Association of Health Outcomes with Gait Characteristics Extracted via Structured Functional Principal Components

Jaroslaw Harezlak, Ph.D.

Professor
Department of Epidemiology and Biostatistics
Indiana University School of Public Health
Bloomington, IN, USA

February 27, 2024
ADDS 2024, Pensacola Beach, FL

INDIANA UNIVERSITY
SCHOOL OF
PUBLIC HEALTH
BLOOMINGTON

Motivation

- Walking → **most accessible** form of physical activity
- Walking **features** → likely **predictors of physical health**
- **Feature extraction** from **hierarchical functional data** (preprocessed accelerometry data)⇒ walking spectra nested within participants
  - Structured functional principal component analysis **(SFPCA)** (Shou et al., 2015)

<u>Results</u>

⇒ **Easily interpretable walking features** on a subject-specific level

⇒ Associations between walking features and **several health indicators**

Data

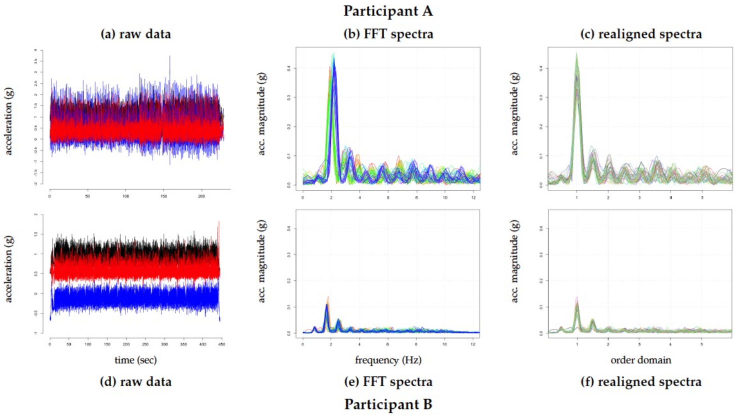
- **Sample:** N = 48 subjects enrolled in the Developmental Epidemiologic Cohort Study (DECOS) at the University of Pittsburgh (Lange-Maia et al., 2015)
  - Age: between 70.0 and 90.0 (median = 78.0, sd = 5.7),
  - BMI: between 20.5 and 37.9 (median = 25.9, sd = 3.9)
- Accelerometry Data:
  - Different activities in-the-lab and free-living data collected for 7 days
  - Device location:
    - Left wrist, right wrist and hip: ActiGraph GT3X+ (80hz)
    - Thigh: activPAL 3 (20hz)

Data pre-processing

- Described in detail in Urbanek et al. (2018), Fadel et al. (2020)

- Briefly:
  - 400m fast-paced walk data extracted from in-the-lab signal
  - Shot-time Fast Fourier Transform (SFFT) applied to 10-second intervals (46 intervals)
  - Spectra aligned in the, so called, **order domain**

Data pre-processing



Statistical Methods – Overview

(1) **Structured functional principal component analysis** (SFPCA) (Shou et al., 2015)

- Dimension reduction on two levels $\Rightarrow$ feature extraction on subject-specific level

(2) **Principal component regressions** (PCRs)

- OLS regression $\Rightarrow$ subject-specific level features ($\approx$ principal scores) related to age and physical health indicators

SFPCA – Model

$$\mathbf{Y_{ij}(t)} = \mu(\mathbf{t}) + \texttt{subject}_{\mathbf{i}}(\mathbf{t}) + \texttt{spectrum}_{\mathbf{ij}}(\mathbf{t}) + \varepsilon_{\mathbf{ij}}(\mathbf{t}),$$

$$\underbrace{i \in \{1, \ldots, 48\}}_{\text{(subjects)}},\ \underbrace{j \in \{1, \ldots, 46\}}_{\text{(spectra)}},\ \underbrace{t \in \{0.30, 0.31, \ldots, 5.75\}}_{\text{(order domain)}}$$

$$\varepsilon_{ij}(t) \sim \mathrm{N}\left(0, \sigma^2\right),\ i.i.d.$$

- $Y_{ij}(t)$ = acceleration magnitude for order domain point t on a spectrum curve j nested within participant i
- subject$_i(t)$ = latent subject-specific process,
- spectrum$_{ij}(t)$ = latent subject-spectrum specific process
- order domain axis sampled in equal steps of 0.01 ⇒ grid length $p = 546$

Covariance separation by SFPCA

$$\begin{aligned}\mathsf{Cov}\left(Y_{ij}(s), Y_{ij}(t)\right) &= \mathsf{Cov}\left(\texttt{subject}_i(s), \texttt{subject}_i(t)\right) + \\ &+ \mathsf{Cov}\big(\texttt{spectrum}_{ij}(s), \texttt{spectrum}_{ij}(t)\big)\end{aligned}$$

