JOHNS HOPKINS
BLOOMBERG SCHOOL of PUBLIC HEALTH

JOHNS HOPKINS
UNIVERSITY

Statistical methods for wearable device data and sample size calculation in complex models

Marta Karas

Advisor: Ciprian M. Crainiceanu

Co-advisor: Jacek K. Urbanek

November 17, 2021

Recording from the defense presentation:

https://youtu.be/xST4dqInxBo

About

Profile

- 5th year PhD candidate in Biostatistics at Johns Hopkins University
    - Wearable and Implantable Technology (WIT) lab
    - ENGAGE lab
- Industry summer internships:
    - 2019: Novartis (Switzerland) @ Digital Solutions
    - 2020: Evidation Health (CA, USA) @ Digital Measures

Interests – methods for wearable devices data

- Pattern identification and quantification
- Accelerometry data preprocessing
- Physical activity digital measures
- R software development

Interests -- other statistical methods

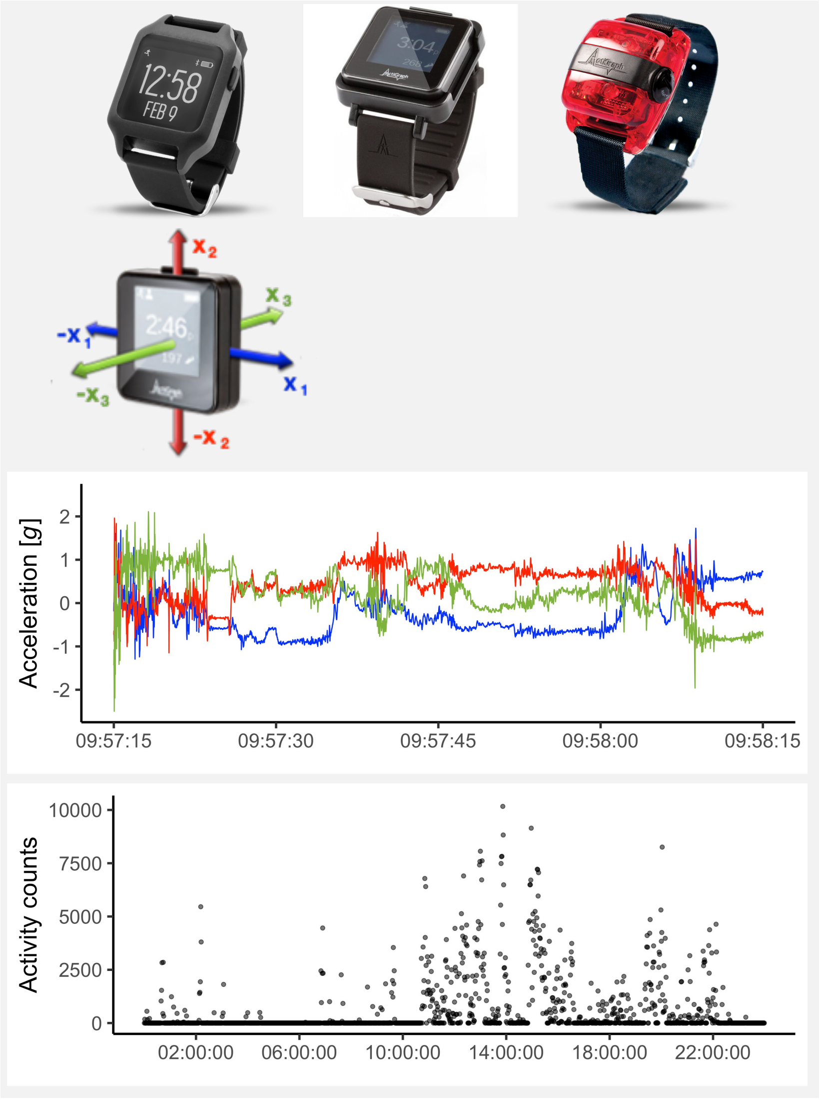
- Power estimation in complex settings
- Regularization, change point detection

Outline

- [16 min] The ADEPT pattern-recognition method with application to walking stride segmentation from raw accelerometry data
- [12 min] Harmonization of accelerometry-based measures of physical activity
- [12 min] The upstrap for power and sample size estimation in complex models

Accelerometry data in health research

- **Wearable monitors** allow for non-invasive, objective monitoring of human motor activity
- **Accelerometer** measures acceleration [*g*] along three orthogonal axes
- Accelerometry data
  - **Raw data**: three-dimensional time series of acceleration
  - **Summary measures**: raw data aggregated in fixed-time windows (e.g., 1 minute intervals)



Figures rows 1 and 2: ActiGraph, LLC (adapted). Sourced from: https://actigraphcorp.com/.

Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation

Karas, M., Straczkiewicz, M., Fadel, W., Harezlak, J., Crainiceanu, C. M., Urbanek, J. K. (2019). *Biostatistics*, 22(2), 331–347. https://doi.org/10.1093/biostatistics/kxz033

Estimation of free-living walking cadence from wrist-worn sensor accelerometry data and its association with SF-36 quality of life scores

Karas, M., Urbanek, J. K. U., Illiano, V. P., Bogaarts, G., Crainiceanu, C. M., Dorn, J. F. (2021). *Physiological measurement*, 42(6). https://doi.org/10.1088/1361-6579/ac067b

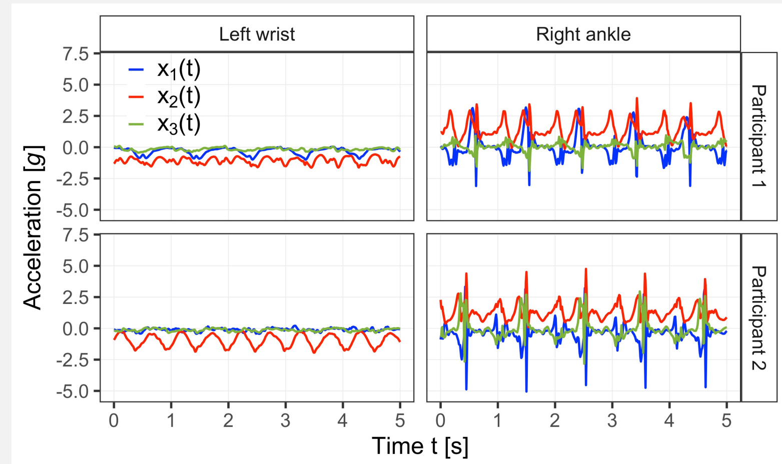
ADEPT: method for automatic, fast and accurate pattern segmentation in time-series

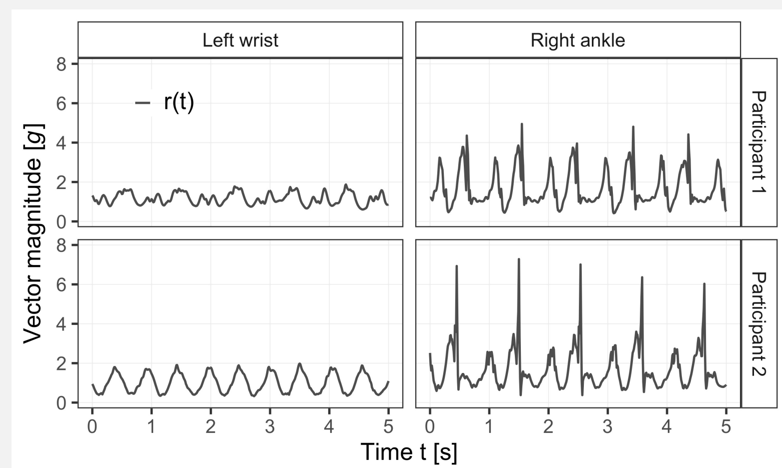
Scientific problem

- Detailed walking characteristics have become increasingly important in health studies
    - Distance covered and speed in a 6 min walk, cadence, stride pattern variability, gait symmetry[1]
- Context: supervised and semi-supervised walking
- Need for automatic, fast and accurate methods for walking strides segmentation from raw accelerometry data

Challenges

- Variations in shape and duration of a pattern within and between individuals
- Different sensor locations: wrist (left, right), lower back, hip, ankle





$$r(t) = \sqrt{x_1^2(t) + x_2^2(t) + x_3^2(t)}$$

1: Studenski et al., 2011; Brown et al., 2014; Urbanek et al., 2017; Del Din et al., 2019

ADEPT: method for automatic, fast and accurate pattern segmentation in time-series

