Wearables: other functional approaches

JSM 2019

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Overview

- Multi-level functional methods
- Matrix-variate multi-level functional methods
- Multi-level methods for generalized (e.g binary) functional curves
- Registration of generalized (e.g. binary) functional curves

Multi-level functional principal component analysis (MFPCA)

NIMH Family Study of Affective Spectrum Disorders

- 350 participants; ages from 10 to 84
- 5 diagnosis groups: BPI, BPII, MDD, Other and Control

Data:

- 2 weeks of follow-up measurements
- minute-by-minute activity counts
- 4945 person-days; 7,120,800 data points

Goal:

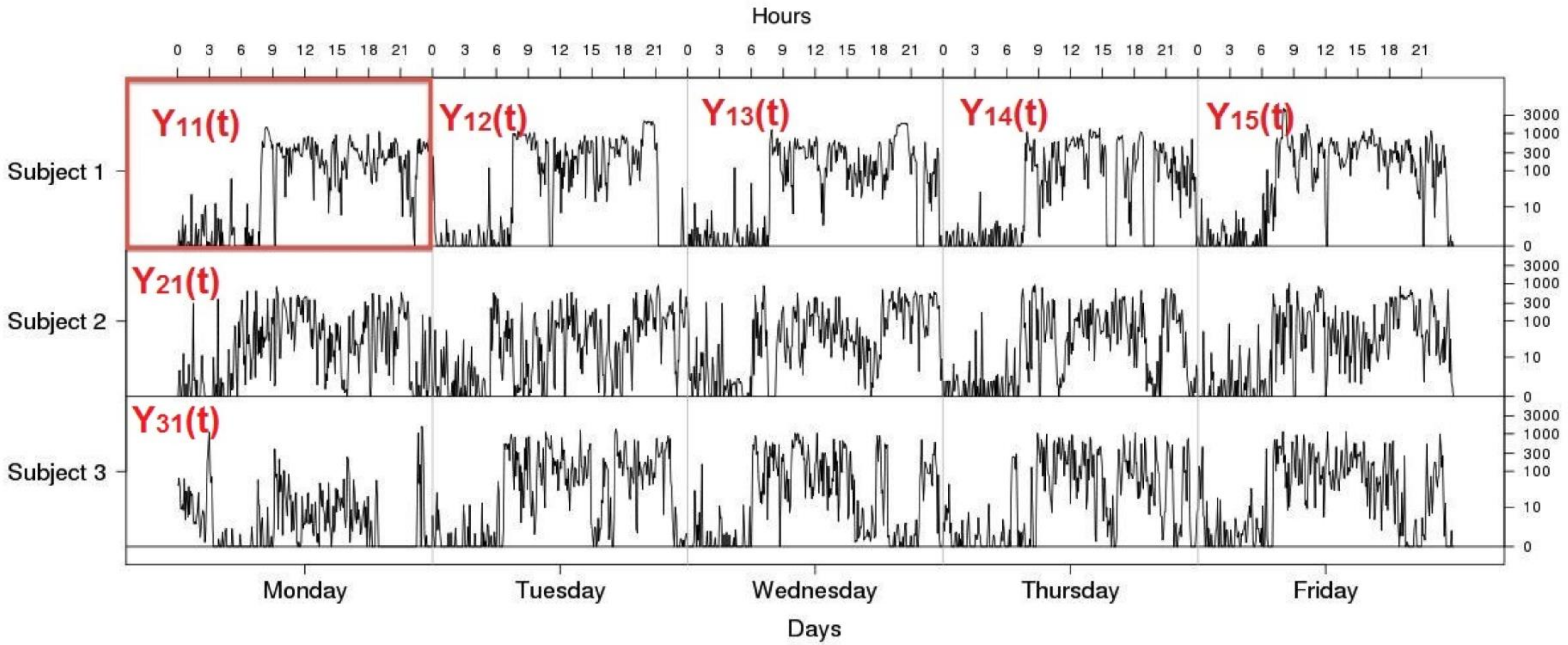
Extract representative patterns that comprise daily activity

Quantify multilevel variations after adjusting age and disease effects

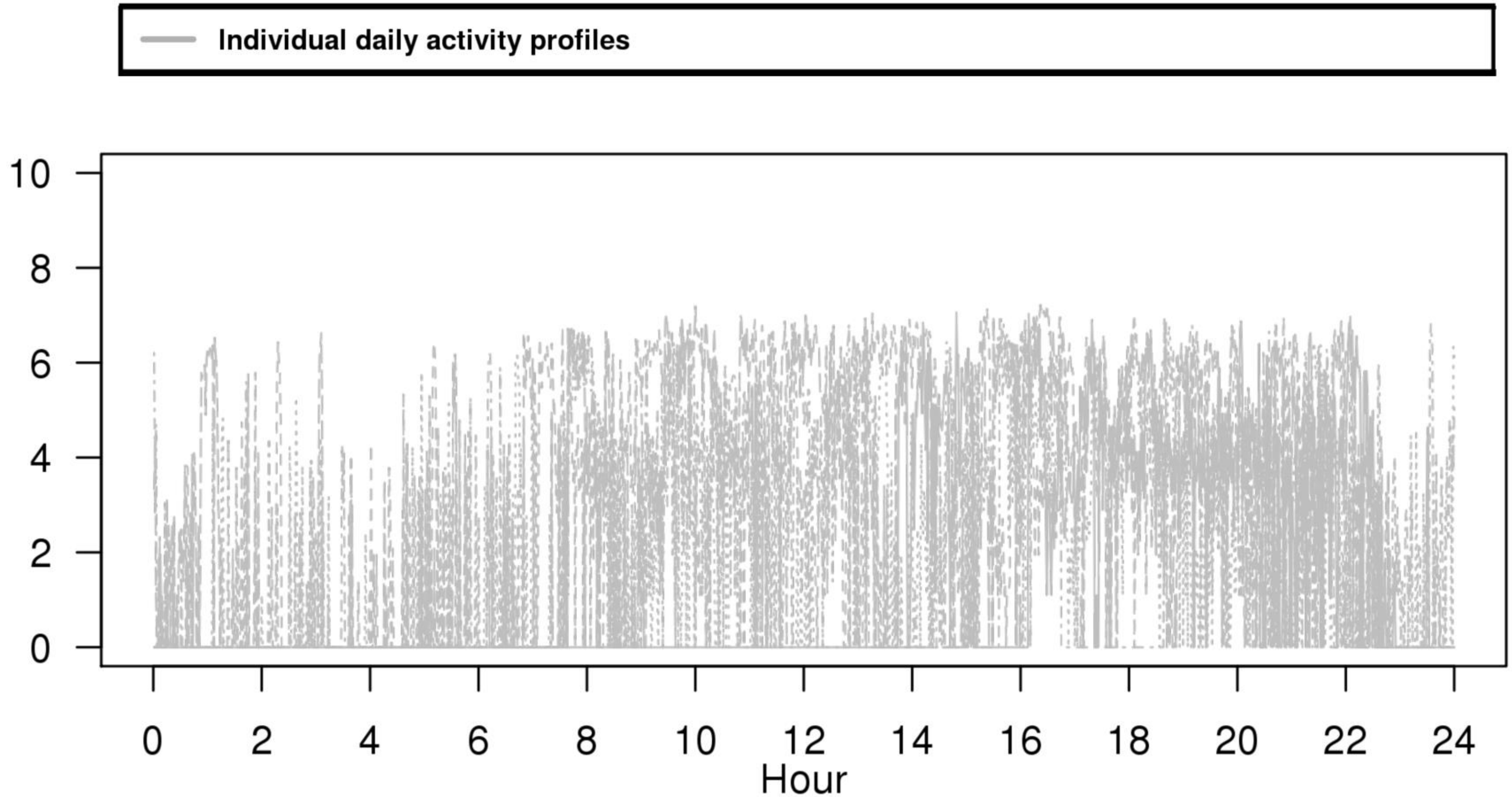
- Subject heterogeneity
- Day-to-day variation
- Age and disease effects in mean and variance components