- BUT `subject` and `spectrum` are latent

$\Rightarrow$ Estimate K_X and K_U using design-specific matrices $\mathbf{G}_X$ and $\mathbf{G}_U$ ("implicit" level separation → w/o estimating level data explicitly)

$$\Rightarrow \widehat{K}_Y = \widehat{K}_X + \widehat{K}_U = Y\mathbf{G}_X Y' + Y\mathbf{G}_U Y'$$

Feature extraction

Level-specific eigendecomposition of covariance → feature extraction

$$
\begin{aligned}
\widehat{K}_X(s,t) &= \sum_{k=1}^{d^*_{\texttt{sub}}} \widehat{\lambda}_k^{\texttt{sub}} \widehat{\phi}_k^{\texttt{sub}}(s) \widehat{\phi}_k^{\texttt{sub}'}(t), \\
\widehat{K}_U(s,t) &= \sum_{\ell=1}^{d^*_{\texttt{spec}}} \widehat{\lambda}_\ell^{\texttt{spec}} \widehat{\phi}_\ell^{\texttt{spec}}(s) \widehat{\phi}_\ell^{\texttt{spec}'}(t)
\end{aligned}
$$

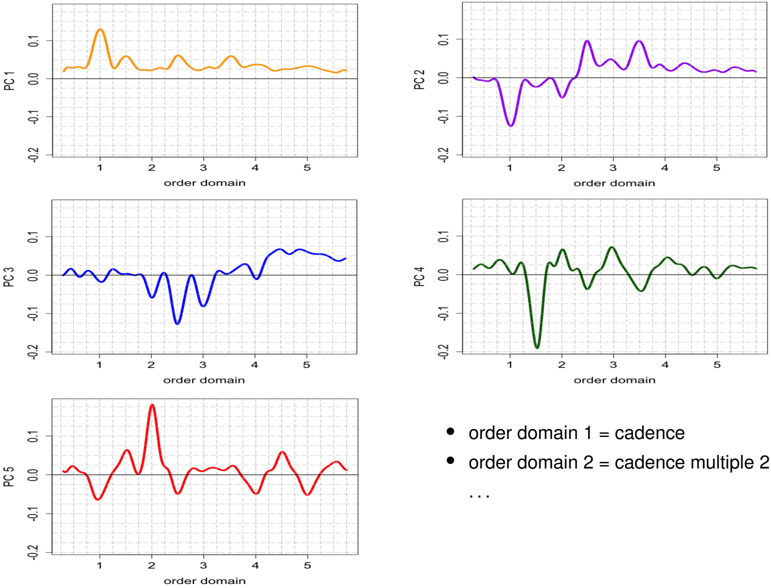
$\widehat{\phi}^{\texttt{sub}}$, $\widehat{\phi}^{\texttt{spec}}$: level-specific eigenfunctions (“walking features”)
$\widehat{\lambda}^{\texttt{sub}}$, $\widehat{\lambda}^{\texttt{spec}}$: level-specific eigenvalues

Scores subject-specific only

$$
\xi_k^{\texttt{sub}} = \int \texttt{subject}(t) \widehat{\phi}_k^{\texttt{sub}}(t) dt
$$

⇒ Subject-specific level scores as BLUP in two-level framework (e.g. Crainiceanu et al., 2009)

Subject-specific level features



- order domain 1 = cadence
- order domain 2 = cadence multiple 2

  ...

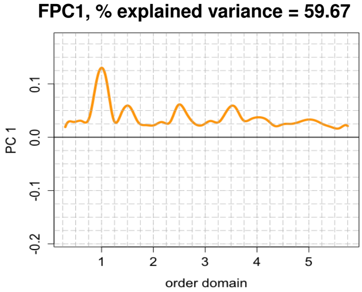
Subject-specific level features

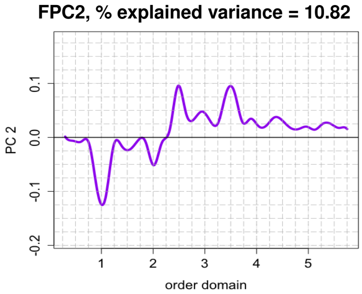
- Participant-specific walking pattern approximated through

$$\widetilde{\mathtt{Y}}_i^{\mathtt{sub}}(t) = \sum_{k=1}^{5} \widehat{\phi}_k^{\mathtt{sub}}(t)\widehat{\xi}_{ik}^{\mathtt{sub}} \quad \text{(Karhunen-Loève)}$$

  where $\widehat{\phi}_k^{\mathtt{sub}}$ = features and $\widehat{\xi}_{ik}^{\mathtt{sub}}$ = scores
- Next: analyze *pronounced* peaks and valleys in $\widehat{\phi}_k^{\mathtt{sub}}$