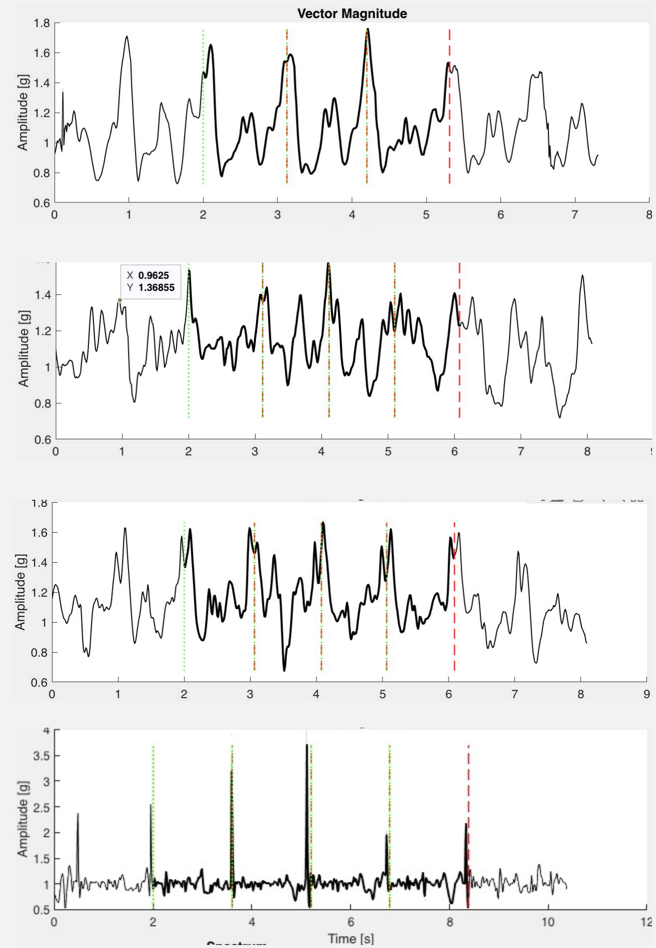
Scientific problem

- Detailed walking characteristics have become increasingly important in health studies
    - Distance covered and speed in a 6 min walk, cadence, stride pattern variability, gait symmetry[1]
- Context: supervised and semi-supervised walking
- Need for automatic, fast and accurate methods for walking strides segmentation from raw accelerometry data

Challenges

- Variations in shape and duration of a pattern within and between individuals
- Different sensor locations: wrist (left, right), lower back, hip, ankle

1: Studenski et al., 2011; Brown et al., 2014; Urbanek et al., 2017; Del Din et al., 2019



4 individuals from STURDY RCT (age mean = 77.3, SD = 5.5).

Data collected at a non-dominant wrist during a 6-minute walk with ActiGraph GT9X at 80 Hz.

ADEPT: method for automatic, fast and accurate pattern segmentation in time-series

Proposed method [1] : ADaptive Empirical Pattern Transformation (ADEPT)

- Uses a predefined pattern template function $\varphi(t)$ to detect pattern repetitions in the observed data, $x(t)$, by maximizing local similarity (e.g., covariance, correlation) between:

  a) the collection of time-translated and rescaled templates, $\left\{\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right)\right\}_{s,\tau}$;

  b) observed data $x(t)$.

- Done by iteratively identifying maxima of similarity (here: covariance) function:

$$W_{\varphi}(s,\tau) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right)dt\ ,$$

where $\varphi(t)$ is non-zero in $[0,1]$, $x(t)$ is non-zero in $[0,T]$.

1: Karas et al., 2019.

ADEPT: method for automatic, fast and accurate pattern segmentation in time-series

Proposed method [1] : ADaptive Empirical Pattern Transformation (ADEPT)

- Uses a predefined pattern template function $\varphi(t)$ to detect pattern repetitions in the observed data, $x(t)$, by maximizing local similarity (e.g., covariance, correlation) between:

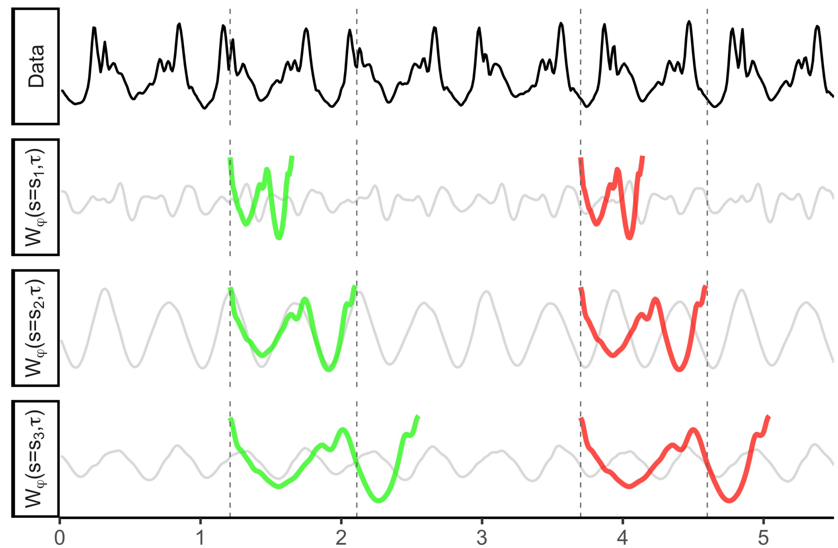
  a) the collection of time-translated and rescaled templates, $\left\{\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right)\right\}_{s,\tau}$;

  b) observed data $x(t)$.

- Done by iteratively identifying maxima of similarity (here: covariance) function:

$$W_\varphi(s,\tau) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right) dt ,$$

where $\varphi(t)$ is non-zero in $[0,1]$, $x(t)$ is non-zero in $[0,T]$.



1: Karas et al., 2019.

ADEPT: method for automatic, fast and accurate pattern segmentation in time-series

Proposed method [1] : ADaptive Empirical Pattern Transformation (ADEPT)

- Uses a predefined pattern template function $\varphi(t)$ to detect pattern repetitions in the observed data, $x(t)$, by maximizing local similarity (e.g., covariance, correlation) between:

    a) the collection of time-translated and rescaled templates, $\left\{\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right)\right\}_{s,\tau}$;

    b) observed data $x(t)$.

- Done by iteratively identifying maxima of similarity (here: covariance) function:

$$W_\varphi(s,\tau) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{s}}\varphi\left(\frac{t-\tau}{s}\right)dt \,,$$

where $\varphi(t)$ is non-zero in $[0,1]$, $x(t)$ is non-zero in $[0,T]$.

Optimized for identification of walking strides in raw accelerometry data:

- Uses 1D vector magnitude $r(t)$ of accelerometry data as $x(t)$ => Invariant to sensor rotation, and robust to sensor placement on wrist
- Template $\varphi(t)$ is data-derived, and allows population- or individual-specific templates
- Maximizes $W_\varphi(s,\tau)$ iteratively => Accounts for changes of stride cadence across time
- Allows multiple distinct templates $\varphi_1, \varphi_2, \varphi_3, \ldots$ simultaneously => Accounts for changes of stride pattern across time
- Uses location fine-tuning step => Returns precise location of start/end of a stride
- Implementation uses rolling statistics and supports parallel computing => Computational speed

1: Karas et al., 2019.

ADEPT: results

ADEPT was validated against manual strides segmentation in continuous outdoor walk

- N = 32, healthy adult
- 4 sensor locations (left wrist, left hip, both ankles)
- Excellent agreement with manual segmentation for hip and ankles, very good for wrist

From ADEPT-segmented walking strides:

- Estimated temporal cadence [steps/s] trajectory
- Estimated subject-specific stride patterns

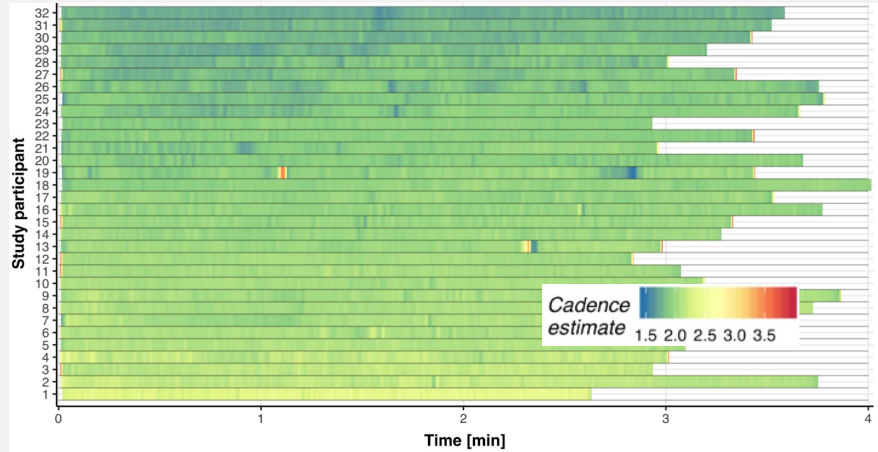


Figure 1. Walking cadence [steps/s] estimates during continuous outdoor walk (N = 32), based on raw accelerometry data collected at left ankle[1].

