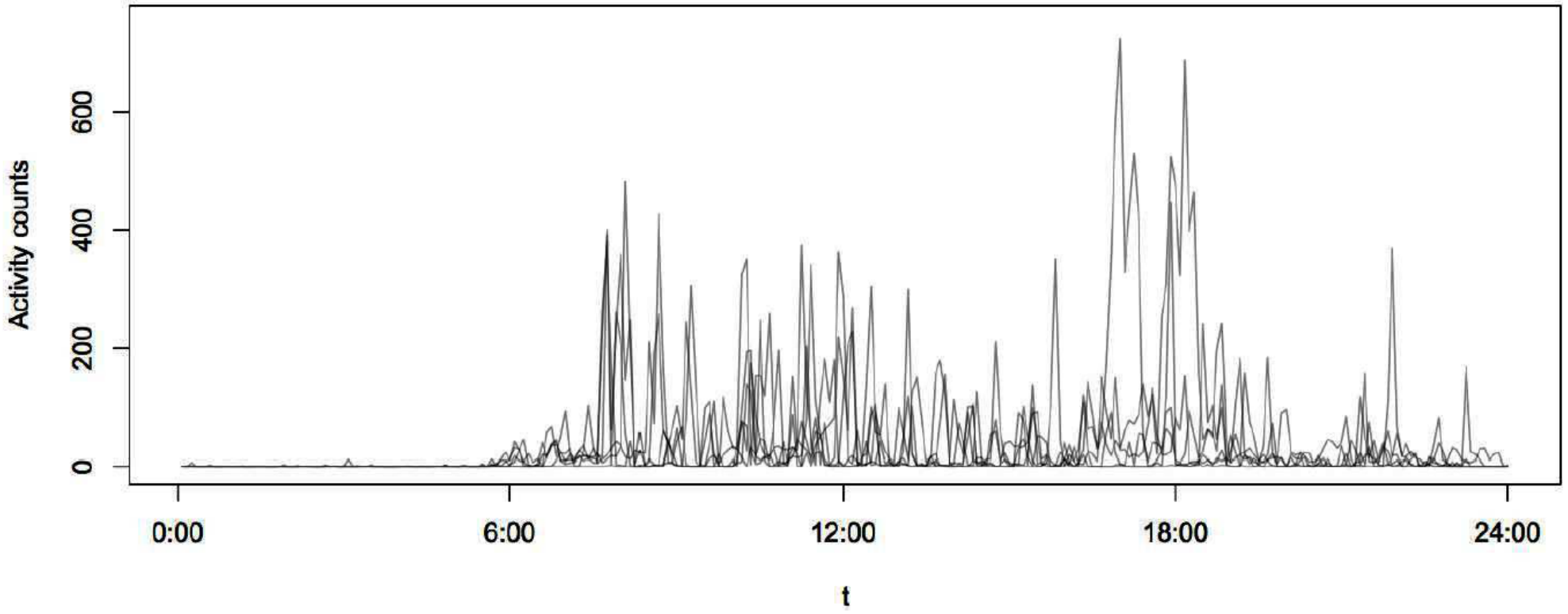
What if either the outcome or covariate is a function, or both?

Functional Data

- Functional Data: data for which units of observation are functions
- These functions can be curves (1D), images (2D or 3D), or higher dimension object data (e.g., structure or functional MRI).

Examples of Functional Data

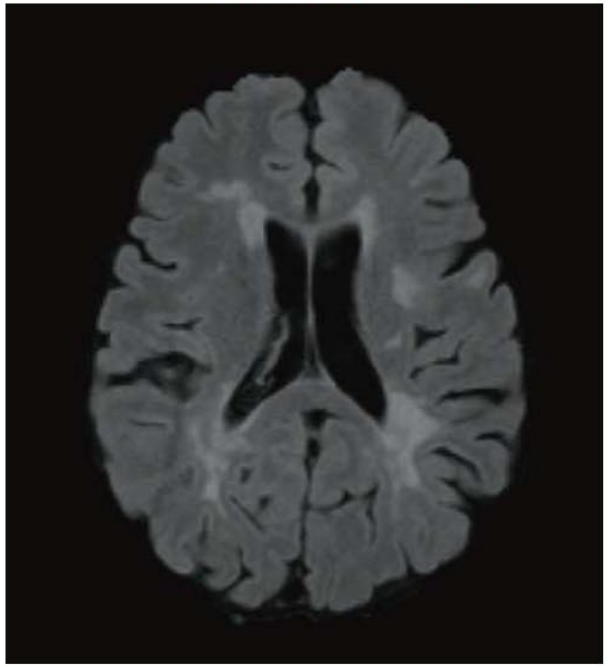
- Physical activity information



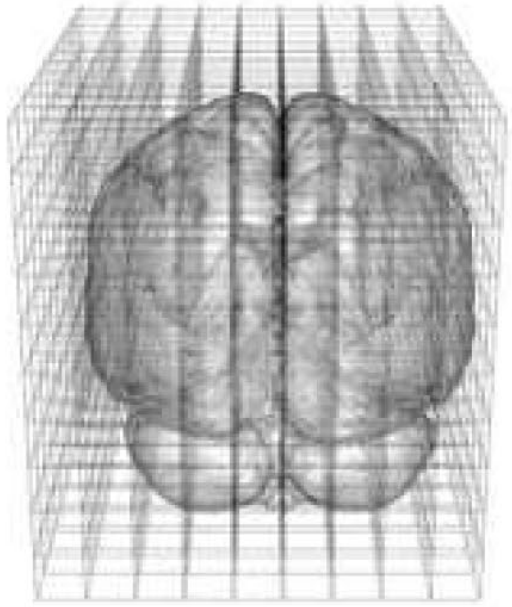
Examples of Functional Data

- Brain imaging

A slice of Magnetic Resonance Imaging (MRI)



Voxel-based whole-brain image



- The analysis of functional data is termed "Functional Data Analysis" (FDA)

Functional Regression

Regression analysis involving functional data.

1. Functional predictor regression (scalar-on-function)
   Ex: how is the minute-level actigraphy activity associated with the physical function?

2. Functional response regression (function-on-scalar)
   Ex: how do sex and age change the minute-level actigraphy activity?

3. Function-on-function regression (function-on-function)
   Ex: how is the minute-level actigraphy activity associated with the MRI data?