Interpretation example





- Subject i with higher score $\xi_{1i}^{\text{sub}} > 0$ has a higher acceleration magnitude at the cadence
- FPC1 captures most of heterogeneity b/w participants → participants vary most strongly by acceleration magnitude at cadence

- Subject i with higher score $\xi_{2i}^{\text{sub}} > 0$ has a lower acceleration magnitude at the cadence and a higher acceleration magnitude at cadence multiples 2.5 and 3.5

Outcome regressions – Algorithm

(1) **Outcome regression:** Regress Z (e.g. BMI) on score matrix $\widehat{\boldsymbol{\xi}}^{\text{sub}}$, and other regressors R.

(2) **Subset selection:** Select best model via an information criterion (R is "forced" into the model).

(3) **Interpretation:** Interpret significant associations.

Outcome regressions

$$\mathbb{E}[Z_i] = \beta_0 + \sum_{k=1}^{5} \widehat{\xi}_{ik}^{\texttt{sub}} \beta_k + \texttt{Age}_i \gamma_1 + \texttt{Male}_i \gamma_2$$

$Z \in \{\texttt{Age}, \texttt{BMI}, \texttt{BPM}, \texttt{Avg\_cadence}, \texttt{PFS\_mental}, \texttt{PFS\_physical}, \texttt{MAP}, \texttt{SPPB}\}$

- BPM = beats per minute (heart rate)
- PFS = Pittsburgh Fatigability Scale
- MAP = Mean Arterial Pressure [(SBP + 2*DBP)/3]
- SPPB = Short Physical Performance Battery

Note: subset regressions consider subset of $\left[\widehat{\xi}_{i1}^{\texttt{sub}}, \ldots, \widehat{\xi}_{i5}^{\texttt{sub}}\right]$

Results – Final models

Coeff/Z	Age	BMI	BPM	Avg_cad	PFS_ment	PFS_phys	MAP	SPPB
Const	78.413	53.007	88.948	1.769	12.735	10.000	126.366	16.764
Age		-0.342 (0.003)	-0.311 (0.298)	0.003 (0.468)	-0.063 (0.817)	0.080 (0.709)	-0.442 (0.165)	-0.078 (0.118)
Male	1.065 (0.464)	0.998 (0.368)	-4.407 (0.124)	-0.067 (0.163)	1.034 (0.701)	-1.383 (0.506)	-6.278 (0.074)	-0.447 (0.343)
sc1 (60%)	-5.691 (<0.001)	-3.747 (0.002)	-3.509 (0.272)	0.194 (<0.001)	-5.054 (0.087)	-3.654 (0.113)		1.234 (0.023)
sc2 (11%)	6.915 (0.017)						16.696 (0.035)	
sc3 (7%)	8.812 (0.026)	-7.349 (0.021)		-0.252 (0.063)				
sc4 (5%)	12.585 (0.007)							
sc5 (4%)								
$\mathbf{R}^2_{\text{adj, b}}$	-0.023	0.048	0.046	0.077	-0.030	0.048	0.038	0.192
$\mathbf{R}^2_{\text{adj, f}}$	0.474	0.324	0.051	0.325	0.027	0.088	0.116	0.272
I	45	45	45	45	39	40	45	45

Interpretation – Feature 1

Z	$\widehat{\beta}_1$
Age	-5.691
BMI	-3.747
Avg_cad	0.194
SPPB	1.234

higher $\widehat{\xi}_{i1}^{\text{sub}}$ associated with

- younger age
- lower BMI
- faster average cadence
- higher SPPB scores

Note that $\widehat{\phi}_1^{\text{sub}} \approx 0.13 \Rightarrow$ higher $\widehat{\xi}_{i1}^{\text{sub}}$ also associated with higher acceleration magnitude at cadence

$\Rightarrow$ Individuals with higher acc. magnitude at cadence predicted to **(i)** be younger, **(ii)** have a lower BMI, **(iii)** make more steps per second, and **(iv)** have a better physical function

Interpretation – Feature 2

Z	$\widehat{\beta}_2$
Age	6.92
MAP	16.70

Higher $\widehat{\xi}_{i2}^{\text{sub}}$ associated with

- older age
- higher MAP

Note that $\widehat{\phi}_2^{\text{sub}} \approx -0.125 \Rightarrow$ higher $\widehat{\xi}_{i2}^{\text{sub}}$ associated with lower acc. magnitude at cadence;
$\widehat{\phi}_2^{\text{sub}} \approx 0.1 \Rightarrow$ higher $\widehat{\xi}_{i2}^{\text{sub}}$ associated with higher acc. magnitude at cadence multiples 2.5 and 3.5