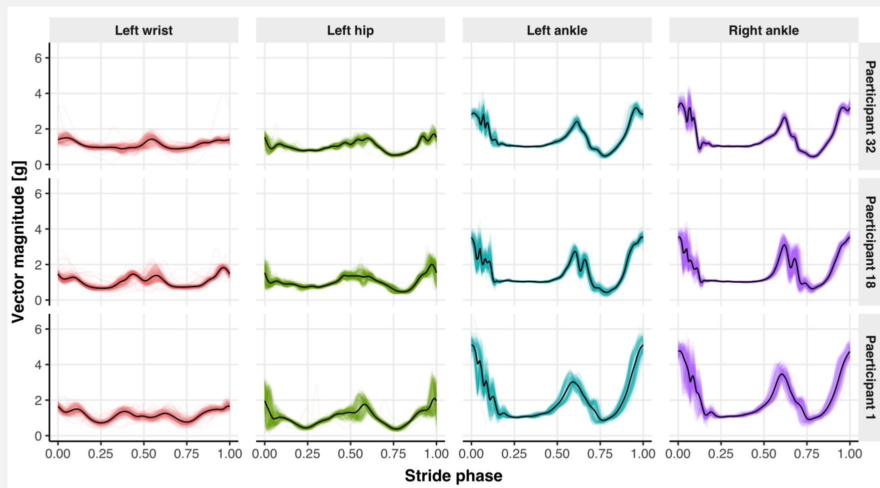


Figure 2. Examples of subject- and sensor-location specific stride patterns for three selected study participants[1].

1: Karas et al., 2019.

ADEPT extension: walking segmentation from data collected in the free-living at a wrist

Scientific problem

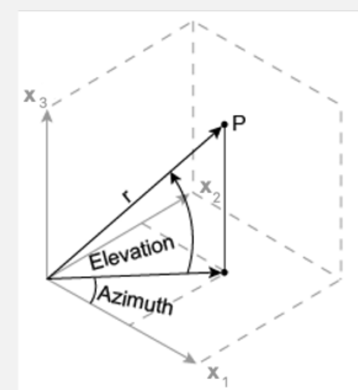
- Walking features measured in the lab are weakly associated with those from the free-living[1]
- Free-living: decreased speed, increased step variability, increased asymmetry[1]
- Need for methods to segment walking strides in the free-living environment

Challenges

- Sensor typically worn at wrist -- a challenging location for walking identification
- Validation is difficult

Proposed method[2]

- Use ADEPT for initial exhaustive segmentation of walking stride patterns
- Filter the results –- accept a pattern if:
    - (a) has high correlation with a template,
    - (b) in consecutive >=3,
    - (c) "looks alike" its neighbouring patterns
- For (c), uses transformation of raw accelerometry data from Cartesian $[x_1, x_2, x_3]$ into spherical $[az, el, r]$ coordinate system



1: Del Din et al., 2016; Mueller et al., 2019; Van Ancum et al., 2019
2: Karas et al., 2021
Figure -- left: ActiGraph (adapted), LLC. Figure – right: MathWorks (adapted).

ADEPT: summary of contributions

Methods

- **Karas et al. (2019).** Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. *Biostatistics*.
- **Karas et al. (2021).** Estimation of free-living walking cadence from wrist-worn sensor accelerometry data and its association with SF-36 quality of life scores. *Physiological Measurement*.

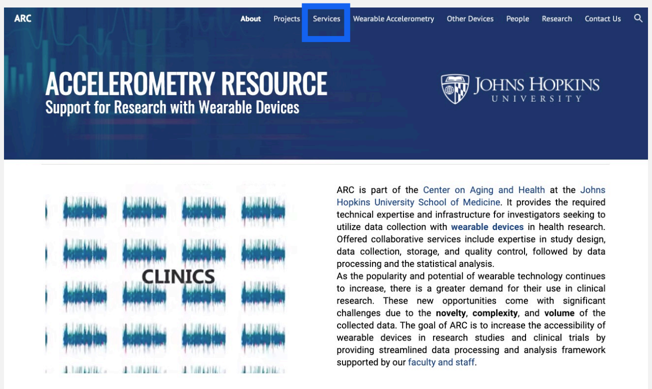
Applications

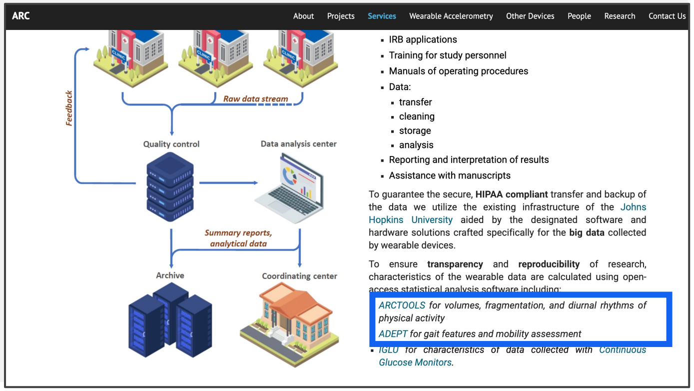
- **Karas et al. (2021):** higher free-living cadence associated with better quality of life score
- **Urbanek et al.** (revised, resubmitted): higher free-living cadence associated with lower fall rates in older individuals
- **Catallini (2020)**: In MS thesis, ADEPT used for segmentation of neuronal activity traces from time series of calcium imaging
- **Rubin lab** (U of Chicago): Integrating ADEPT for iOS for semi-supervised experiments
- **Qiao** (U of Pittsburgh; PhD thesis): novel markers to identify performance fatigability during a fast-paced 400 m walk

Resources

- adept R package:
  - Implements ADEPT and its extension for free-living
  - Data examples and tutorials

https://www.accelerometry.org/





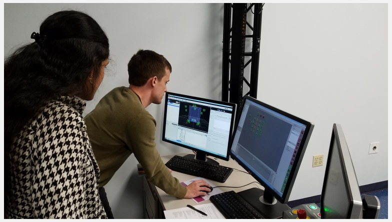
ADEPT: potential future directions

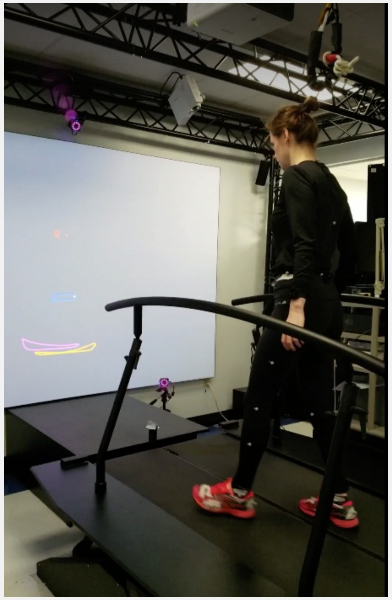
1. Functional registration of individual walking stride patterns (characterization of gait asymmetry etc.)

2. Functional registration of temporal walking characteristics from standardized tests (e.g. cadence in 6 minute walk test)

At Center For Movement Studies, Kennedy Krieger Institute, Baltimore with

Drs Purnima Padmanabhan and Ryan Roemmich





Harmonization of accelerometry-based measures of physical activity

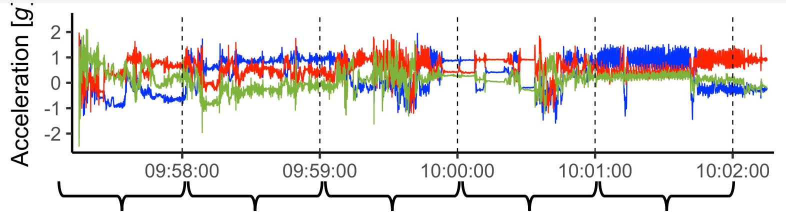
Karas, M.*, Muschelli, J. *, Leroux, A., Urbanek, J.K., Wanigatunga, A.A., Bai, J., Crainiceanu, C.M., Schrack, J.A. (2021). Submitted.

* : Shared co-first authorship.

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Scientific problem

- Summary measures of raw accelerometry data are commonly used in health research[1] to characterize physical activity
- Widely-used: ActiGraph "activity counts" (**AC**)
  - ActiGraph hardware and licensed software needed to derive from raw data
- Recently, open-source statistics have been proposed to aggregate raw data: **MIMS, ENMO, MAD, AI**[2]
- Comparability to previously published research is unknown
  - **AC** cut-off points, **AC** population quantiles



Measure	Based on
MIMS	AUC of interpolated, extrapolated, bandpass-filtered $x_m(t)$; then added across axes $m = 1,2,3$
ENMO	Mean of $r(t)$ vector magnitude from pre-calibrated raw data $[x_1(t), x_2(t), x_3(t)]$
MAD	Mean amplitude deviation of $r(t)$ vector magnitude
AI	Variance of $x_m(t)$; then averaged across axes $m = 1,2,3$

MIMS -- Monitor-Independent Movement Summary
ENMO -- Euclidean Norm Minus One
MAD -- Mean Amplitude Deviation
AI -- Activity Index

1: Karas et al., 2019b.
2: MIMS: John et al., 2019; ENMO: van Hees et al., 2013; MAD: Vähä-Ypyä et al., 2015; AI: Bai et al., 2012

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**

2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI

3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping

4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**
2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping
4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Challenges

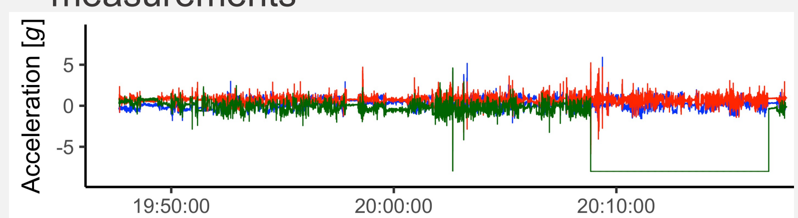
- Large volume of raw accelerometry data needs quality check
  - 700 participants x 7 days x 1440 minutes x 60 seconds x 80 obs./s x 3 sensor axes = 101,606,400,000 (one hundred billion+)