Daily Activity Profiles

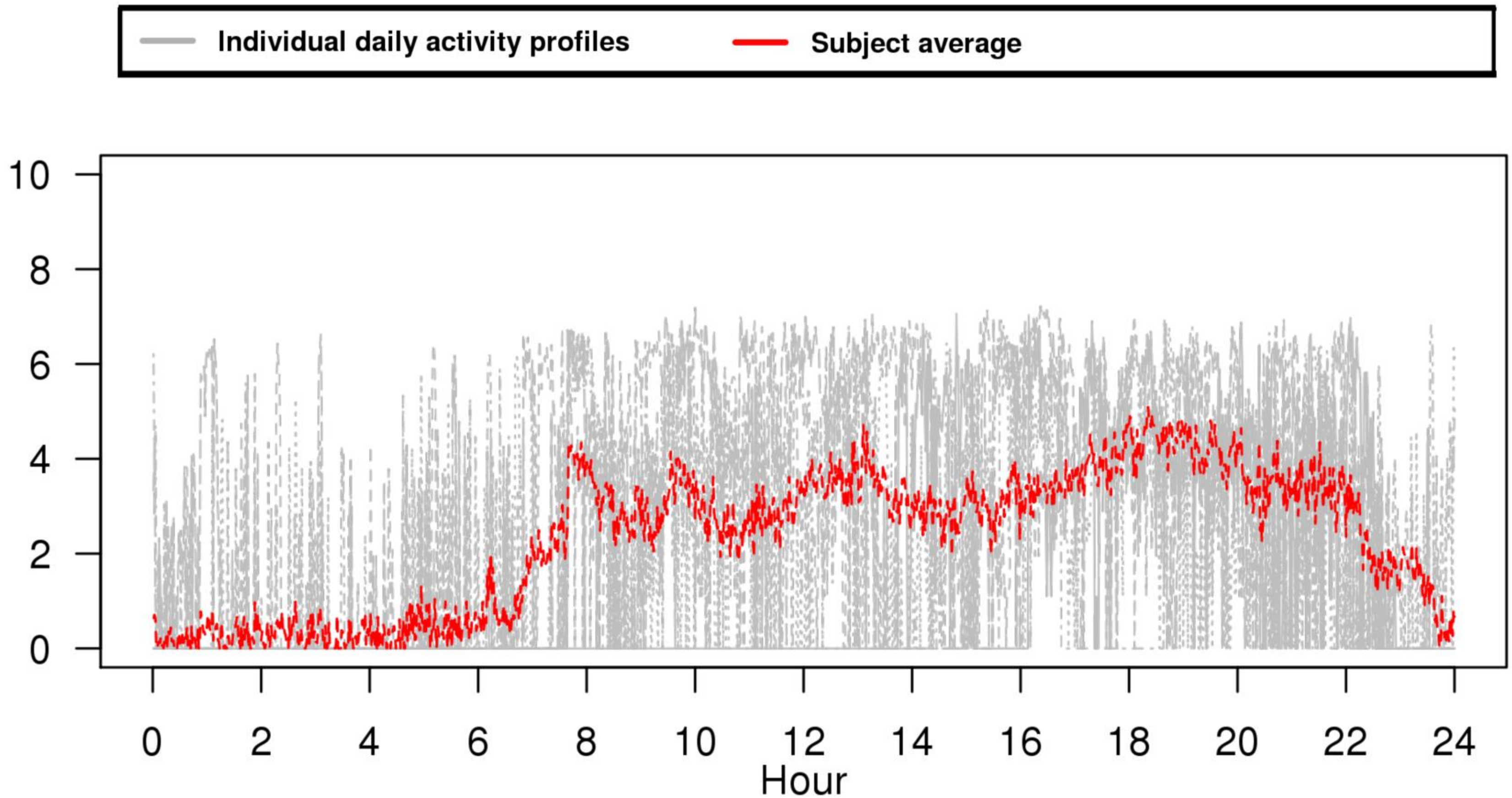
$$Y_{ij}(t),\ i = 1, 2, \cdots, I; j = 1, 2, \cdots, J_i$$



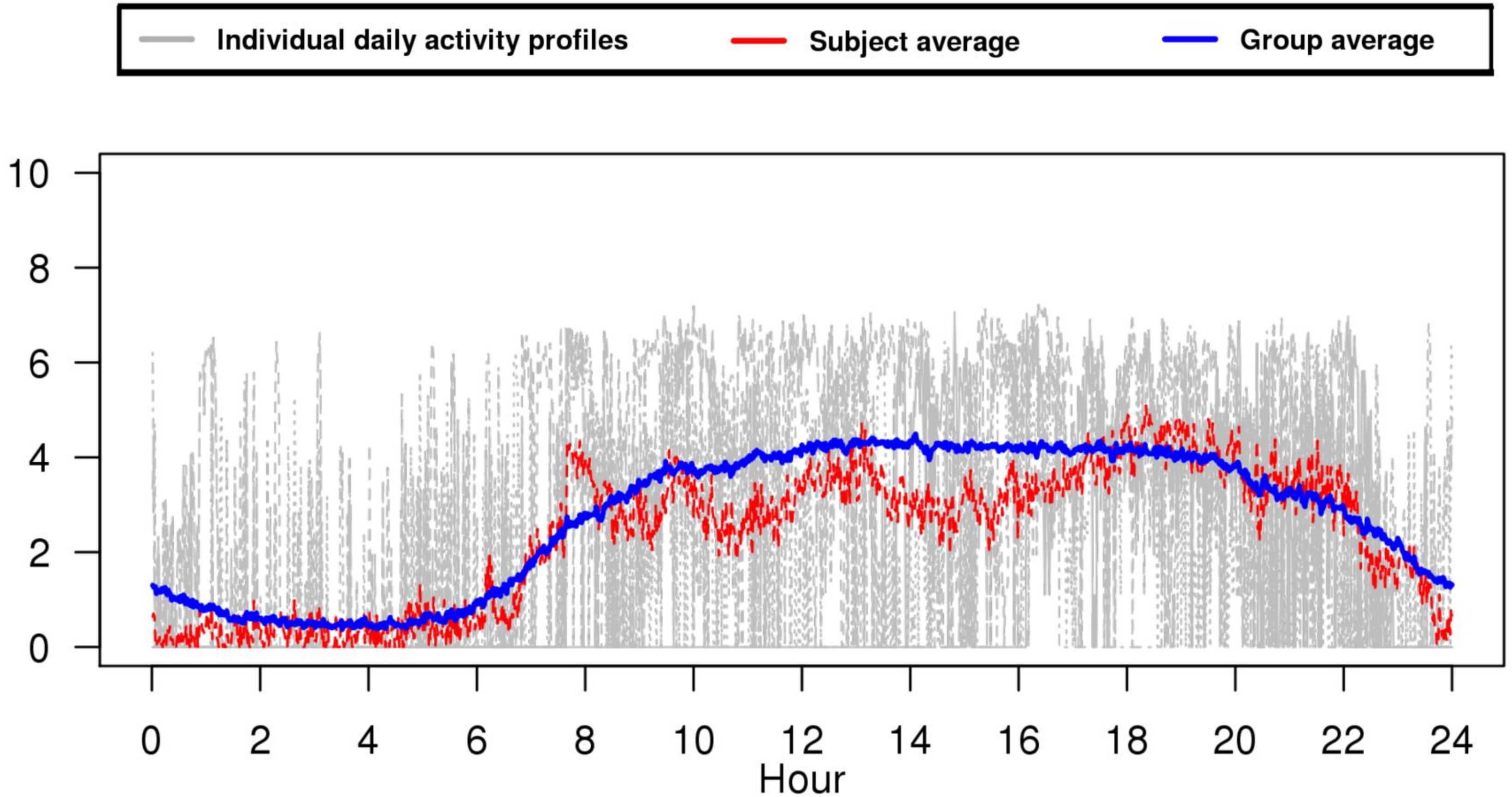
Control and age 67 (> 60)