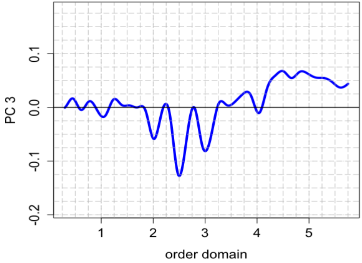
$\Rightarrow$ Individuals with lower acc. magnitude at cadence and higher acc. magnitude at cadence multiples 2.5 and 3.5 predicted to **(i)** be older, **(ii)** have a higher MAP

Interpretation – Feature 3

Z	$\widehat{\beta}_3$
Age	8.81
BMI	-7.35

Higher $\widehat{\xi}_{i3}^{\mathrm{sub}}$ associated with

- older age
- lower BMI



Note that $\widehat{\phi}_1^{\mathrm{sub}} \approx -0.13 \Rightarrow$ higher $\widehat{\xi}_{i3}^{\mathrm{sub}}$ associated with lower acc. magnitude at cadence multiple 2.5

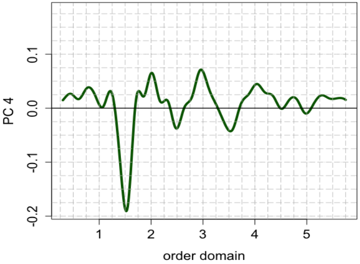
$\Rightarrow$ Individuals with lower acc. magnitude at cadence multiple 2.5 predicted to **(i)** be older, and **(ii)** have a lower BMI

Interpretation – Feature 4

Z	$\widehat{\beta}_4$
Age	12.59

Higher $\widehat{\xi}_{i4}^{\mathrm{sub}}$ associated with

- older age



Note that $\widehat{\phi}_1^{\mathrm{sub}} \approx -0.185 \Rightarrow$ higher $\widehat{\xi}_{i4}^{\mathrm{sub}}$ associated with lower acc. magnitude at cadence multiple 1.5

$\Rightarrow$ Individuals with lower acc. magnitude at cadence multiple 1.5 predicted to be older

Discussion

- Processed accelerometry data ≈ functional data → curves provide **time-ordered intra-walk information**
  - **Cadence:** alignment of curves; proxy-indicator for walking intensity (Tudor-Locke et al., 2018)
  - **Quantification of walking asymmetry:** more energy at high frequencies → unstable walk
- Subject-specific level features **significantly related** to several indicators of physical health
  - ⇒ Individual **walking pattern** may shed light on subject's **subclinical disease status**
  - ⇒ Potentially: 400m corridor walk performance for older adults → **prognostic factor for health outcomes**

Who

- Dr. Verena Werkmann (PhD student and now a postdoc)
- Dr. Nancy Glynn (UPitt collaborator)

Literature

- Crainiceanu, C. M., A.-M. Staicu, and C.-Z. Di (2009). Generalized multilevel functional regression. *Journal of the American Statistical Association* 104(488), 1550–1561.
- Di, C.-Z., C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics* 3(1), 458.
- Fadel, W. F., J. K. Urbanek, N. W. Glynn, and J. Harezlak (2020). Use of functional linear models to detect associations between characteristics of walking and continuous responses using accelerometry data. *Sensors 20(21)*, 6394.
- Glynn, N. W., A. J. Santanasto, E. M. Simonsick, R. M. Boudreau, S. R. Beach, R. Schulz, and A. B. Newman (2015). The Pittsburgh fatigability scale for older adults: development and validation. *Journal of the American Geriatrics Society* 63(1), 130–135.
- Lange-Maia, B. S., A. B.Newman, E. S. Strotmeyer, T. B. Harris, P. Caserotti, andN.W. Glynn (2015). Performance on fast-and usual-paced 400-m walk tests in older adults: are they comparable? *Aging Clinical and Experimental Research* 27(3), 309–314.
- Shou, H., V. Zipunnikov, C. M. Crainiceanu, and S. Greven (2015). Structured functional principal component analysis. *Biometrics* 71(1), 247–257.
- Tudor-Locke, C., H. Han, E. J. Aguiar, T. V. Barreira, J. M. Schuna Jr, M. Kang, and D. A. Rowe (2018). How fast is fast enough? walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *British Journal of Sports Medicine* 52(12), 776–788.
- Urbanek, J. K., V. Zipunnikov, T. Harris,W. Fadel, N. Glynn, A. Koster, P. Caserotti, C. Crainiceanu, and J. Harezlak (2018). Prediction of sustained harmonic walking in the free-living environment using raw accelerometry data. *Physiological measurement* 39(2), 02NT02.