Methods

- Adapted raw data quality flags from recently published NHANES protocol[1]
- Implemented flags to detect acceleration spikes, values at the sensor's dynamic range

Results

- Identified few flagged cases of raw measurements



1: NHANES 2011-2012 Data Documentation (2020)

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**
2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping
4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Methods

- Linear regression with subject-specific correlation between measures as an outcome

Results

- Very high correlation between: AC and MIMS, AC and AI
- Significant but small effects of covariate(s)

	Unadjust. model	Model adjusted for: age, BMI, sex			
	Intercept	Intercept	Age	BMI	Sex (is male)
Response var.	Coef. est. (se)	Coef. est. (se)	Coef. est. (se)	Coef. est. (se)	Coef. est. (se)
corr (AC, MIMS)	0.988 (0.0002)	0.988 (0.0017)	< 0.001 (<0.0001)	< 0.001 (<0.0001)	-0.002 (0.0005)*
corr (AC, ENMO)	0.867 (0.0018)	0.887 (0.0138)	-0.001 (0.0001)*	0.001 (0.0004)	> -0.001 (0.0037)
corr (AC, MAD)	0.913 (0.0013)	0.892 (0.0099)	< 0.001 (0.0001)	0.001 (0.0003)*	-0.010 (0.0026)*
corr (AC, AI)	0.970 (0.0007)	0.962 (0.0050)	< 0.001 (< 0.0001)	< 0.001 (0.0001)*	-0.010 (0.0013)*

Table. "*" symbol is used to denote model coefficients (excluding intercept) for which the corresponding p-value was <0.05.

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**

2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI

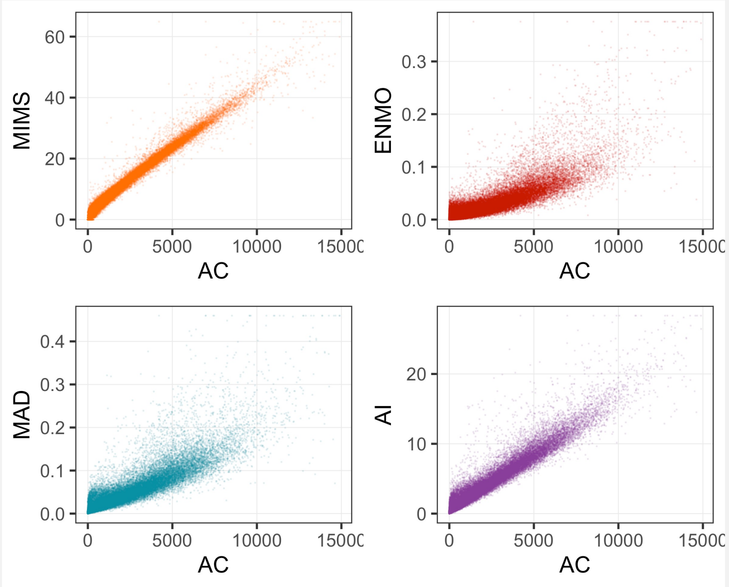
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping

4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Challenges

- Estimate the relation between pairs of minute-level measures $(x_{ij}(t), y_{ij}(t))$ e.g., $(AC_{ij}(t), MIMS_{ij}(t))$ as a smooth function f while accounting for correlation structure (i-th participant, j-th day, t-th minute)

- Volume of minute-level data = 700 participants x 7 days x 1440 minutes = 7,056,000



Showing 1% of the data.

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**
2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping
4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Methods

- Estimated f via additive model $y_{ij}(t) = f(x_{ij}(t)) + \varepsilon_{ij}(t)$, assuming independence
- Used "case bootstrap"[1] (sampling units at the highest level and then sampling within these units without replacement) to get 95% CI for $\hat{f}$
- Used $\hat{f}$ to define one-to-one harmonization mapping
- Evaluated $\hat{f}$ in tasks of: (a) predicting total AC, (b) classifying a minute into active vs non-active

1: Ren et al., 2010

Harmonization of open-source and proprietary accelerometry-based physical activity measures

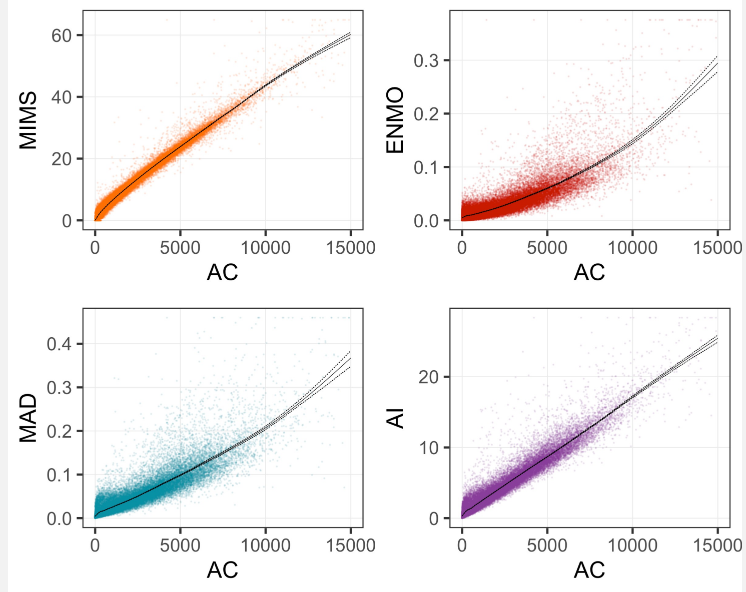
Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**
2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping
4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Results

Black solid line: $\hat{f}(AC)$, dashed lines: 95% CI



AC cut-off proposed in	AC	$\hat{f}_{MIMS}(AC)$ (95% CI)
Koster (2016)	1853	10.56 [10.53, 10.59]
Montoye (2020)	2860	15.05 [15.02, 15.07]
Montoye (2020)	3940	19.61 [19.58, 19.65]

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

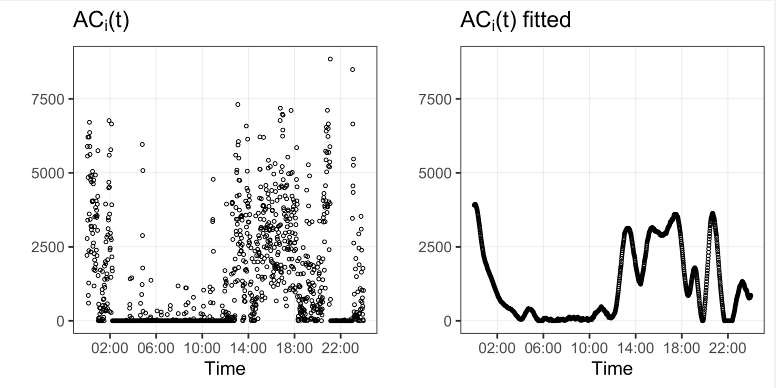
1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**
2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI
3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping
4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures

Challenges

- Required to impute minute-level data missingness (up to 10% per 24 h)

Methods

- For each measure separately, use FPCA model $Y_i(t) = \mu(t) + \sum_{k=1}^{ncp} \xi_{ik}\phi_k(t) + \varepsilon_i(t)$ to estimate $\widehat{Y}_i(t)$ -- smoothed (fitted) version of each i–th participant-day functional observation[1]
- Use $\widehat{Y}_i(t)$ for imputation



1: Leroux et al., 2020

Harmonization of open-source and proprietary accelerometry-based physical activity measures

Contributions

Data from ~700 participants in the Baltimore Longitudinal Study on Aging (BLSA), each monitored for a week with a wrist-worn PA sensor

1. Summarized raw data at minute-level: **AC** and open-source **MIMS**, **ENMO**, **MAD**, **AI**

2. Quantified association between **AC** and open-source measures marginally and conditionally on age, sex and BMI

3. Harmonized minute-level **AC** with open-source measures via one-to-one mapping

4. Reproduced some of the published BLSA results that used **AC** with the use of the open-source measures[1]

Results

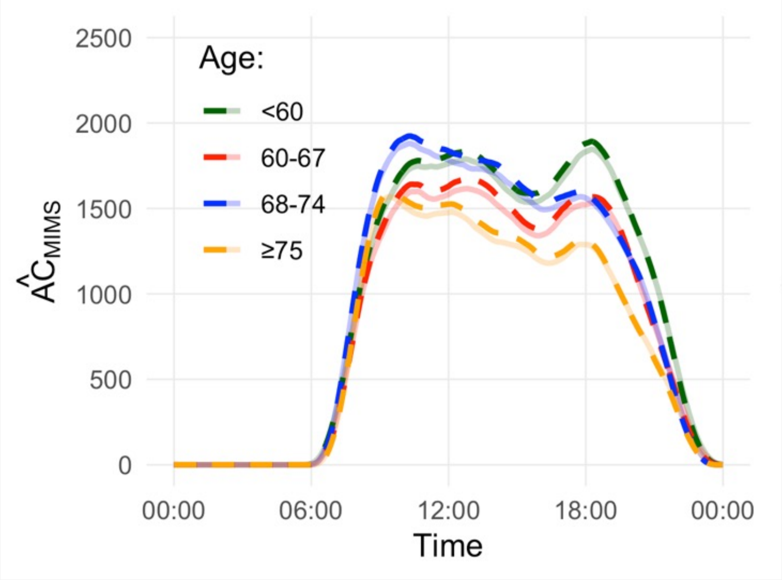


Figure. Smoothed 24-hour median activity counts per minute across four age group. Solid semi-transparent colour lines: AC. Dashed colour lines: results obtained with $\widehat{\mathrm{AC}}$ values mapped into AC from MIMS using the harmonization mapping.