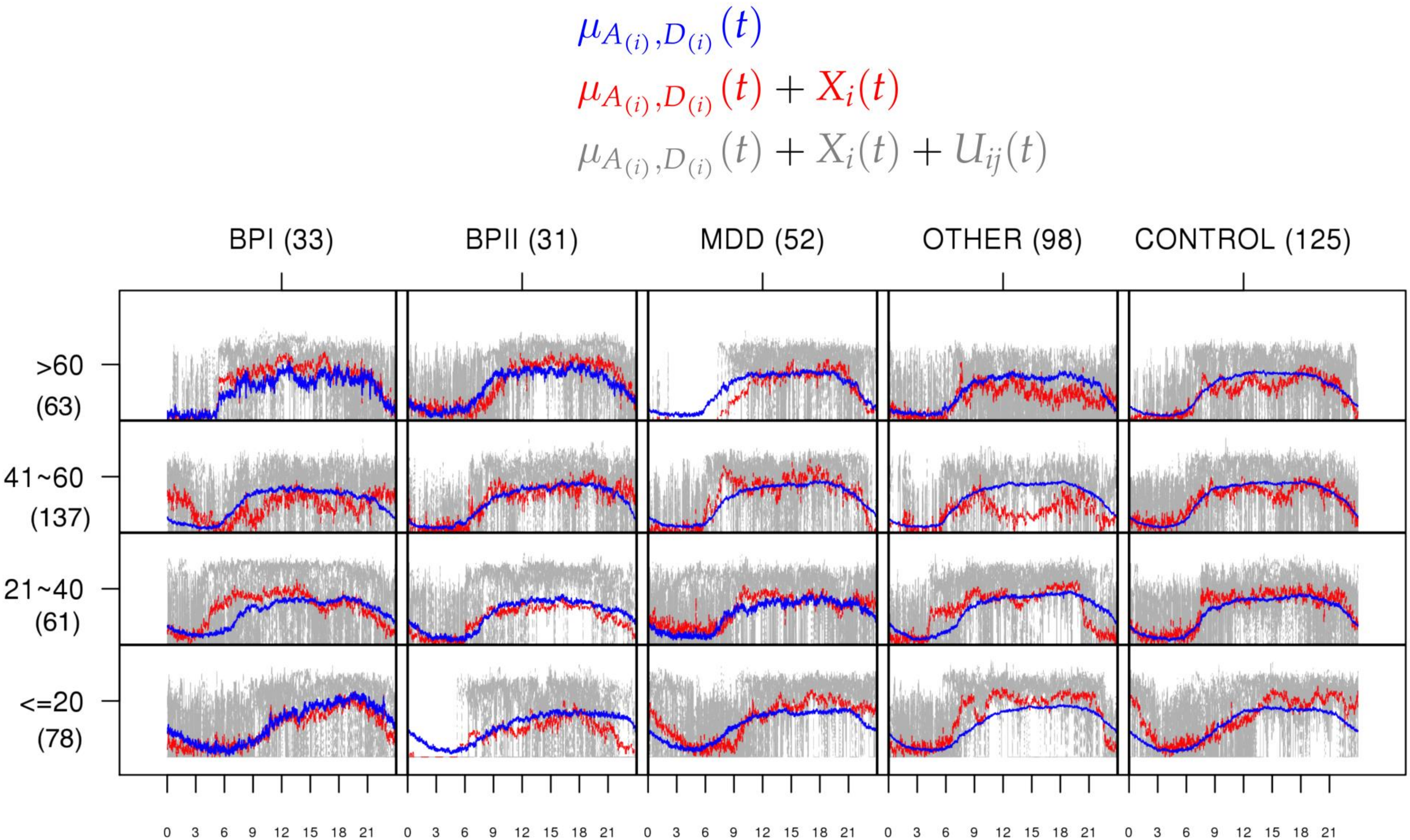


Control and age 67 (> 60)



Control and age > 60





$\mu_{A_{(i)},D_{(i)}}(t)$
$\mu_{A_{(i)},D_{(i)}}(t) + X_i(t)$
$\mu_{A_{(i)},D_{(i)}}(t) + X_i(t) + U_{ij}(t)$
BPI (33)
BPII (31)
MDD (52)
OTHER (98)
CONTROL (125)
>60
(63)
41~60
(137)
21~40
(61)
<=20
(78)
0 3 6 9 12 15 18 21

Without loss of generality, assume $Y_{ij}(t) = Y'_{ij}(t) - \mu(t, v_i)$

- Decomposability and additivity

$$Y_{ij}(t) = X_i(t) + U_{ij}(t)$$

- Identifiability

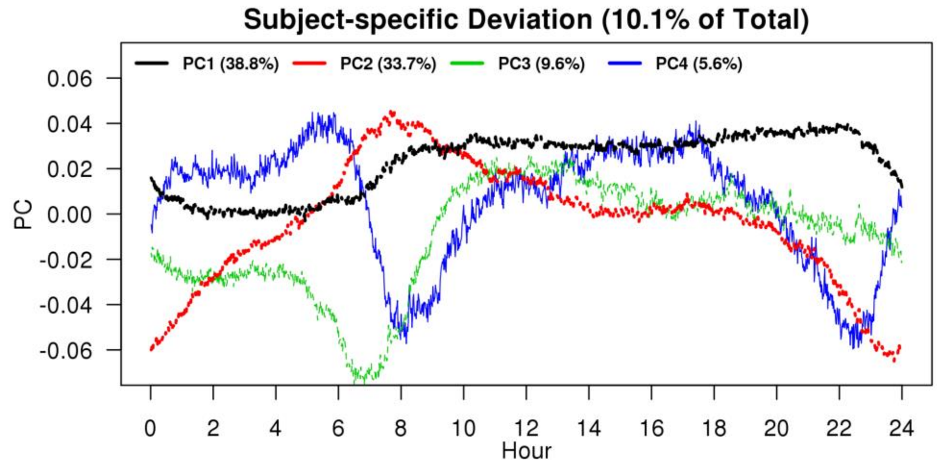
  Latent processes are mean zero and mutually uncorrelated.