Part I: Novel analytic approaches to investigate minute-level actigraphy and associations with physical function

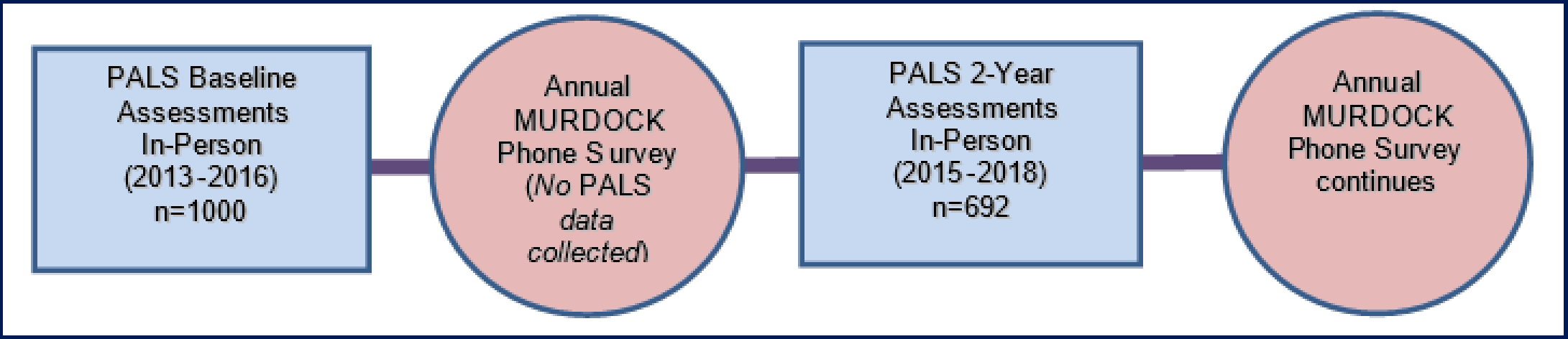
Motivation

- Low levels of physical activity and declined physical function have implications for dementia risk, premature disability in older adults.
- Accelerometers provide objective and convenient measurement of physical activity.
- Previous studies examined the associations between accelerometry-derived physical activity and physical function, but they reduced data into average means of total daily physical activity (e.g., daily step counts).
- We used FDA methods to investigate the association between physical activity and physical function.

Physical performance Across the Lifespan Study (PALS)

- Longitudinal cohort study of community-dwelling adults aged 30-90+ residing in southwest region of North Carolina.

- Participants completed an extensive functional battery and wore an accelerometer as a measure of activity for 7 days. Assessments were completed at baseline and again 2 years later with 69% retention rate.



Outcomes

1. Gait Speed (m/sec): measures how quickly someone can walk within a specified distance (i.e., 4 meters) in normal pace and rapid pace.
2. Single Leg Stance (sec): measures the time participants are able to stand unassisted on one leg with eyes open.
3. Chair Stands in 30 seconds (n): measures lower extremity strength. The score is the number of completed stands in 30 seconds.
4. 6-minute Walk (feet): the total distance walked in 6 minutes as a measure of aerobic endurance and capacity.

Data Summary at Baseline (n=669)

Table 1. Data Summary at Baseline

	Statistics*
Age (year)	66.0 (±16.1)
Gender (Male)	284 (42.5%)
Race (White)	593 (88.6%)
BMI	27.1 (±4.8)
Gait Speed: rapid pace (m/sec)	1.81 (±0.47)
Gait Speed: normal pace (m/sec)	1.23 (±0.28)
Single Leg Stance (sec)	32.6 (±23.5)
Chair Stands in 30 secs (n)	15.0 (±6.2)
6-minute Walk (feet)	1741.86 (±463.78)

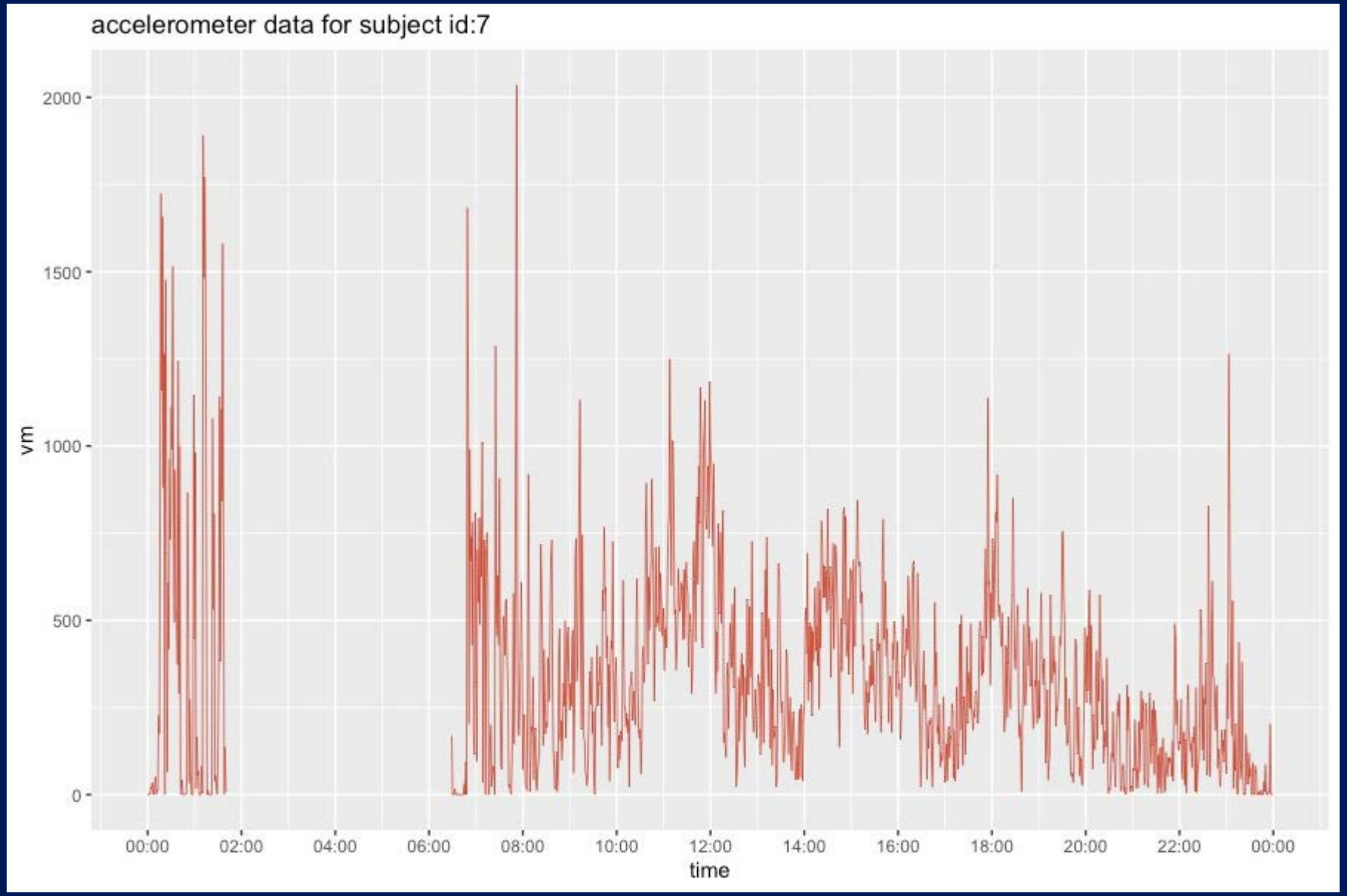
*Statistics given by means SDs, N (percent).