1: Schrack et al., 2013

Summary measures of physical activity: potential future directions

1. Harmonization of data from large- and mega-size studies collecting minute-level measurements of physical activity
2. Data imputation for continuously collected measurements of physical activity

BLSA BALTIMORE LONGITUDINAL STUDY OF AGING

AC

nhats National Health & Aging Trends Study

how daily life changes as we age

AC

biobank uk

Improving the health of future generations

ENMO

nhanes

National Health and Nutrition Examination Survey

MIMS

NHANES most recent release (cohorts 2011-2012, 2013-2014)

Upstrap for estimating power and sample size in complex models

Karas, M., Crainiceanu, C.M. (2021). Submitted.
https://doi.org/10.1101/2021.08.21.457220

Upstrap method for power and sample size estimation in complex models

Scientific problem

Given an observed data sample $\mathbf{x}$ of sample size N, a null and an alternative hypothesis and a test statistic, assuming significance level α,

- estimate the sample size M required to achieve power $(1-\beta)$ (i.e., to achieve probability of rejecting the null hypothesis when the null is true)

Here, we consider power to detect:

- (a) an effect size observed in the sample $\mathbf{x}$;
- (b) an effect size chosen by a researcher.

and aim to address complex settings, including testing significance of model coefficients in:
LM, GLM, LMM, GLMM.

Upstrap method for power and sample size estimation in complex models

Scientific problem

Given an observed data sample $\mathbf{x}$ of sample size N, a null and an alternative hypothesis and a test statistic, assuming significance level α,

- estimate the sample size M required to achieve power $(1 - \beta)$ (i.e., to achieve probability of rejecting the null hypothesis when the null is true)

Here, we consider power to detect:

- (a) an effect size observed in the sample $\mathbf{x}$;
- (b) an effect size chosen by a researcher.

and aim to address complex settings, including testing significance of model coefficients in: LM, GLM, LMM, GLMM.

```
# Example: one-sample t-test

set.seed(123)

# simulate observed sample
x <- rnorm(n = 30, mean = 0.2, sd = 1)

mean(x)
# [1] 0.1528962

# effect size observed in sample x
("observed power" estimation)
power.t.test(n = 30, delta = mean(x),
 sd = sd(x), type = "one.sample")$power
# [1] 0.1283329
```

Upstrap method for power and sample size estimation in complex models

Scientific problem

Given an observed data sample $\mathbf{x}$ of sample size N, a null and an alternative hypothesis and a test statistic, assuming significance level α,

- estimate the sample size M required to achieve power $(1 - \beta)$ (i.e., to achieve probability of rejecting the null hypothesis when the null is true)

Here, we consider power to detect:

- (a) an effect size observed in the sample $\mathbf{x}$;
- (b) an effect size chosen by a researcher.

and aim to address complex settings, including testing significance of model coefficients in: LM, GLM, LMM, GLMM.

```
# Example: one-sample t-test

set.seed(123)

# simulate observed sample
x <- rnorm(n = 30, mean = 0.2, sd = 1)

mean(x)
# [1] 0.1528962

# effect size observed in sample x
("observed power" estimation)
power.t.test(n = 30, delta = mean(x),
 sd = sd(x), type = "one.sample")$power
# [1] 0.1283329

# effect size chosen by researcher
(here: 0.1)
power.t.test(n = 30, delta = 0.1,
 sd = sd(x), type = "one.sample")$power
# [1] 0.07781938
```

Upstrap method for power and sample size estimation in complex models

Methods

- Upstrap -- resampling with replacement fewer or more observations than in original sample **x** -- was proposed[1] as a general solution
    - For effect size observed in the sample x
    - Not evaluated in numerical experiments
- Contributions of this project:
    - Extend upstrap approach for estimating power to detect an effect size chosen by a researcher
    - Evaluate method in a simulation study

```
# Example: bootstrap vs upstrap

# observed sample
x <- 1:10

# bootstrap resample
xb <- sample(x, size = 10, replace = TRUE)

# upstrap resample
# (sample size larger than in original x)
xu1 <- sample(x, size = 20, replace = TRUE)

# upstrap resample
# (sample size smaller than in original x)
xu2 <- sample(x, size = 8, replace = TRUE)
```

1: Crainiceanu and Crainiceanu (2020)

Upstrap-based algorithm for estimating power

Methods

Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$

1. Generate B upstrap resamples of size M;
2. Perform hypothesis test on each resample;
3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.

- Case (b): effect size chosen by researcher
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$
  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.
- Case (b): effect size chosen by researcher
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

Toy example: one-sample t-test

```
# simulate observed sample
x <- rnorm(n = 30, mean = 0.2, sd = 1)


# Case (a): effect size observed

out <- rep(NA, B)
for (bb in 1 : B){
  x_bb <- sample(x, size = M, replace = T)
  out[bb] <- (t.test(x_bb)$p.value < 0.05)
}
mean(out)
```

1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$
  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.
- Case (b): effect size chosen by researcher
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

Toy example: one-sample t-test

```
# simulate observed sample
x <- rnorm(n = 30, mean = 0.2, sd = 1)


# Case (a): effect size observed

out <- rep(NA, B)
for (bb in 1 : B){
  x_bb <- sample(x, size = M, replace = T)
  out[bb] <- (t.test(x_bb)$p.value < 0.05)
}
mean(out)


# Case (b): effect size chosen to 0.4

x <- x + (0.4 - mean(x))
out <- rep(NA, B)
for (bb in 1 : B){
  x_bb <- sample(x, size = M, replace = T)
  out[bb] <- (t.test(x_bb)$p.value < 0.05)
}
mean(out)
```

1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

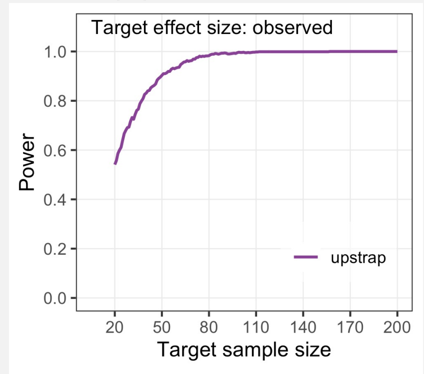
Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$
  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.
- Case (b): effect size chosen by researcher
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

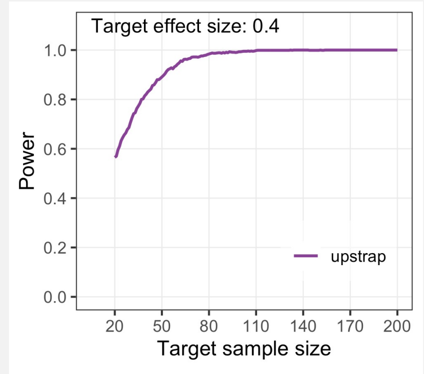
Toy example: one-sample t-test

Consider: $\mathrm{x_i} \sim_{\mathrm{iid}} \mathcal{N}(0.3, 1)$, $\mathrm{i} = 1, \ldots, 50$,
1 repetition

Case (a)



Case (b)



1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

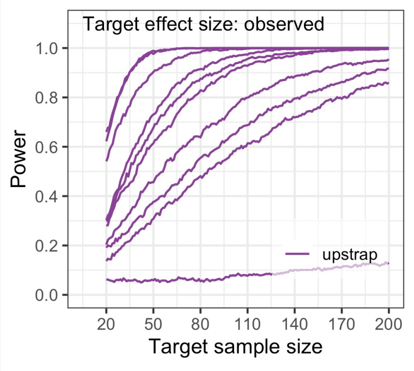
Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$
  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.
- Case (b): effect size chosen by researcher
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

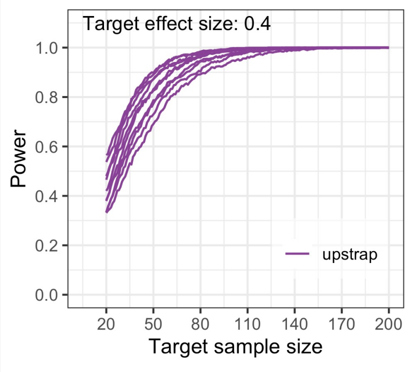
Toy example: one-sample t-test

Consider: $\mathrm{x_i} \sim_{\mathrm{iid}} \mathcal{N}(0.3, 1),\ \mathrm{i} = 1, \dots, 50,$
15 repetitions

Case (a)



Case (b)



1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$

  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.