- Data correlation captured by covariance operators

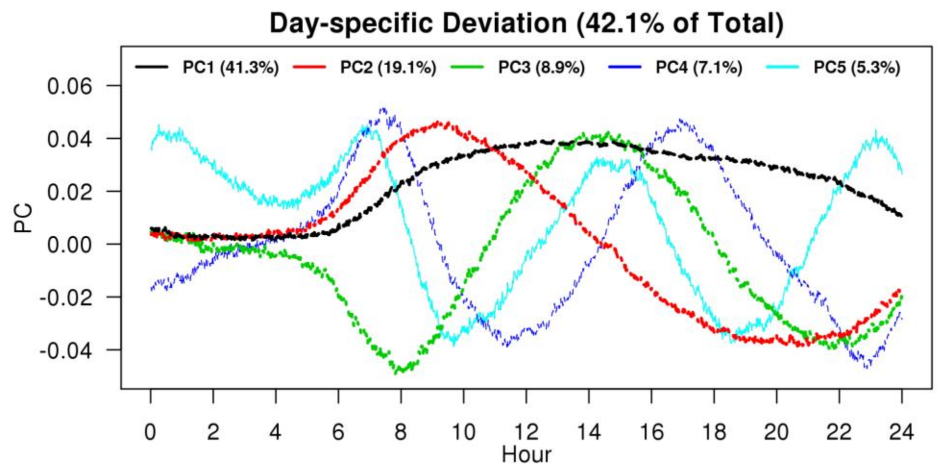
$$E\{Y_{ij}(t) - Y_{kl}(t)\}\{(Y_{ij}(s) - Y_{kl}(s)\}^T = \begin{cases} 2K_U(t,s) & i = k, j \neq l \\ 2\{K_X(t,s) + K_U(t,s)\} & i \neq k \end{cases}$$

  where $K_X(t,s) := \mathrm{Cov}\{X(t), X(s)\}$, similar definitions for K_U.

- Subject heterogeneity accounts for 10.1% of total variability
- The first 4 principal components explain 87.8% of the subject heterogeneity



- Day-specific deviation and random noise along the curve together accounts for the remaining 89.9% of total variability
- The first 5 principal components explain 81.8% of day-to-day variation



References

- Di, C.Z., Crainiceanu, C.M., Caffo, B.S. and Punjabi, N.M., 2009. Multilevel functional principal component analysis. *The annals of applied statistics*, *3*(1), p.458.
- Zipunnikov V., Caffo B.S., Yousem D.M, Davatzikos C., Schwartz B.S., Crainiceanu C. (2011),Multilevel Functional Principal Component Analysis for High-Dimensional Data. Journal of Computational and Graphical Statistics, 20(4), pp. 852-873
- Shou, H., Zipunnikov, V., Crainiceanu C., Greven, S. (2015) Structured Functional Principal Component Analysis
  Biometrics, 71 (1), pp. 247-757

Multi-level Matrix-Variate Analysis (MMVA)

Background

- **Heart failure** (HF) is a leading chronic disease in the elderly
- Lifetime risk is 20% for those over age 40 in the US
- HF burden exceeds $30 billion (> 50% on hospitalization costs)
- Identifying subjects with increased risk of hospitalization is important

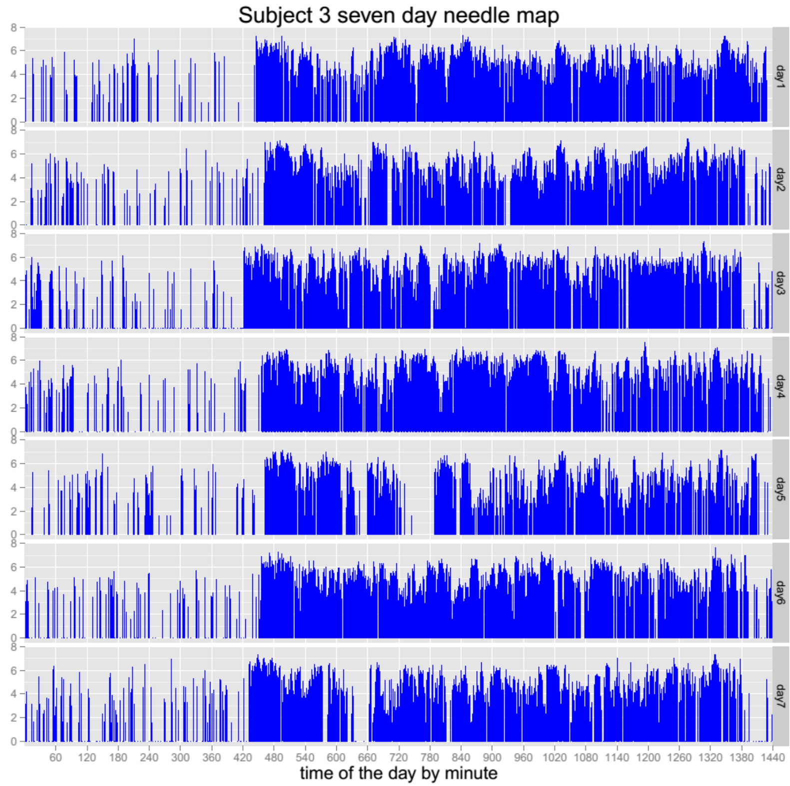
Background

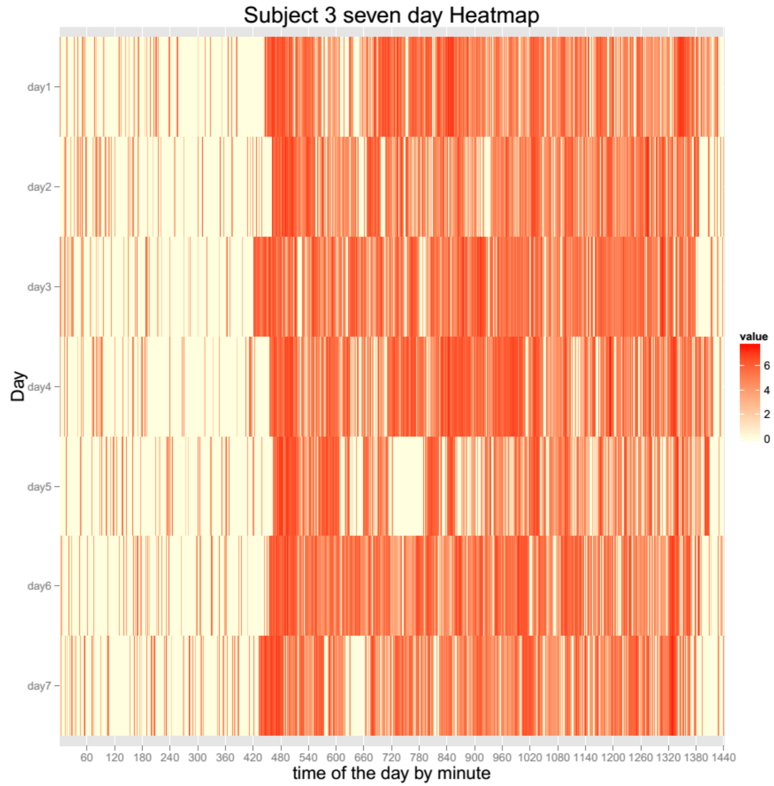
- **Static risk models** include demographics, comorbidities (AFib, hypertension, diabetes mellitus), income, etc.
- **Dynamic risk models** may be more accurate by including real-time data from wearables
- A prospective longitudinal cohort study Advanced Cardiac Care Center of Columbia University Medical Center
- 59 individuals with clinical diagnosis of congestive heart failure (CHF)
- 3-9 months of follow up
- Actical (Respironics)
- up to one month of minute-level activity counts