Vector Magnitude Data (activity counts)

date	epoch	axis1	axis2	axis3	vm
2/5/2014	9:11:00 AM	516	330	468	771
2/5/2014	9:12:00 AM	229	236	572	660
2/5/2014	9:13:00 AM	2499	2036	2973	4385
2/5/2014	9:14:00 AM	359	784	1065	1370
2/5/2014	9:15:00 AM	34	4	164	168
2/5/2014	9:16:00 AM	91	361	723	813
2/5/2014	9:17:00 AM	63	679	759	1020
2/5/2014	9:18:00 AM	12	250	23	251
2/5/2014	9:19:00 AM	0	170	0	170
2/5/2014	9:20:00 AM	0	98	0	98
2/5/2014	9:21:00 AM	0	20	0	20
2/5/2014	9:22:00 AM	0	38	0	38
2/5/2014	9:23:00 AM	0	235	23	236
2/5/2014	9:24:00 AM	0	145	0	145

Subject ID: 7

Vector Magnitude Data (activity counts)

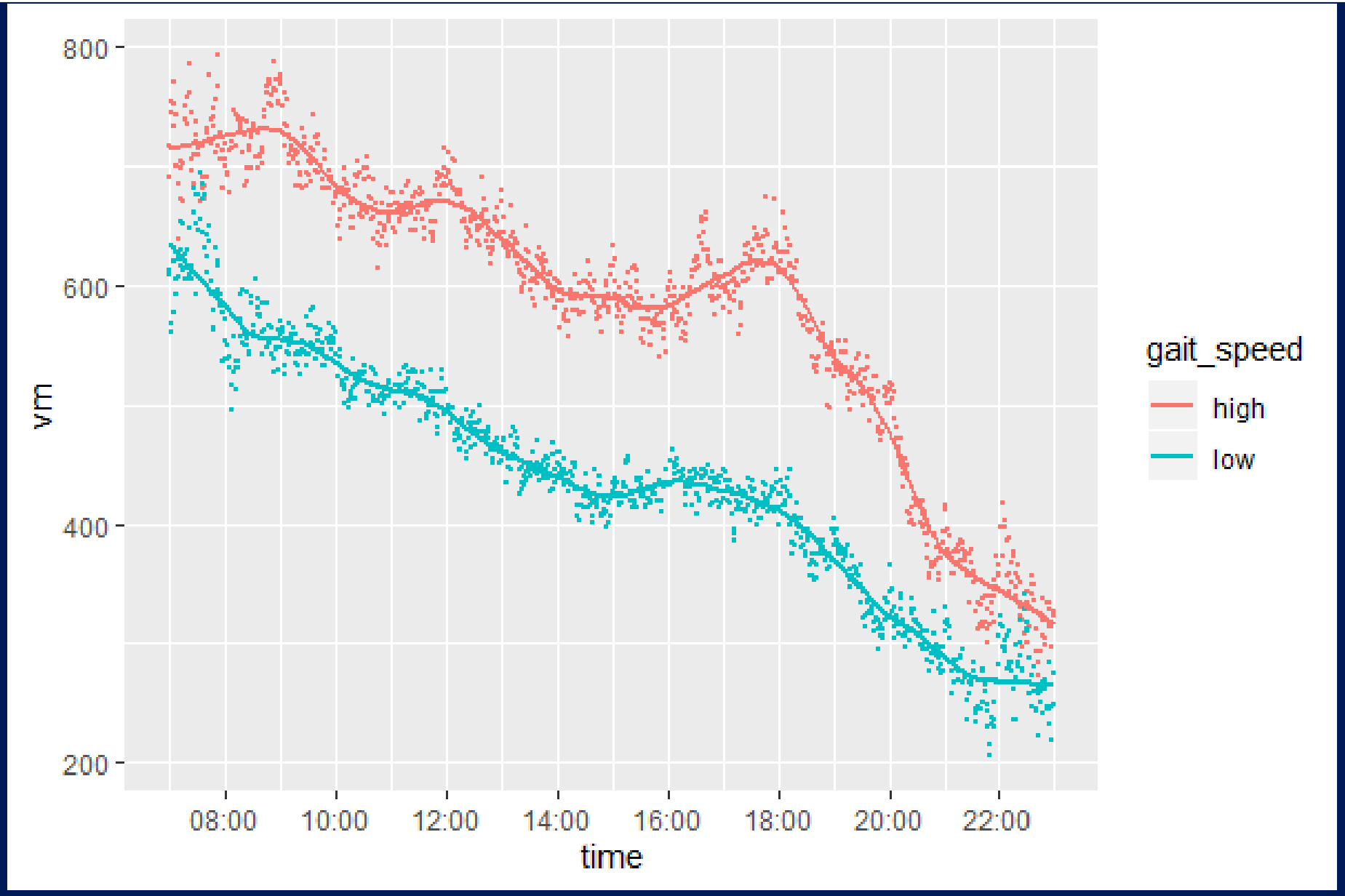


Subject ID: 7

Objectives

Aim 1: Investigate the functional associations between physical activity features and physical functions (gait speed, single leg stance, chair stands, and 6-minute walk test) at baseline.