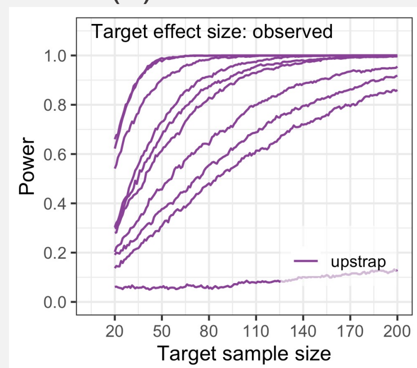
- Case (b): effect size chosen by researcher

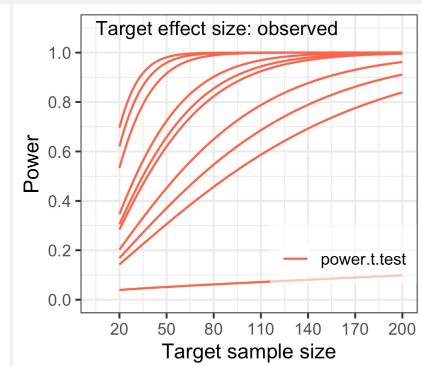
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

Toy example: one-sample t-test

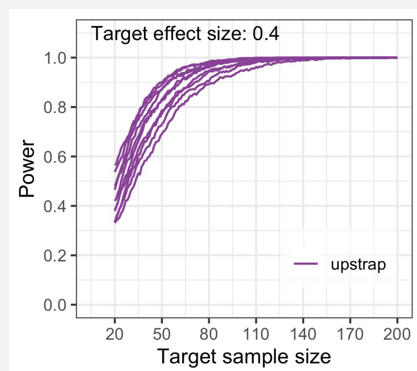
Consider: $\mathrm{x_i} \sim_{\mathrm{iid}} \mathcal{N}(0.3, 1)$, $\mathrm{i} = 1, \ldots, 50$,
15 repetitions

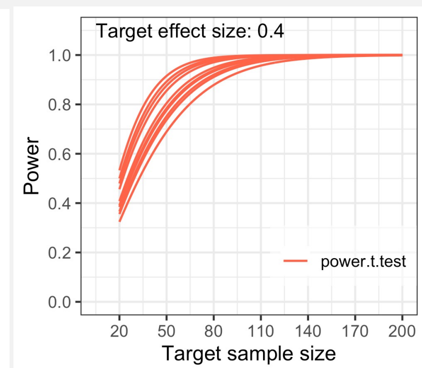
Case (a)





Case (b)





1: Details in the preprint

Upstrap-based algorithm for estimating power

Methods

Given observed data sample $\mathbf{x}$ of size N, to estimate power for a target sample size M:

- Case (a): effect size observed in the sample $\mathbf{x}$

  1. Generate B upstrap resamples of size M;
  2. Perform hypothesis test on each resample;
  3. Estimate power as the proportion of B resamples where the null hypothesis was rejected.

- Case (b): effect size chosen by researcher

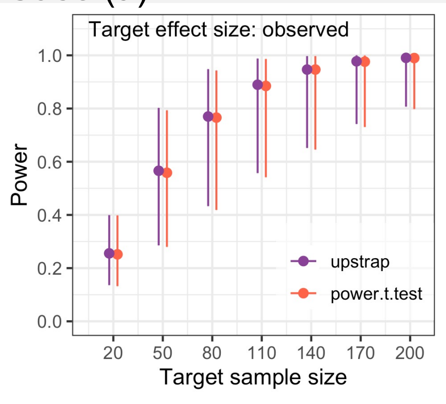
  - As above (a), but update response variable values in sample $\mathbf{x}$ (/resample) so as the updated sample (/resample) reflects the target effect size[1]

1: Details in the preprint

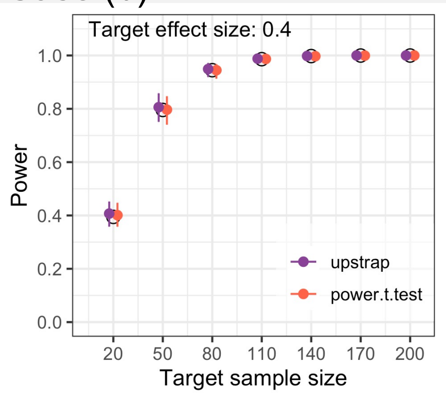
Toy example: one-sample t-test

Consider: $x_i \sim_{iid} \mathcal{N}(0.3, 1),\ i = 1, \ldots, 50$,
1000 repetitions, aggregated power values (mean, 25th and 75th percentiles)

Case (a)



Case (b)



Upstrap-based algorithm for estimating power: simulation study

- Summary of the simulation setup across six different problems:

	Data-generating model (effect being tested highlighted in color)	Observed sample size	Target effect size	Comparator to upstrap
1	$Y_i = \beta_0 + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
2	$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
3	$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$	50	0.5, 1, observed	SIMR
4	$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$	50	0.5, 1, observed	SIMR
5	$Y_{ij} = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \varepsilon_{ij}$	50	0.5, 1, observed	SIMR
6	$\text{logit}(\pi_{ij}) = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij}$	50	0.5, 1, observed	SIMR

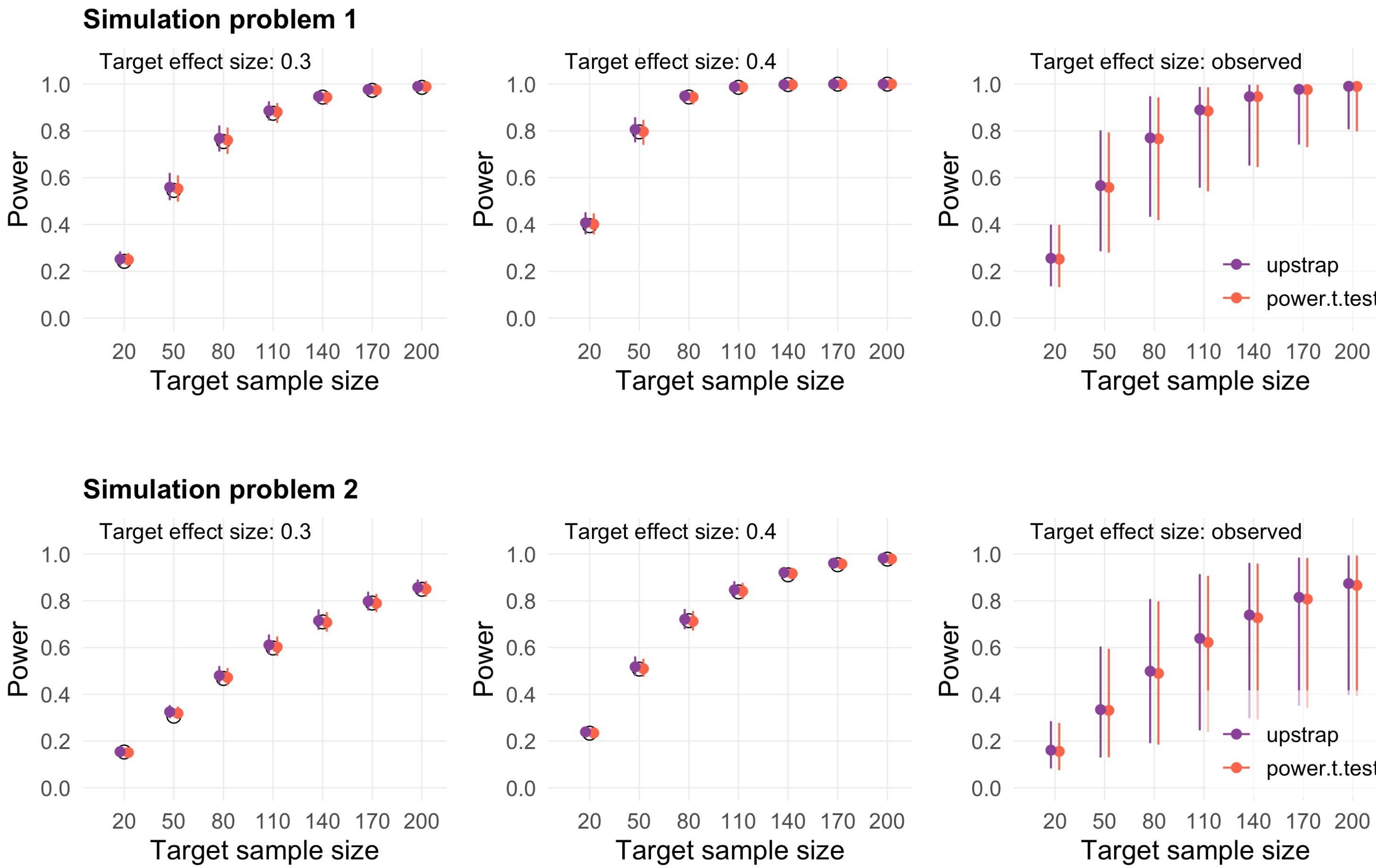
- In each of the six problems:
  - X_{1i} / X_{1ij} is defined as dichotomous variable
  - 1,000 independent experiment repetitions (generating a sample and power estimation)
  - Two-sided test used to test H_0: $\beta = 0$ versus H_1: $\beta \neq 0$ at significance level $\alpha = 0.05$
  - "True power" estimated by proportion of null rejected from 10,000 samples