Background

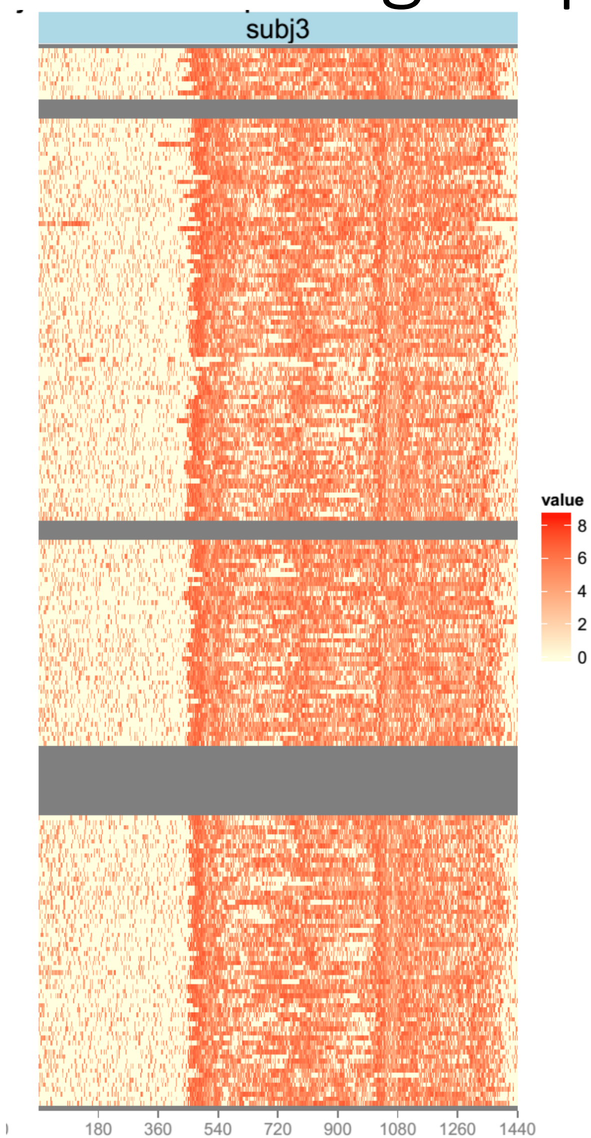
- 24 individuals had adverse clinical events
  - 14 hospitalizations
  - 10 emergency room visits
- **Goal:** model within-subject pre/post event change in patients status
- **Solution:** track week-to-week variability

Minute-level activity profiles



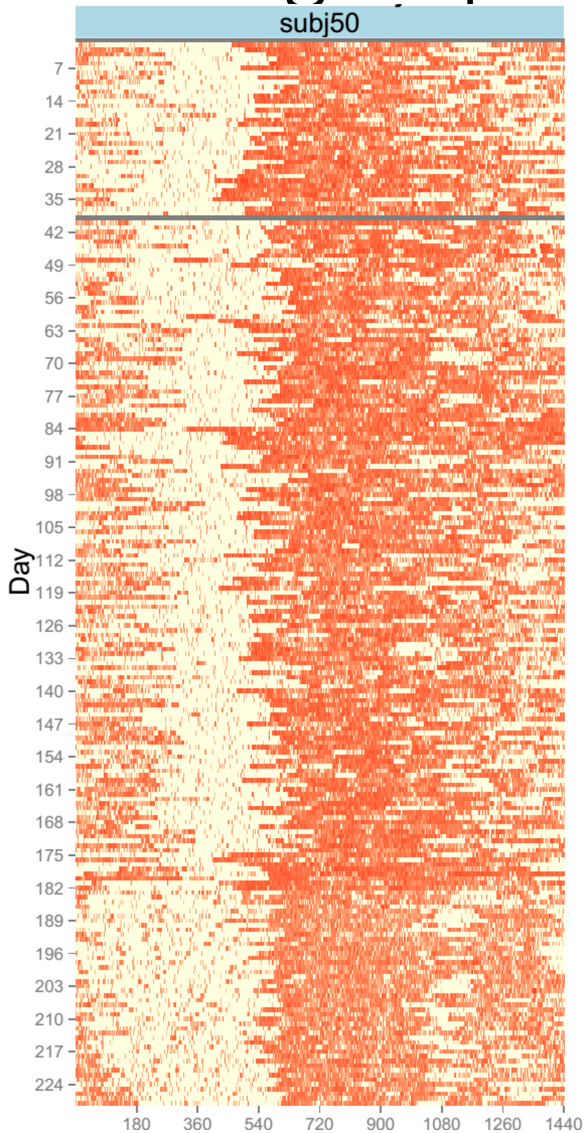


No-event group subject



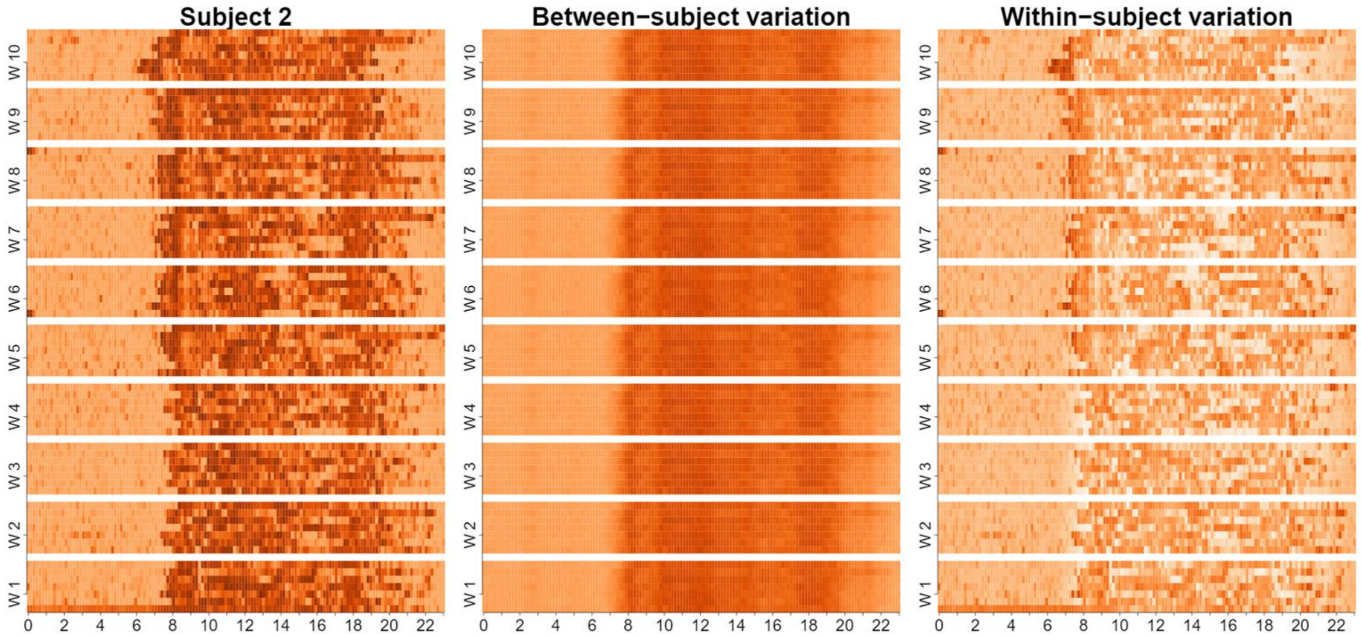
- 8 months of monitoring
- **Low week-to-week variability**
- **Had no hospitalizations**

Event-group subject



- 8 months of monitoring
- **High week-to-week variability**
- **Had a hospitalization**

Multilevel Matrix-Variate Analysis (MMVA)



- Keep within-week temporal structure across days & within-day
- Models subject-specific week-to-week variability
- uses a linear mixed effect model to account for the multilevel design
- 2D structure is handled via a matrix-variate distribution

MMVA

Model:

$$\begin{cases} \mathbf{Y}_{ij} = \boldsymbol{M} + \mathbf{X}_i + \mathbf{W}_{ij}, i = 1, \ldots, I, j = 1, \ldots, n_i \\ \mathbf{X}_i \sim \mathrm{MD}_{D,T}(\mathbf{0}, \mathbf{C}_X, \mathbf{R}_X), \\ \mathbf{W}_{ij} \sim \mathrm{MD}_{D,T}(\mathbf{0}, \mathbf{C}_W, \mathbf{R}_W), \end{cases}$$