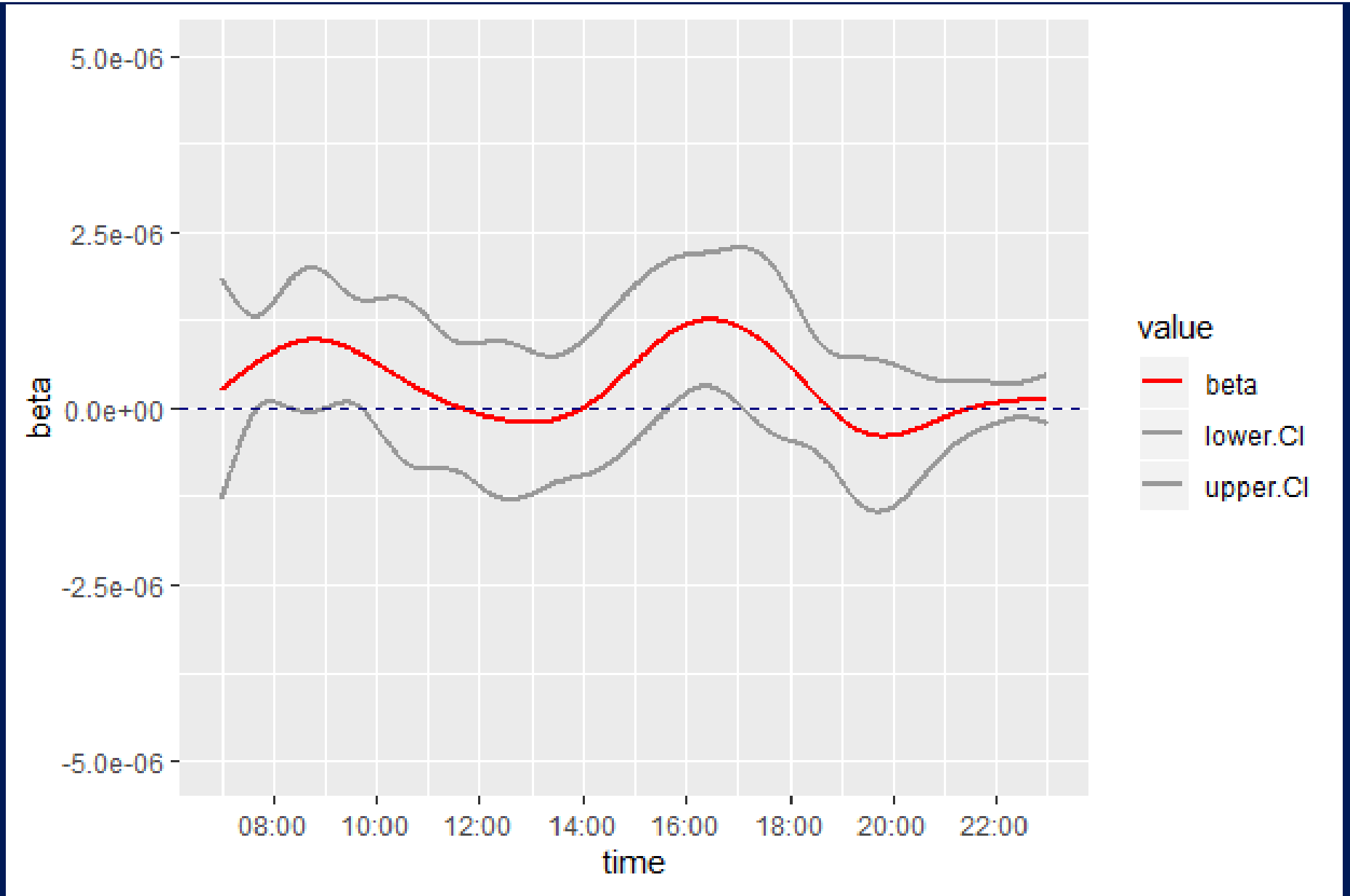
Lowess Curves for VM by High/Low Rapid Pace



Functional Regression for Baseline Rapid Pace

Coefficient	Estimate	SE	t value	Pr(>|t|)
Intercept	2.976	0.131	22.791	< 2e-16
Male Sex*	0.172	0.030	5.746	1.4e-08
Age*	-0.015	0.001	-13.311	< 2e-16
BMI*	-0.013	0.003	-3.885	1.13e-04
White Race	0.068	0.048	1.438	0.151

Estimated Coefficient Function for Baseline Rapid Pace



Findings of Aim 1

Increased physical activity at specific times of day was associated with increased physical functions

1. Rapid gait speed: 8AM, 9:30AM, 2:30-5PM
2. Normal gait speed: 9-10:30AM, 3-4:30PM
3. Single leg stand: 9-10:30AM
4. Chair stand: 9:30-11:30AM, 3-6PM
5. 6-min walk: 3-6:30PM

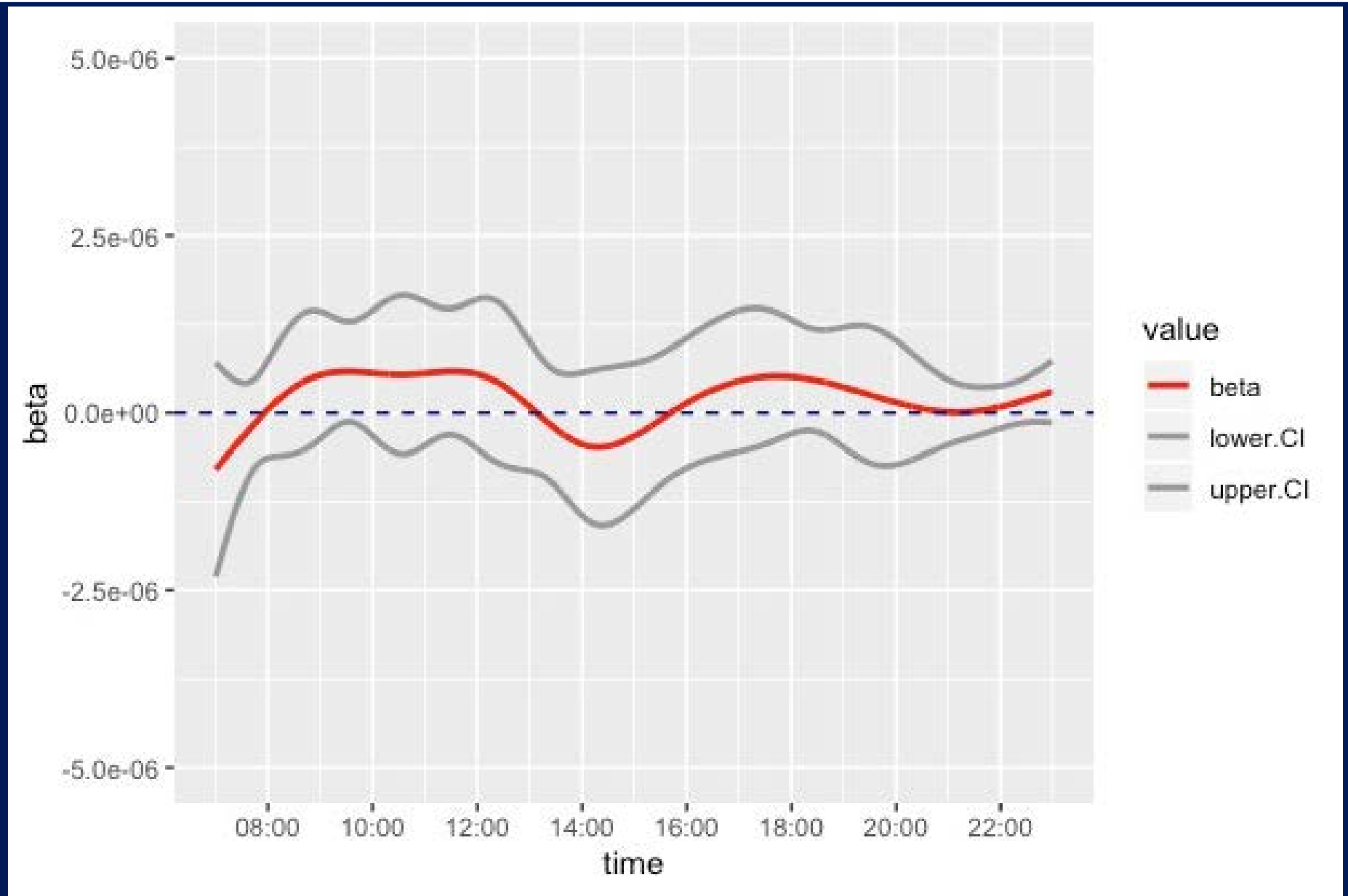
Objectives

Aim 2: Investigate the functional associations between the baseline physical activity features and the physical function at two years.

Functional Regression for Rapid Pace after 2 Years

Coefficient	Estimate	SE	t value	Pr(>|t|)
(Intercept)	1.013	0.181	5.611	3.50e-08
Baseline RP*	0.673	0.042	15.969	< 2.00e-16
Male Sex	0.046	0.030	1.511	0.13
Age*	-0.006	0.001	-4.560	6.58e-06
BMI	-0.006	0.003	-1.711	0.09
White Race	0.056	0.052	1.070	0.29

Estimated Coefficient Function for Rapid Pace Change



Findings of Aim 2

No significant association between baseline physical activity and physical functions after 2 years.

Conclusion

Functional data analysis (FDA) provides new insight into the relationship between minute-by-minute daily activity and physical functions.

Part II: Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease

Alzheimer's Disease (AD)

- A neurodegenerative disorder of the brain and No. 1 leading cause of dementia.
- No disease-modifying treatments for AD.
- The most expensive disease in America.
- In 2018, 5.8 million American with AD and $277 billion in payment (1.35% of 2018 GDP!).
- The number of Americans with AD will reach 7.7 million by 2030 and the corresponding total cost of care for AD will increase to $1.08 trillion each year.

NIH All of US Research Program

NIH National Institutes of Health
All of Us Research Program

ABOUT FUNDING NEWS, EVENTS, & MEDIA JoinAllofUs.org Search

The future of health begins with you

The *All of Us* Research Program is a historic effort to gather data from one million or more people living in the United States to accelerate research and improve health. By taking into account individual differences in lifestyle, environment, and biology, researchers will uncover paths toward delivering precision medicine.

JOIN NOW

Key Scientific Questions

Quote from NIH All of US Research Program:

- Develop ways to measure risk for a range of diseases based on environmental exposures, genetic factors and interactions between the two
- Discover biological markers that signal increased or decreased risk of developing common diseases.

The tool is Personalized Risk Prediction!

Our Research Question

- Objective: To develop a prognostic model, based on multivariate longitudinal markers, for predicting progression-free survival in patients with mild cognitive impairment.

Predictive Models

- Most of the predictive models are static model, e.g., logistic regression, Cox model.
- Pros
  - Simple
  - Low computing cost
- Cons
  - Prediction can not be updated in a real-time fashion.

Dynamic Prediction

- What is dynamic prediction?
  - Predictions are conducted on a real-time basis so that the predictions can be updated with new data.
- Why is it important?
  1. Predict patients prognoses and make medical decisions in a real-time fashion.
  2. Answer important predictive questions:
     - For a particular person, what are the most likely outcome trajectories in the next 6 months?
     - What is the risk of developing AD?
  3. Enable personalized prevention, treatment, and care.

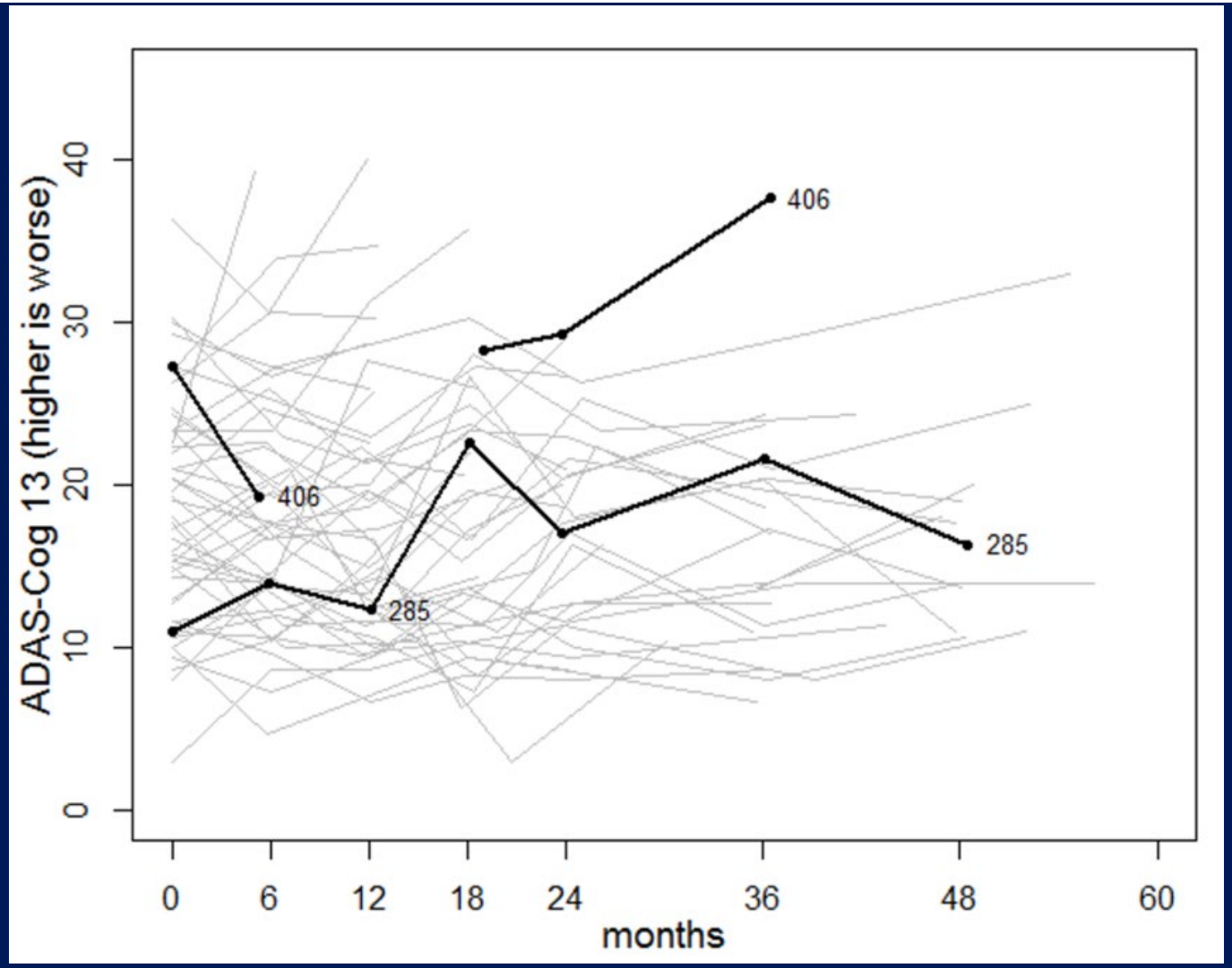
ADNI Study

- Alzheimer's Disease Neuroimaging Initiative (ADNI) study: a longitudinal observational study investigating whether serial brain imaging, clinical, and neuropsychological assessments can be combined to measure the progression of AD.
- Focus on 355 MCI patients who started from ADNI-1 and were reassessed at 6, 12, 18, 24, 36 months.
- 180 patients were diagnosed with AD (survival event) and 175 had stable MCI over a mean follow-up period of 2.3 years and 4.2 years, respectively.

Data Source: Longitudinal Markers

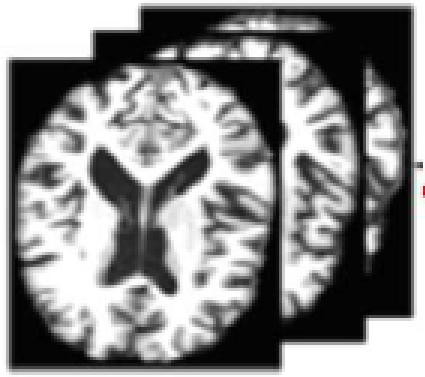
- Longitudinal **AD Assessment Scale-Cognitive** (ADAS-Cog) score and **Hippocampal volume** (HV) are the strongest predictors of AD conversion from MCI in neurocognitive and neuroimaging domain.
- Enormous information lost occurs when the high dimensional image data are reduced to a single volume.
- Surface-based morphology analysis retains more information about Hippocampus atrophy.
  - Hippocampal radial distance (HRD): the distance from the medial core of the hippocampus to points on the surface and quantifies the thickness of hippocampus relative to its center line.

Longitudinal ADAS-Cog



Longitudinal trajectories of ADAS-Cog 13: 50 MCI patients from the ADNI study

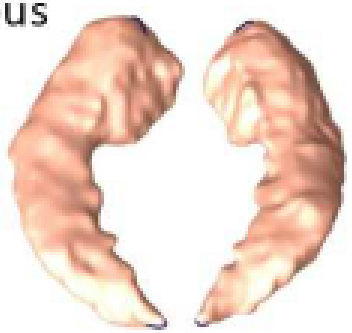
Hippocampus Image Processing



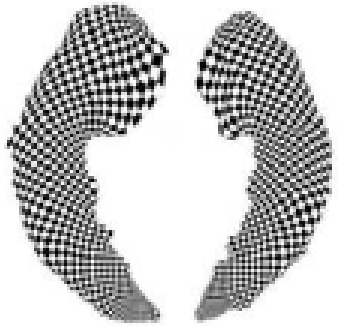
MRI data

Hippocampus surface extraction

Step 1



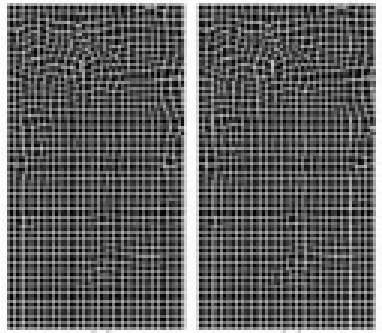
Surface-based subcortical



Surface meshes

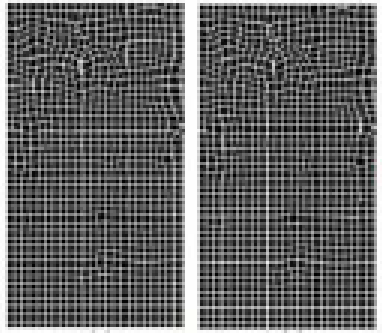
Surface conformal mapping and parameterization

Step 2



Feature images

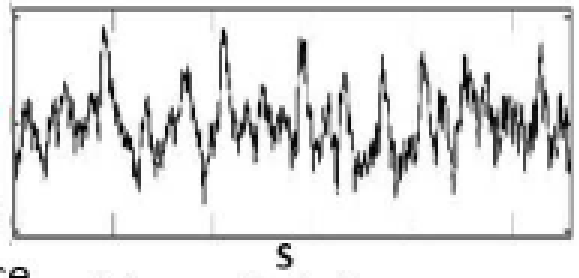
Step 3 Registration with a template



Compute HRD of each point using its coordinates

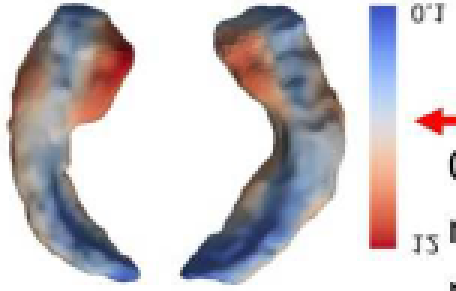
Align the points

Step 4



$g_i(s)$: radial distance

One-to-one mapping of the points to surface



Application to the ADNI Study

Functional Joint Model (FJM) structure

- Survival sub-model: time from first visit to AD diagnosis
- Longitudinal sub-model: ADAS-Cog 13
- The baseline hippocampal radial distance (*HRD*) as the functional predictor.
- Baseline hippocampal volume, age, gender, years of education and presence of the apolipoprotein E (*APOE*) ε4 allele as scalar covariates.

Model Comparison

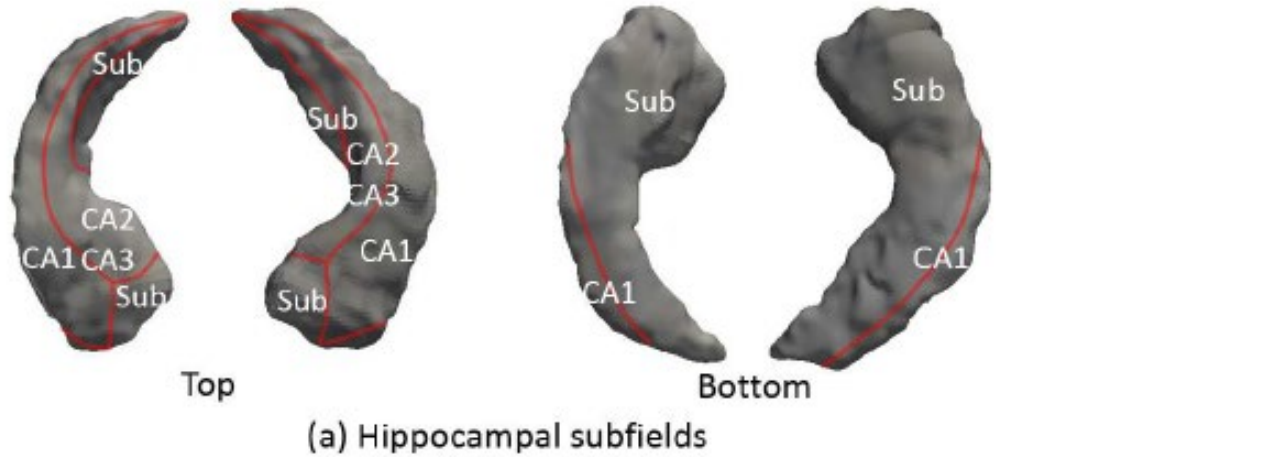
- Compare the two candidate models by time-dependent AUCs, at different time points over the follow-up period.

Δt	t	*JM*	*FJM*
6m	12m	0.768	0.772
	18m	0.710	0.752
	24m	0.781	0.844
12m	12m	0.761	0.781
	18m	0.744	0.792
	24m	0.712	0.770

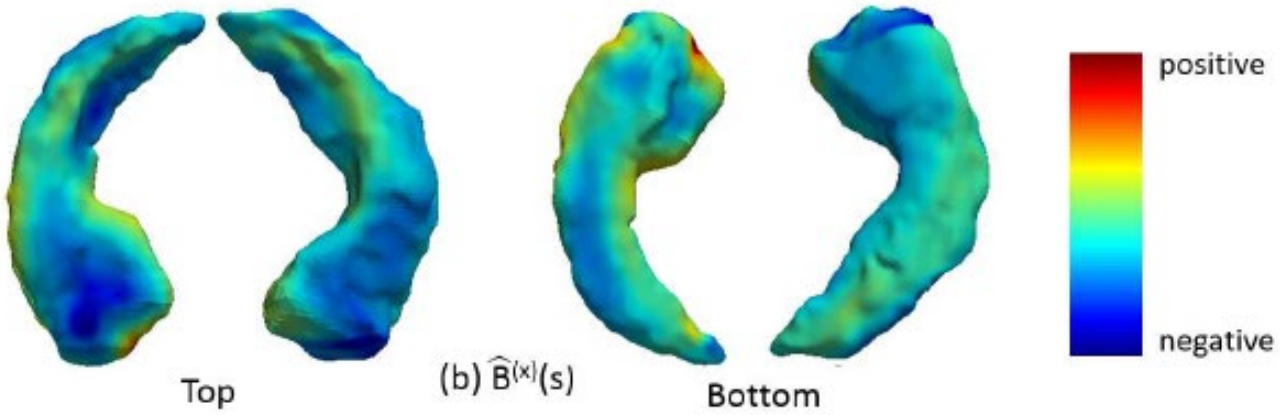
Parameter Estimation

- Parameter estimates from model *FJM* with HRD in both longitudinal and survival sub-models.

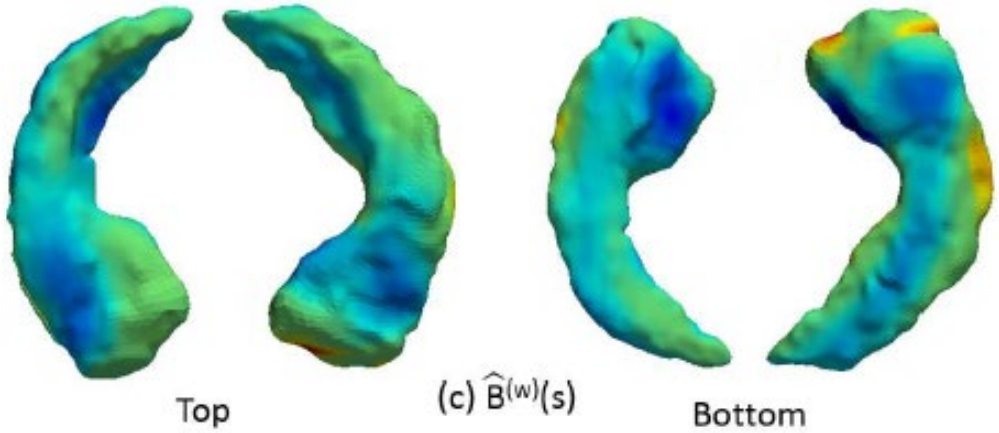
	Parameters	Mean	SE	2.5%	97.5%
For longitudinal outcome					
ADAS-Cog 13	Time (Years)	0.428	0.045	0.338	0.521
	bAge	−0.364	0.260	−0.885	0.156
	bHV (mm^3)	−1.617	0.295	−2.201	−1.051
For survival process					
MCI to AD	Female	−0.088	0.173	−0.397	0.270
	bAge	−0.283	0.042	−0.423	−0.109
	Edu (years)	0.028	0.016	−0.002	0.062
	APOE-ε4	0.533	0.125	0.239	0.728
	bHV (mm^3)	0.056	0.114	−0.185	0.276
	α	0.134	0.022	0.079	0.177



(a) Hippocampal subfields



(b) $\widehat{B}^{(x)}(s)$



(c) $\widehat{B}^{(w)}(s)$

Estimated coefficient functions for *HRD* in the sub-models are mapped back to the hippocampal surfaces.

Dynamic prediction for new patients using FJM

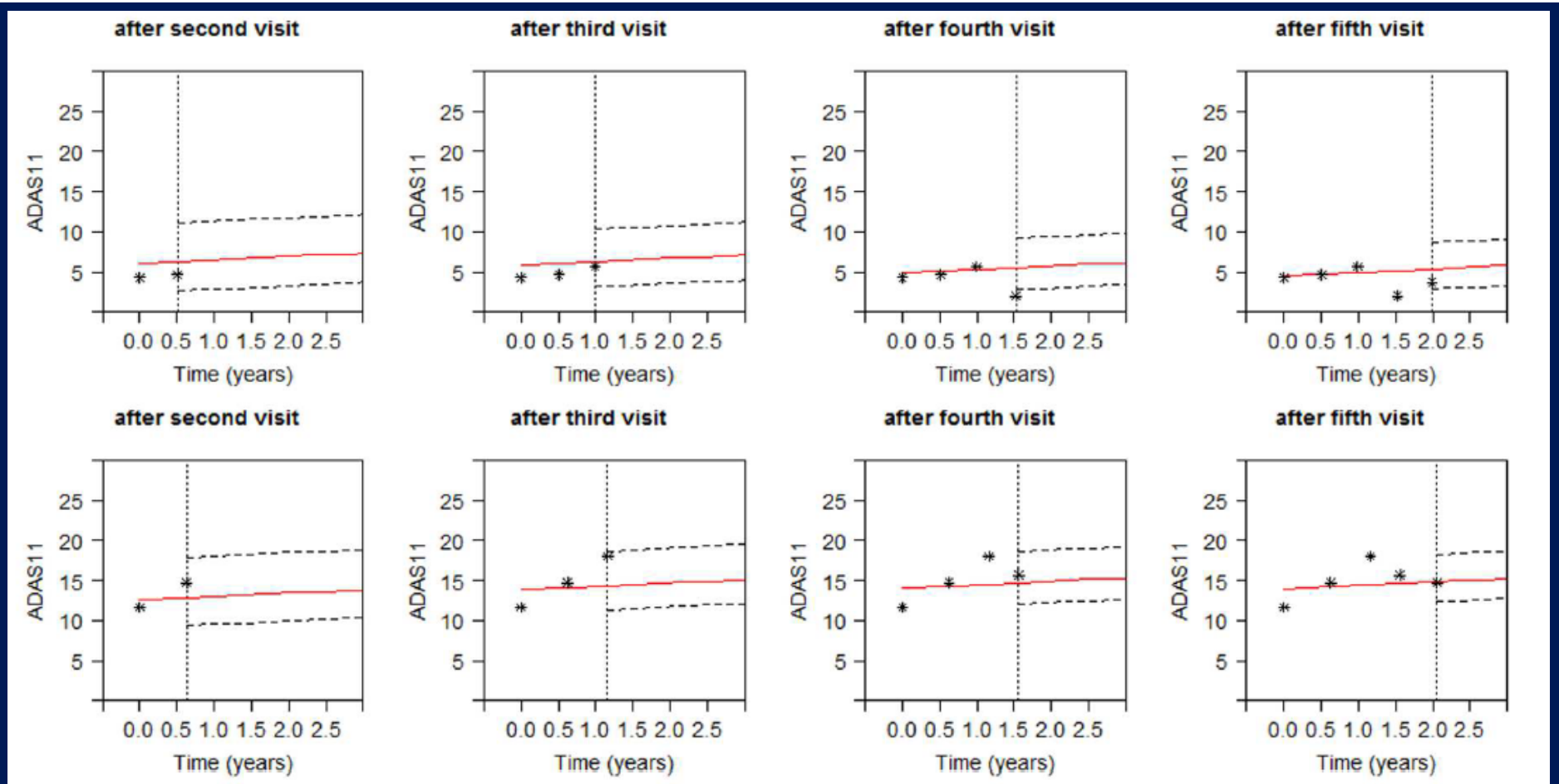


Figure: Predicted ADAS-Cog 11 for Patient A (upper panels) and Patient B (lower panels). Solid line is predicted longitudinal trajectories. Dashed lines construct a 95% pointwise uncertainty band. The dotted vertical line represents the time of prediction t.

Dynamic prediction for new patients using FJM

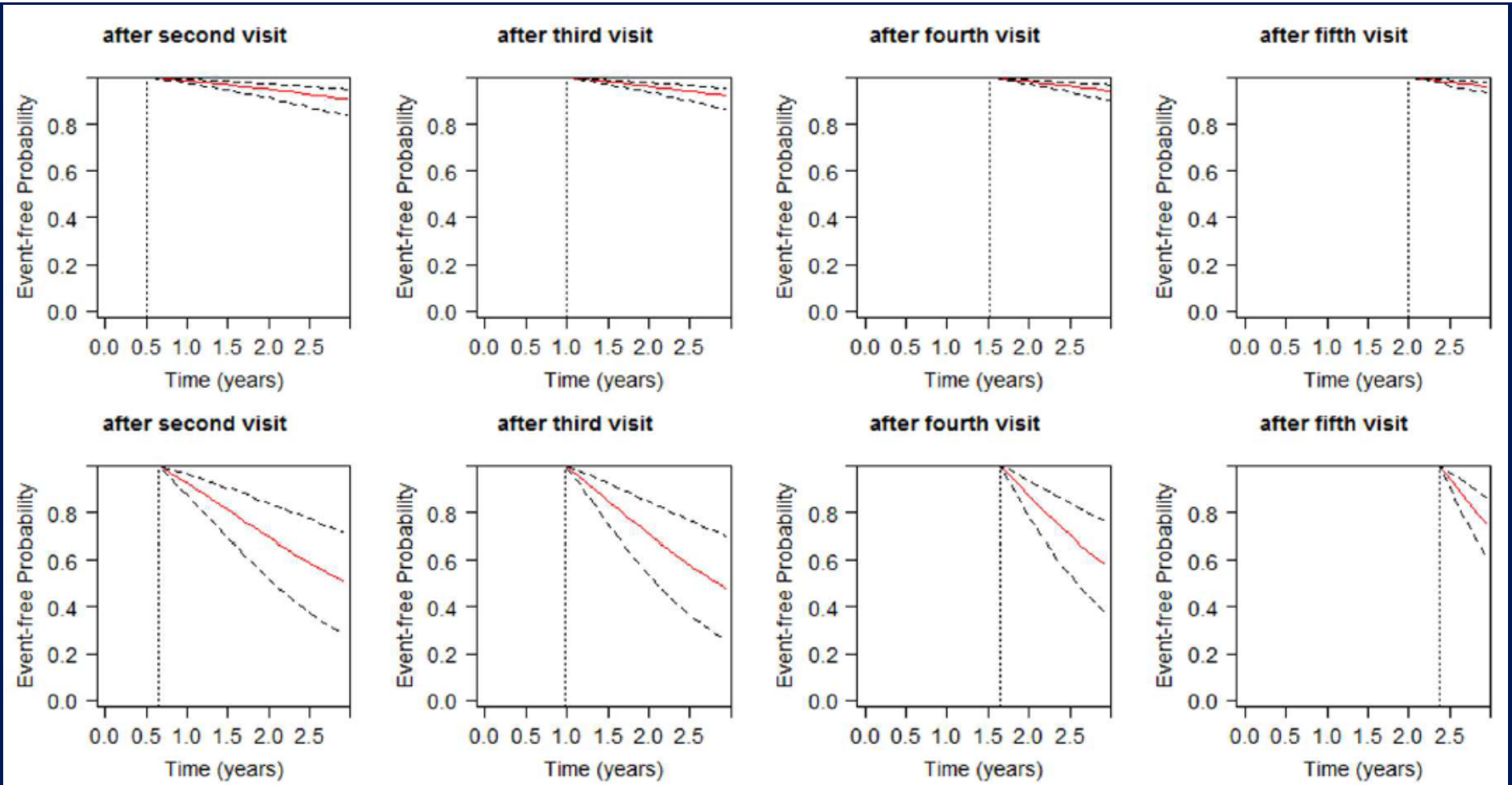


Figure: Predicted event-free probability with 95% pointwise uncertainty band for Patient A (upper panels) and Patient B (lower panels).

Conclusion

- Including baseline *HRD* as a functional predictor in the dynamic prediction framework can improve the predictive performance.
- Regional radial atrophy in the CA1 subfield and the subiculum subfield is a good predictor of AD progression among patients with MCI.
- The proposed FJM can readily include multiple brain regions, and even genotype profiles, as functional predictors.

Related Publications

1. Li L, **Luo S**, Hu B, Greene T. Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease, *Statistics in Biosciences*, 2017;9(2), 357-78.
2. Li K, Chan W, Doody RS, Quinn J, **Luo S**. Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data, *J Alzheimers Dis*. 2017;58(2):361-371.
3. Li K, **Luo S**. Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease, *Stat Med*, 2017;36(22), 3560-72.
4. Li K, **Luo S**. Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease. *Stat Methods Med Res*. 2019;28(2), 327-42.
5. Li K, **Luo S**. Dynamic predictions of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data, *Stat Med*, 38(24), 4804-18.

Q: What are the application areas of Functional Data Analysis (FDA)?

When to use FDA?

- When you have functions (1D, 2D, 3D, or 4D)
- Longitudinal data: sparse functional data
- Multivariate longitudinal data
- Longitudinal –omics data: high-dimensional sparse functional data

Q: How to do Functional Data Analysis (FDA)?

Available software

- refund package in *R* is the best.
- Talk to a statistician with strong expertise in Functional Data Analysis.

Acknowledgement

- Katherine Hall, PhD
- Miriam Morey, PhD
- Harvey Cohen, MD
- Kaiyuan Hua, MS
- Kan Li, PhD