Upstrap-based algorithm for estimating power: simulation study

- Summary of the simulation setup across six different problems:

	Data-generating model (effect being tested highlighted in color)	Observed sample size	Target effect size	Comparator to upstrap
1	$Y_i = \beta_0 + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
2	$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
3	$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$	50	0.5, 1, observed	SIMR[1]
4	$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$	50	0.5, 1, observed	SIMR
5	$Y_{ij} = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \varepsilon_{ij}$	50	0.5, 1, observed	SIMR
6	$\text{logit}(\pi_{ij}) = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij}$	50	0.5, 1, observed	SIMR

- In each of the six problems:
  - X_{1i} / X_{1ij} is defined as dichotomous variable
  - 1,000 independent experiment repetitions (generating a sample and power estimation)
  - Two-sided test used to test H_0: $\beta = 0$ versus H_1: $\beta \neq 0$ at significance level $\alpha = 0.05$
  - "True power" estimated by proportion of null rejected from 10,000 samples

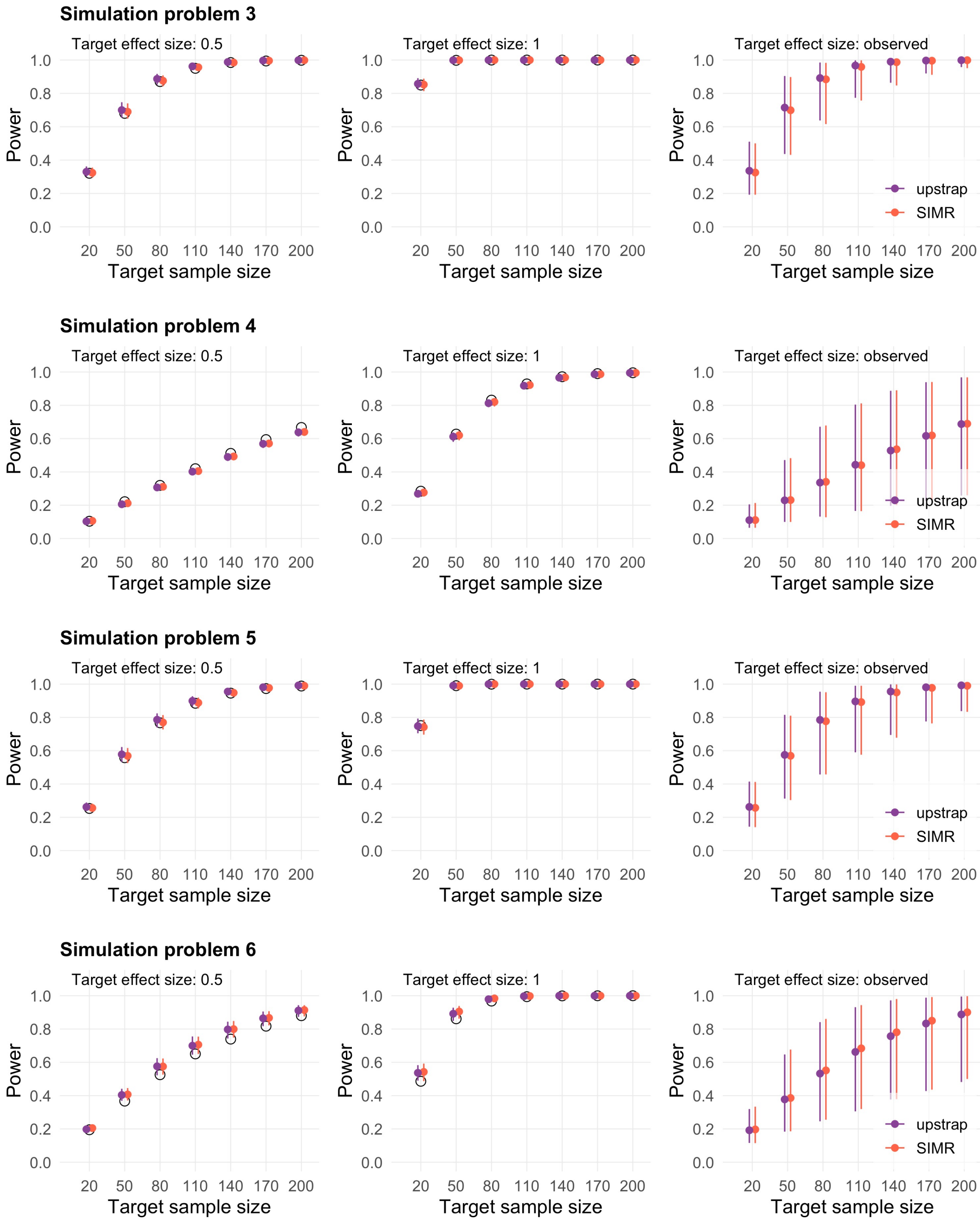

Simulation problem 1
Target effect size: 0.3
Target effect size: 0.4
Target effect size: observed
Power
1.0
0.8
0.6
0.4
0.2
0.0
20 50 80 110 140 170 200
Target sample size
upstrap
power.t.test
Simulation problem 2
Target effect size: 0.3
Target effect size: 0.4
Target effect size: observed
Power
Target sample size
upstrap
power.t.test

Upstrap-based algorithm for estimating power: simulation study

- Summary of the simulation setup across six different problems:

	Data-generating model (effect being tested highlighted in color)	Observed sample size	Target effect size	Comparator to upstrap
1	$Y_i = \beta_0 + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
2	$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$	50	0.3, 0.4, observed	power.t.test()
3	$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$	50	0.5, 1, observed	SIMR
4	$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$	50	0.5, 1, observed	SIMR
5	$Y_{ij} = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \varepsilon_{ij}$	50	0.5, 1, observed	SIMR
6	$\text{logit}(\pi_{ij}) = b_{0i} + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij}$	50	0.5, 1, observed	SIMR

- In each of the six problems:
  - X_{1i} / X_{1ij} is defined as dichotomous variable
  - 1,000 independent experiment repetitions (generating a sample and power estimation)
  - Two-sided test used to test H_0: $\beta = 0$ versus H_1: $\beta \neq 0$ at significance level $\alpha = 0.05$
  - "True power" estimated by proportion of null rejected from 10,000 samples


Simulation problem 3
Target effect size: 0.5
Target effect size: 1
Target effect size: observed
upstrap
SIMR
Simulation problem 4
Target effect size: 0.5
Target effect size: 1
Target effect size: observed
upstrap
SIMR
Simulation problem 5
Target effect size: 0.5
Target effect size: 1
Target effect size: observed
upstrap
SIMR
Simulation problem 6
Target effect size: 0.5
Target effect size: 1
Target effect size: observed
upstrap
SIMR
Power
0.0
0.2
0.4
0.6
0.8
1.0
20 50 80 110 140 170 200
Target sample size

Upstrap-based algorithm for estimating power: summary

- Simulation results:
  - For one- and two-sample t-test, the upstrap performed essentially identical to the well-established analytical solutions for power estimation
  - In complex scenarios, the upstrap performed similarly the existing method from SIMR R package; both approaches demonstrated very high agreement with the true power estimates
- The upstrap method is "read-and-use"
  - It can be implemented by any analyst who is familiar with software allowing to: (a) resample data, (b) run the statistical test of interest

Upstrap method for power and sample size estimation: potential future directions

1. Upstrap for estimating power to detect an effect while (a) preserving or (b) changing covariate class proportions

2. Resampling (up/down) cluster-specific observations in longitudinal data

Summary of contributions

The main methodological contributions of this thesis are:

1. development and validation of ADEPT, a novel statistical pattern-segmentation method;
2. introduction of harmonization methods of objective summary measures of physical activity;
3. study of the upstrap properties in complex scenarios;
4. development of four R software packages.