Z follows a matrix-variate distribution: $\boldsymbol{Z} \sim \mathrm{MD}_{D,T}(\boldsymbol{M}, \mathbf{C}, \mathbf{R})$

if $\mathrm{vec}(\boldsymbol{Z}) \sim Q_{DT}(\mathrm{vec}(\boldsymbol{M}), \mathbf{R} \otimes \mathbf{C})$

Normal matrix-variate distribution

$$p(\boldsymbol{Z}|\boldsymbol{M}, \mathbf{C}, \mathbf{R}) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{R}^{-1}(\boldsymbol{Z}-\boldsymbol{M})^T\mathbf{C}^{-1}(\boldsymbol{Z}-\boldsymbol{M})\right]\right)}{(2\pi)^{DT/2}\|\mathbf{R}\|^{D/2}\|\mathbf{C}\|^{T/2}}$$

$$\mathbf{R} = E[(\boldsymbol{Z}-\boldsymbol{M})^T(\boldsymbol{Z}-\boldsymbol{M})]/\mathrm{tr}(\mathbf{C})$$

$$\mathbf{C} = E[(\boldsymbol{Z}-\boldsymbol{M})(\boldsymbol{Z}-\boldsymbol{M})^T]/\mathrm{tr}(\mathbf{R})$$

MMVA

- Between-subject matrix-variate distance

$$d(i,k) = \mathrm{dist}(\mathbf{X}_i, \mathbf{X}_k) = \|vec(\mathbf{\Gamma}_i^X) - vec(\mathbf{\Gamma}_k^X)\|.$$

- Within-subject matrix-variate distance

$$d_i(j_1, j_2) = \mathrm{dist}(\mathbf{W}_{ij_1}, \mathbf{W}_{ij_2}) = \|vec(\mathbf{\Gamma}_{ij_1}^W) - vec(\mathbf{\Gamma}_{ij_2}^W)\|.$$

MMVA

- **Between-subject scores** can be used as a “static” biomarker to enrich and potentially improve accuracy of currently used “static" risk score models.

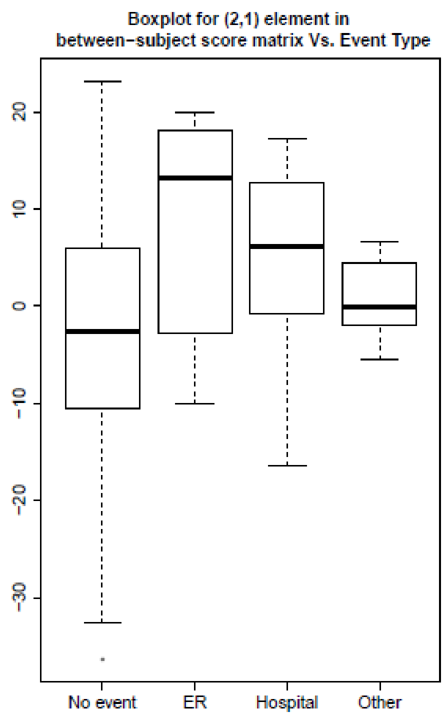
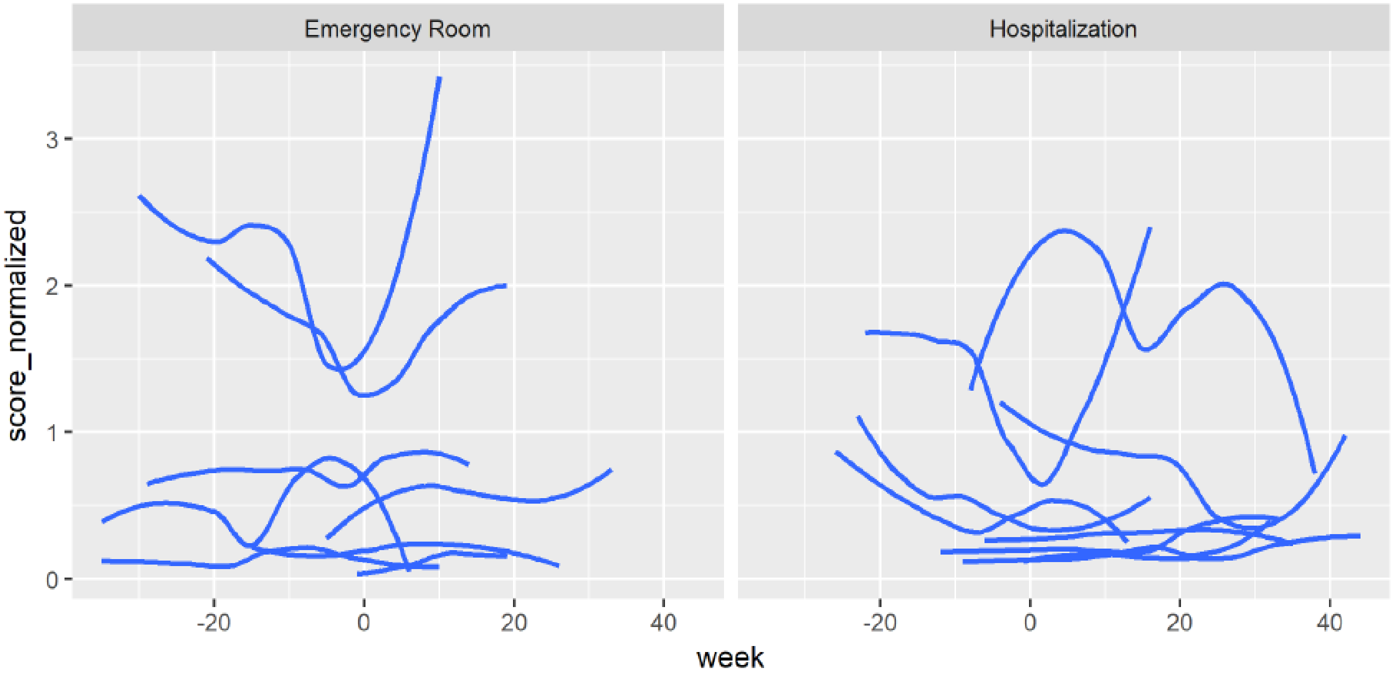


Table 3: Logistic models for between-subject scores; an asterisk indicates significance at level 0.05

	Model 1	Model 2	Model 3	Model 4
Score (2,1)	0.012(0.005)*	0.011(0.005)*	0.010(0.005)*	0.011(0.035)*
Score (2,3)	0.016(0.012)	0.015(0.013)	0.015(0.013)	0.013(0.013)
Score (4,2)	0.024(0.054)	0.014(0.056)	0.026(0.057)	0.029(0.057)
Sex		0.113(0.140)	0.099(0.141)	0.056(0.147)
Age			0.004(0.004)	0.005(0.005)
BMI				0.012(0.011)

MMVA

- **Within-subject scores** can be used as “dynamic” biomarkers that inform about weekly changes in patient status.



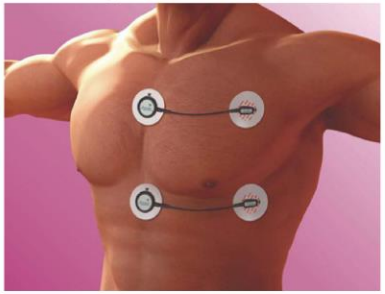
Reference

Huang, L., Bai, J., Ivanescu, A., Harris, T., Maurer, M., Green, P. and Zipunnikov, V., 2019. Multilevel matrix-variate analysis and its application to accelerometry-measured physical activity in clinical populations. *Journal of the American Statistical Association*, *114*(526), pp.553-564.

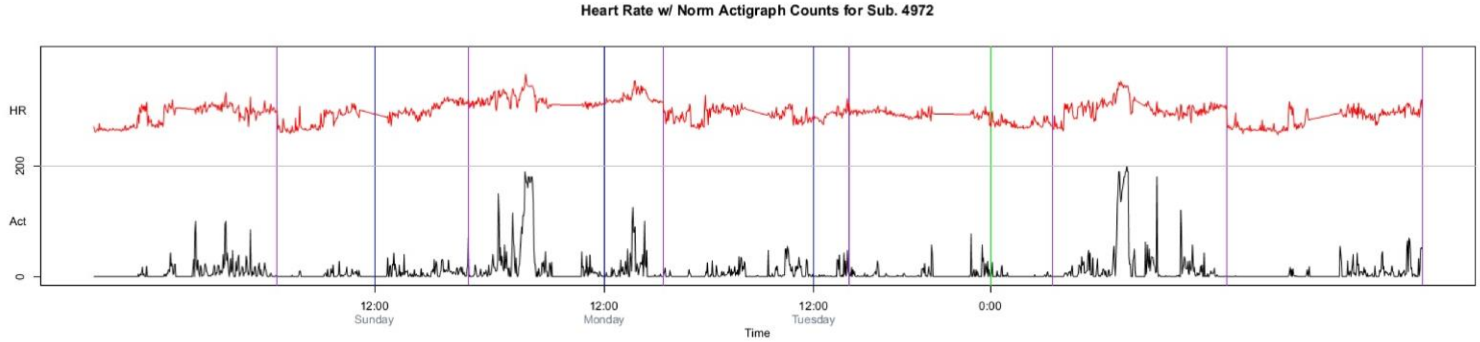
Data is freely available with the submission

Multi-level Generalized Function-on-Scalar Regression and Principal Component Analysis

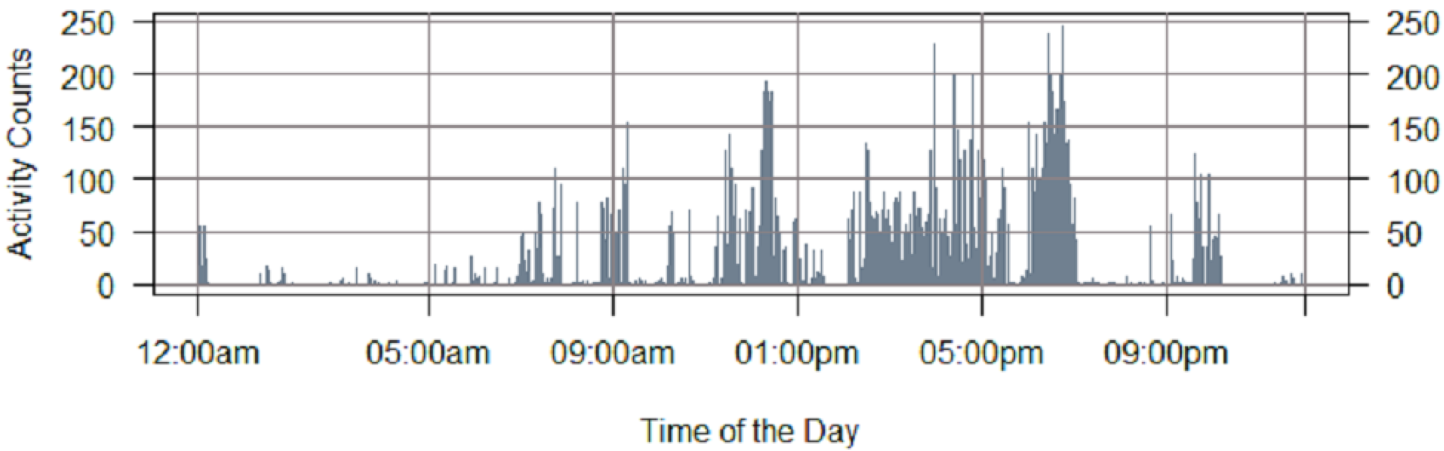
- normative human aging
- followed for life, visits every 1-4 years
- 631 subjects, wear for 7 days
- **Goal:** daily patterns of physical activity



- Every minute: activity counts and heart rate



- Binary time series $Y_{ij}(t)$: activity/inactivity
- Covariates $x_{ij,k}$: age, gender, BMI, etc



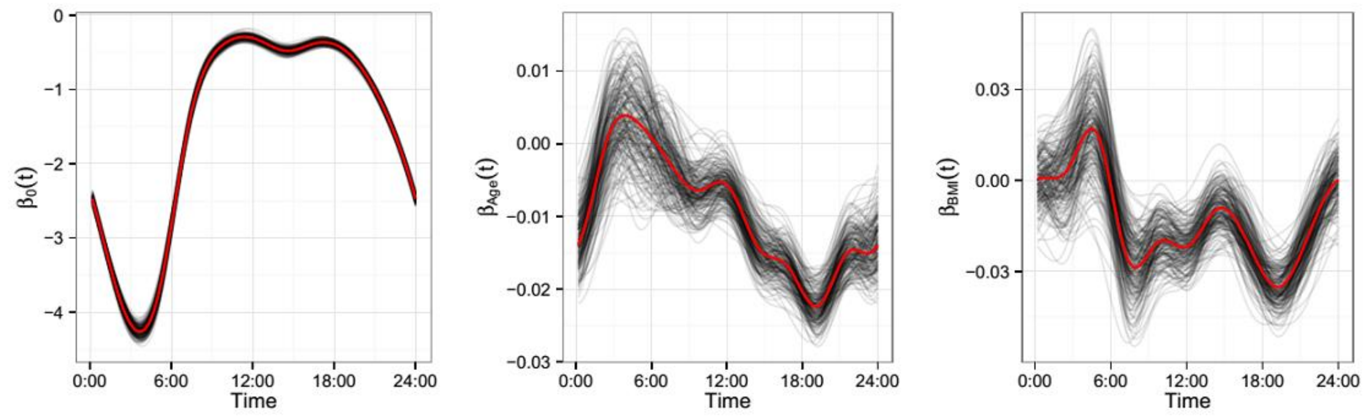
- Generalized Multilevel Functional-on-Scalar Regression
  - $Y_{ij}(t)$ is a generalized response curve
  - $Y_{ij}(t)$ comes from an exponential family

$$\begin{aligned} \mathrm{E}[Y_{ij}(t)|b_i(t), v_{ij}(t)] &= \mu_{ij}(t) \\ g(\mu_{ij}(t)) &= \beta_0(t) + \sum_{k=1}^{p} x_{ij,k}\beta_k(t) + b_i(t) + v_{ij}(t) \\ &\approx \beta_0(t) + \sum_{k=1}^{p} x_{ij,k}\beta_k(t) + \sum_{k=1}^{K^{(1)}} c_{ik}^{(1)}\psi_k^{(1)}(t) + \sum_{k=1}^{K^{(2)}} c_{ijk}^{(2)}\psi_k^{(2)}(t). \end{aligned}$$

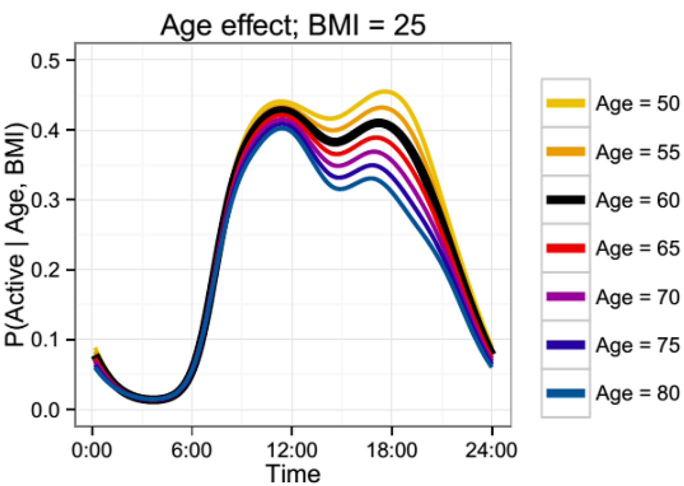
- Generalized Multilevel Functional-on-Scalar Regression
  - $\boldsymbol{\Theta}$: cubic B-spline basis functions
  - computation using Stan (Hamiltonian MC)

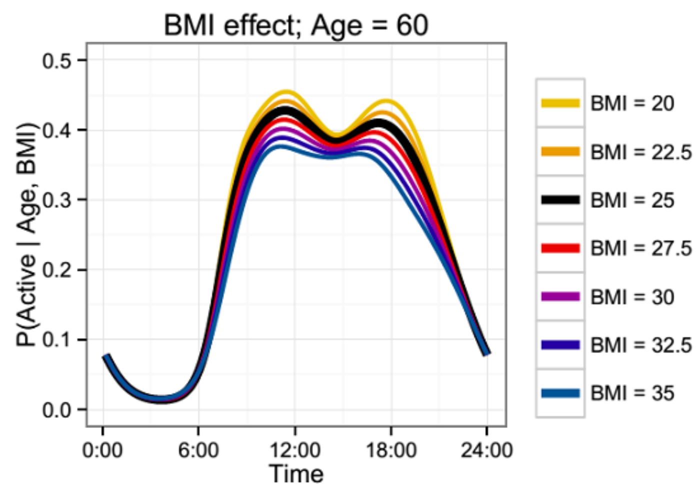
$$
\begin{aligned}
\mathrm{E}[\boldsymbol{Y}|\boldsymbol{b},\boldsymbol{v}] &= \boldsymbol{\mu} \\
g(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{v} \\
&= \boldsymbol{X}\boldsymbol{B}_X^T\boldsymbol{\Theta}^T + \boldsymbol{Z}\boldsymbol{C}^{(1)}\boldsymbol{B}_{\psi^{(1)}}^T\boldsymbol{\Theta}^T + \boldsymbol{C}^{(2)}\boldsymbol{B}_{\psi^{(2)}}^T\boldsymbol{\Theta}^T.
\end{aligned}
$$

- Generalized Multilevel Functional-on-Scalar Regression
    - Estimated functional mean, bmi, and age effects



- Estimated functional mean, bmi, and age effects



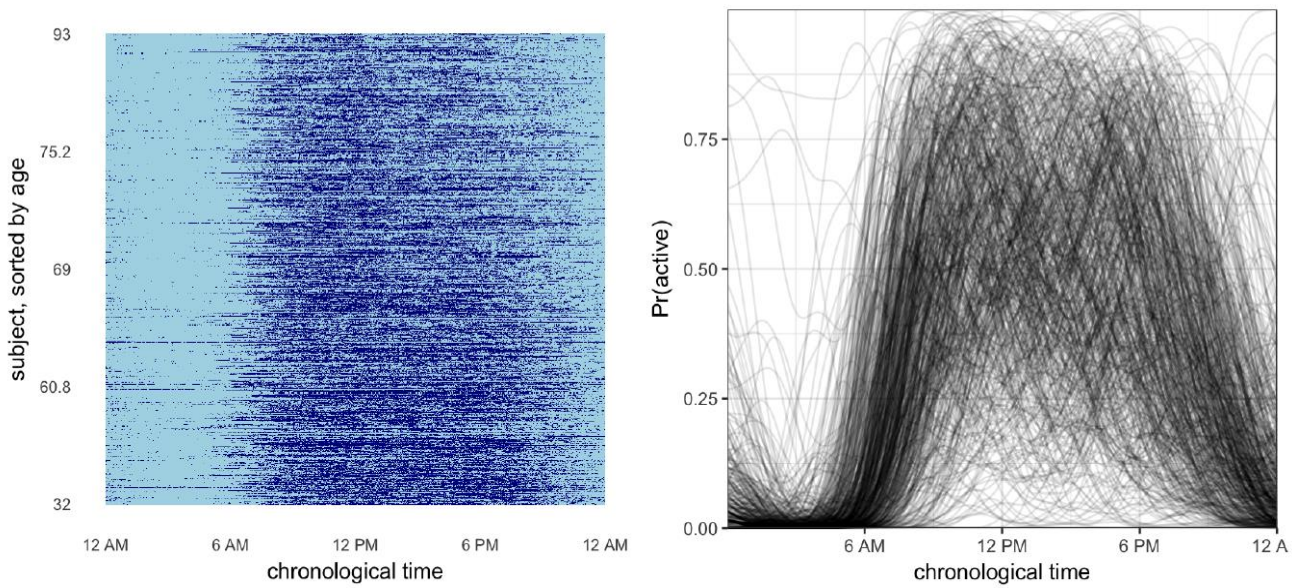


References:

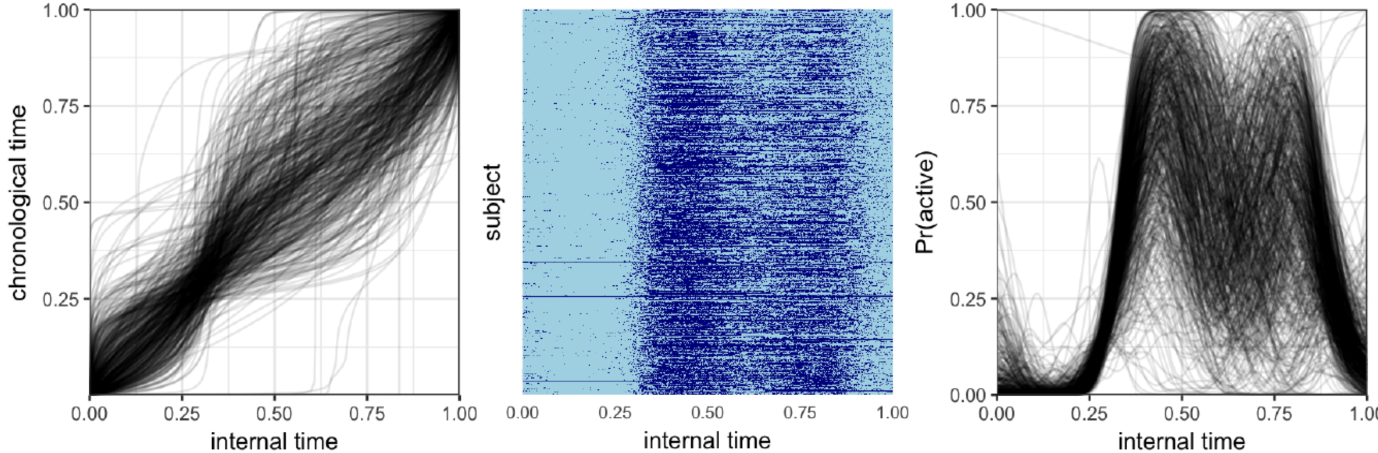
- Goldsmith, J., Zipunnikov, V, Schrack, J., Generalized Multilevel Function-on-Scalar Regression and Principal Component Analysis Biometrics, 71 (2), pp. 344-353

Registration of binary (0/1) actigraphy profiles

Registration of binary (0/1) actigraphy profiles



Registration of binary (0/1) actigraphy profiles



References:

- Wrobel, J., Zipunnikov, V., Schrack, J. and Goldsmith, J., 2019. Registration for exponential family functional data. *Biometrics*, *75*(1), pp.48-57.