R software packages:

- **runstats**: Fast Computation of Running Statistics for Time Series (CRAN)
- **adept**: Adaptive Empirical Pattern Transformation (CRAN)
- **adeptdata**: Accelerometry Data Sets (CRAN)
- **arctools**: Processing and Physical Activity Summaries of Minute Level Activity Data (CRAN)

All projects code publicly available:

- ADEPT (GitHub repo 1, GitHub repo 2)
- Harmonization of measures (GitHub repo)
- Upstrap (GitHub repo)

Reviewer for:

- Plos One, Digital Biomarkers, Scandinavian Journal of Medicine & Science in Sports (Publons)

References

1. Bai, J., Di, C., Xiao, L., Evenson, K. R., LaCroix, A. Z., Crainiceanu, C. M., & Buchner, D. M. (2016). An Activity Index for raw accelerometry data and its comparison with other activity metrics. PLOS ONE, 11(8), 1–14. https://doi.org/10.1371/journal.pone.0160644

2. Brown, J. C., Harhay, M. O., & Harhay, M. N. (2014). Walking cadence and mortality among community-dwelling older adults. Journal of General Internal Medicine, 29(9), 1263–1269. https://doi.org/10.1007/s11606-014-2926-6

3. Crainiceanu, C. M., & Crainiceanu, A. (2018). The upstrap. Biostatistics, 21(2), e164–e166. https://doi.org/10.1093/biostatistics/kxy054

4. Del Din, S., Godfrey, A., Galna, B., Lord, S., & Rochester, L. (2016). Freeliving gait characteristics in ageing and parkinson's disease: Impact of environment and ambulatory bout length. Journal of NeuroEngineering and Rehabilitation, 13(46). https://doi.org/10.1186/s12984-016-0154-5

5. Del Din, S., Elshehabi, M., Galna, B., Hobert, M. A., Warmerdam, E., Suenkel, U., Brockmann, K., Metzger, F., Hansen, C., Berg, D., Rochester, L., & Maetzler, W. (2019). Gait analysis with wearables predicts conversion to Parkinson disease. Annals of Neurology, 86(3), 357–367. https://doi.org/10.1002/ana.25548

6. Green, P., MacLeod, C.J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution 7(4), 493–498.

7. John, D., Tang, Q., Fahd, A., & Intille, S. (2019). An open-source MonitorIndependent Movement Summary for accelerometer data processing. Journal of Measuring Physical Behaviors, 2(4), 268–281. https://doi.org/10.1123/jmpb.2018-0068

8. **Karas, M**., Bai, J., Straczkiewicz, M., Harezlak, J., Glynn, N. W., Harris, T., Zipunnikov, V., Crainiceanu, C., & Urbanek, J. K. (2019). Accelerometry data in health research: Challenges and opportunities. review and examples. Statistics in Biosciences, 11, 210–237. https://doi.org/10.1007/ s12561-018-9227-2

9. **Karas, M.,** Straczkiewicz, M., Fadel, W., Harezlak, J., Crainiceanu, C. M., & Urbanek, J. K. (2019). Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. Biostatistics, 22(2), 331–347. https://doi.org/10.1093/biostatistics/kxz033

10. **Karas, M.,** Urbanek, J. K. U., Illiano, V. P., Bogaarts, G., Crainiceanu, C. M., Dorn, J. F. (2021). Estimation of free-living walking cadence from wrist-worn sensor accelerometry data and its association with SF-36 quality of life scores. Physiological measurement, 42(6). https://doi.org/10.1088/1361-6579/ac067b

11. **Karas, M.*,** Muschelli, J.*, Leroux, A., Urbanek, J.K., Wanigatunga, A.A., Bai, J., Crainiceanu, C.M., Schrack, J.A. (2021). Harmonization of accelerometry-based measures of physical activity. Submitted

12. **Karas, M**., Crainiceanu, C.M. (2021). Upstrap for estimating power and sample size in complex models. Submitted. https://doi.org/10.1101/2021.08.21.457220

13. Leroux, A., Xu, S., Kundu, P., Muschelli, J., Smirnova, E., Chatterjee, N., & Crainiceanu, C. (2020). Quantifying the predictive performance of objectively measured physical activity on mortality in the UK Biobank. The Journals of Gerontology: Series A, 76(8), 1486–1494. https://doi.org/ 10.1093/gerona/glaa250

13. Mueller, A., Hoefling, H. A., Muaremi, A., Praestgaard, J., Walsh, L. C., Bunte, O., Huber, R. M., Fürmetz, J., Keppler, A. M., Schieker, M., Böcker, W., Roubenoff, R., Brachat, S., Rooks, D. S., & Clay, I. (2019). Continuous digital monitoring of walking speed in frail elderly patients: Noninterventional validation study and longitudinal clinical trial. JMIR mHealth and uHealth, 7(11), e15191. https://doi.org/10.2196/15191

14. NHANES 2011-2012 Data Documentation. (2020). National Health and Nutrition Examination Survey 2011-2012. Data Documentation, Codebook, and Frequencies: Quality Assurance and Quality Control. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/PAXMIN_G.htm

15. Ren, S., Lai, H., Tong, W., Aminzadeh, M., Hou, X., & Lai, S. (2010). Nonparametric bootstrapping for hierarchical data. Journal of Applied Statistics, 37(9), 1487–1498. https://doi.org/10.1080/02664760903046102

16. Straczkiewicz, M., Glynn, N. W., & Harezlak, J. (2019). On placement, location and orientation of wrist-worn tri-axial accelerometers during free-living measurements. Sensors, 19(9), 2095. https://doi.org/10.3390/s19092095

17. Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., Brach, J., Chandler, J., Cawthon, P., Connor, E. B., Nevitt, M., Visser, M., Kritchevsky, S., Badinelli, S., Harris, T., Newman, A. B., Cauley, J., Ferrucci, L., & Guralnik, J. (2011). Gait speed and survival in older adults. JAMA: the Journal of the American Medical Association, 305(1), 50–58. https://doi.org/10.1001/jama.2010.1923

18. Schrack, J. A., Zipunnikov, V., Goldsmith, J., Bai, J., Simonsick, E. M., Crainiceanu, C., & Ferrucci, L. (2013). Assessing the "physical cliff": Detailed quantification of age-related differences in daily patterns of physical activity. The Journals of Gerontolo

19. Urbanek, J. K., Harezlak, J., Glynn, N. W., Harris, T., Crainiceanu, C., & Zipunnikov, V. (2017). Stride variability measures derived from wrist- and hip-worn accelerometers. Gait and Posture, 52, 217–223. https://doi.org/10.1016/j.gaitpost.2016.11.045

20. Urbanek, J. K., Roth, D. L., **Karas, M**., Wanigatunga, A. A., Mitchell, C. M., Juraschek, S. P., Cai, Y., Appel, L. J., Schrack, J. A. (2021). Free-living gait cadence measured by wearable accelerometer: a promising alternative to traditional measures of mobility for assessing fall risk. Revised, resubmitted.

21. van Hees, V. T., Gorzelniak, L., Dean León, E. C., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renström, F., Franks, P. W., Horsch, A., & Brage, S. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS ONE, 8(4), e61691. https://doi.org/10.1371/journal.pone.0061691

22. Vähä-Ypyä, H., Vasankari, T., Husu, P., Suni, J., & Sievänen, H. (2015). A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. Clinical Physiology and Functional Imaging, 35(1), 64–70. https://doi.org/10.1111/cpf.12127

* : Shared co-first authorship.

Acknowledgements [1/4]

I am grateful to everyone who has taught, mentored, helped, and supported me over the last 4+ years.

- My academic advisor, Ciprian Crainiceanu
- My academic co-advisor, Jacek Urbanek
- Thesis Committee Primary Readers: Vadim Zipunnikov, Jennifer Schrack, and Alternates: Jiawei Bai, Adam Spira
- Wearable and Implantable Technology (WIT) group members, including: Andrew Leroux, Erjia Cui, Jiawei Bai, John Muschelli, Junrui Di, Lacey Etzkorn
- ENGAGE group members, including: Amal Wanigatunga, Fangyu Liu, Francesca Marino, Nikki Shen, Perry Kuo
- Fellow students, including: Albert Kuo, Athena Chen, Debangan Dey, Elizabeth Sarker, Haley Grant, Ji Soo Kim, Lauren Lan, Linda Gai, Lucy Qiao, Trang Nguyen, Yuji Chen
- Department of Biostatistics, including: Debra Moffitt, Fallon Bachman, Karen Bandeen-Roche, Maria Beeson, Mark Miller, Marti Gilbert, Mary Joy Argo, Patty Hubbard
- My class teachers and teaching assistants

Acknowledgements [2/4]

I am grateful to everyone who has taught, mentored, helped, and supported me over the last 4+ years.

- My industry internship advisors, Ieuan Clay and Jonas Dorn
- My MS thesis advisor, Malgorzata Bogdan, and my first research advisor in the USA, Jaroslaw Harezlak
- My friend, Maria Kudela
- My family in USA, including: Anna Rykojc, Stanislaw Rykojc, Michelle Rykojc, Mark Rykojc
- My partner, Osman Veledar, and his family, including Neda Veledar and Sead Veledar
- My friends: Haroldo Rodriguez, Lacey Etzkorn, Magdalena Urbanek, Michael Mengyang Gu, Paulina Banas, Paulina Chalan, Zohaib Akhtar

Acknowledgements [3/4]

I am grateful to everyone who has taught, mentored, helped, and supported me over the last 4+ years.

- My sister, Anna Karas
- My friends in Poland: Agnieszka Rackiewicz, Andzelika Pajchert-Kozlowska, Cezary Dynak, Katarzyna Perska, Marta Herman, Marta Koruba, Michal Kielbowicz, Olga Graca-Pazura, Olga Golacinska

Acknowledgements [4/4]

I am grateful to everyone who has taught, mentored, helped, and supported me over the last 4+ years.

- My parents, Malgorzata Karas and Dariusz Karas

Thank you!


SCHOOL OF HYGIENE AND PVBLIC HEALTH
JOHNS HOPKINS
BLOOMBERG SCHOOL OF PUBLIC HEALTH
JOHNS HOPKINS
100 YEARS AND BEYOND
JOHNS HOPKINS
BLOOMBERG SCHOOL OF PUBLIC HEALTH
615


JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH