Wearable devices and self-reported Patient Reported Outcomes (PROs) via mobile apps with applications to observational and clinical trials

Ciprian Crainiceanu

Professor of Biostatistics

www.ciprianstats.org

Twitter: @ciprianstats

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Disclaimer

Professor Crainiceanu is consulting for Bayer, Johnson and Johnson, and Cytel on methods development for wearable and implantable technologies. The details of these contracts are disclosed through the Johns Hopkins University eDisclose system. Several devices used by Dr. Crainiceanu during research are identified in the presentation for illustration purposes. Professor Crainiceanu does not have any direct or indirect financial interest related to any of the devices discussed in this talk.

What
kind of
sensors?

# Ranking predictors of five-year all-cause mortality in the US

Rank	Variable	AUC	Rank	Variable	AUC
1	TAC	0.770	16	sPC6	0.657
2	Age	0.757	17	$TLAC_{6-8am}$	0.633
3	$TLAC_{8-10pm}$	0.753	18	Education	0.611
4	MVPA	0.748	19	Drinking	0.593
5	$TLAC_{4-6pm}$	0.740	20	Smoking	0.574
6	$TLAC_{12-2pm}$	0.735	21	CHF	0.569
7	ASTP	0.734	22	BMI	0.550
8	$TLAC_{10am-12pm}$	0.734	23	Cancer	0.559
9	$TLAC_{2-4pm}$	0.730	24	Diabetes	0.556
10	ST	0.728	25	Gender	0.554
11	TLAC	0.722	26	Stroke	0.548
12	$TLAC_{8-10am}$	0.684	27	CHD	0.548
13	Mobil. Prob.	0.672	28	Race	0.514
14	$TLAC_{8-10pm}$	0.671	29	$TLAC_{12am-2am}$	0.519
15	SATP	0.660	30	Wear time	0.459

NHANES 2003-2006, age: 50-84, total: 2969, cases: 294

Micro- and macro-level data
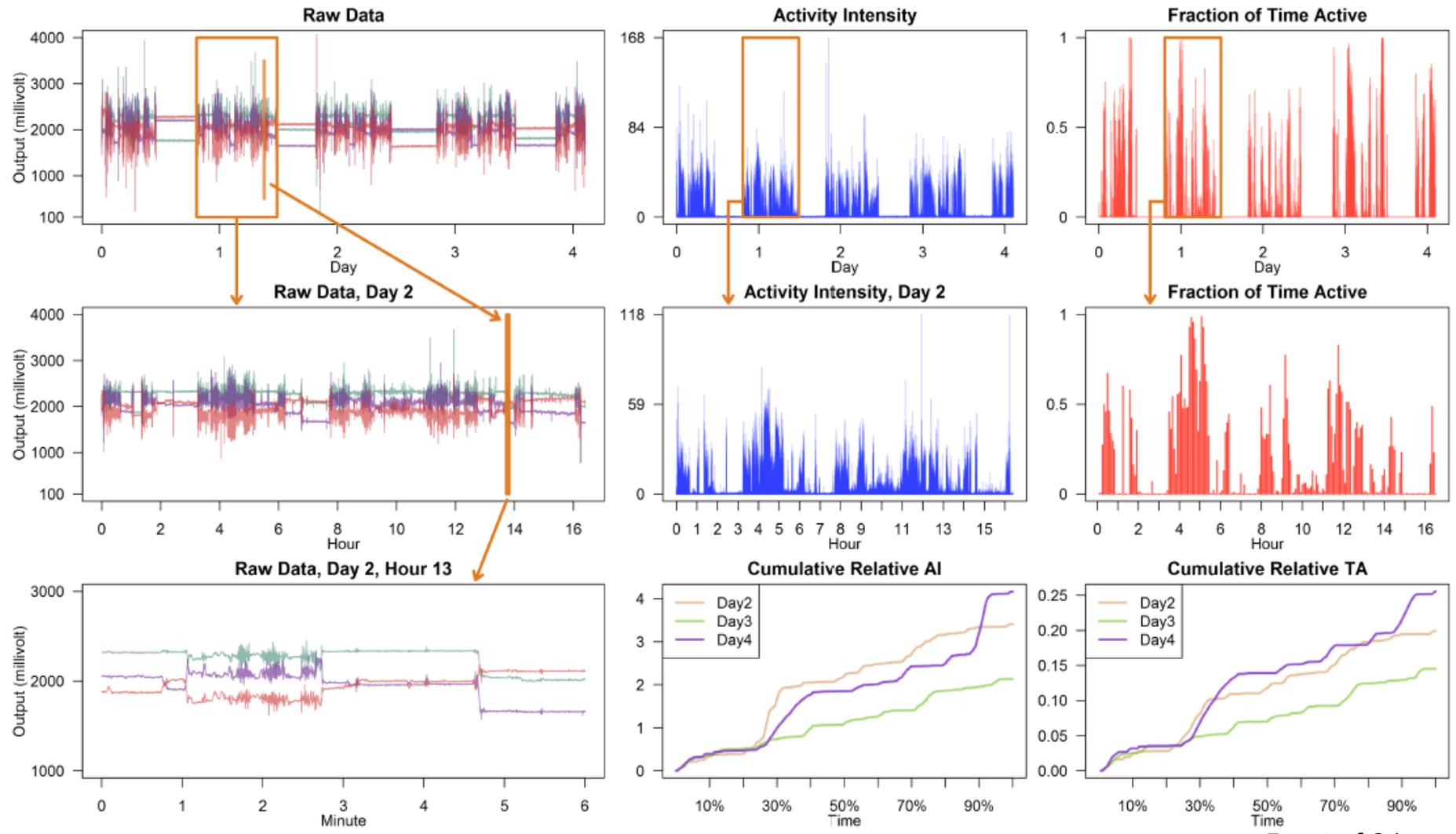
Daily patterns of activity counts
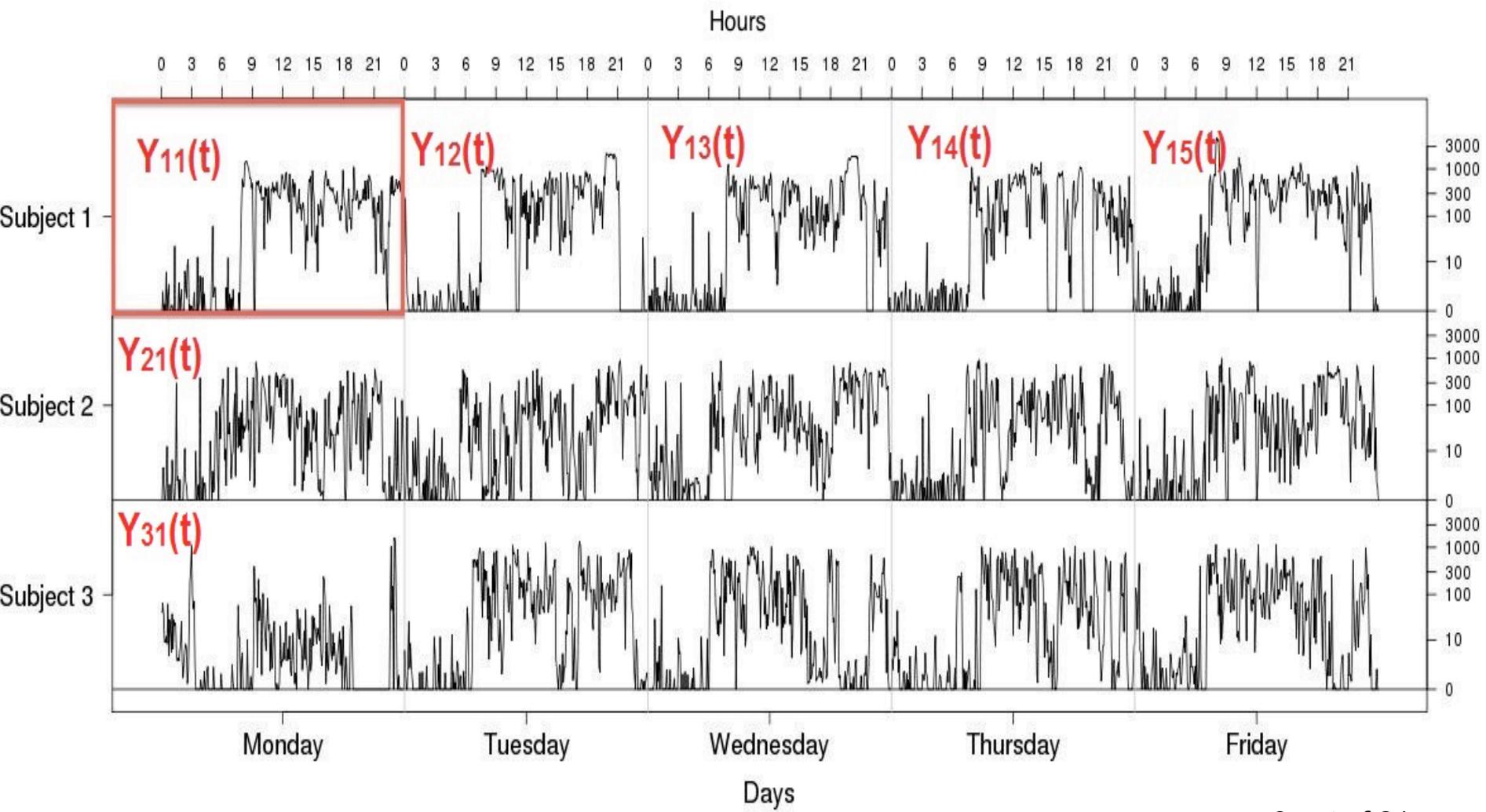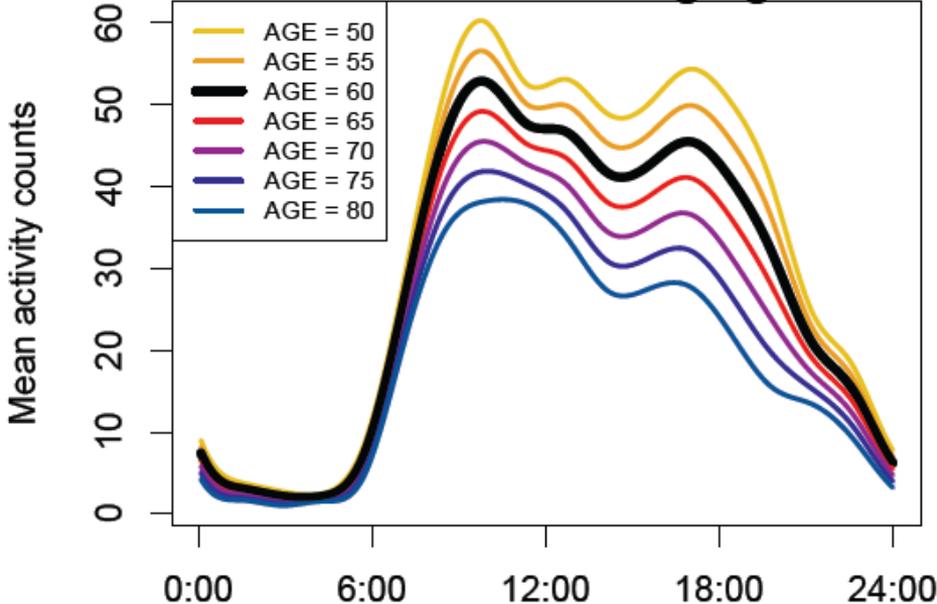
Baltimore Longitudinal Study of Aging (BLSA)

WIT: organized the BLSA data to the 1440+ standard

- Study participants : 773 (394 females, 379 males): i
- Average number of days/subjects : 7 : j
- Daily profile : 1440 minutes : t
- Age : between 31 and 96 : x
- Data set : 5478 by 1440

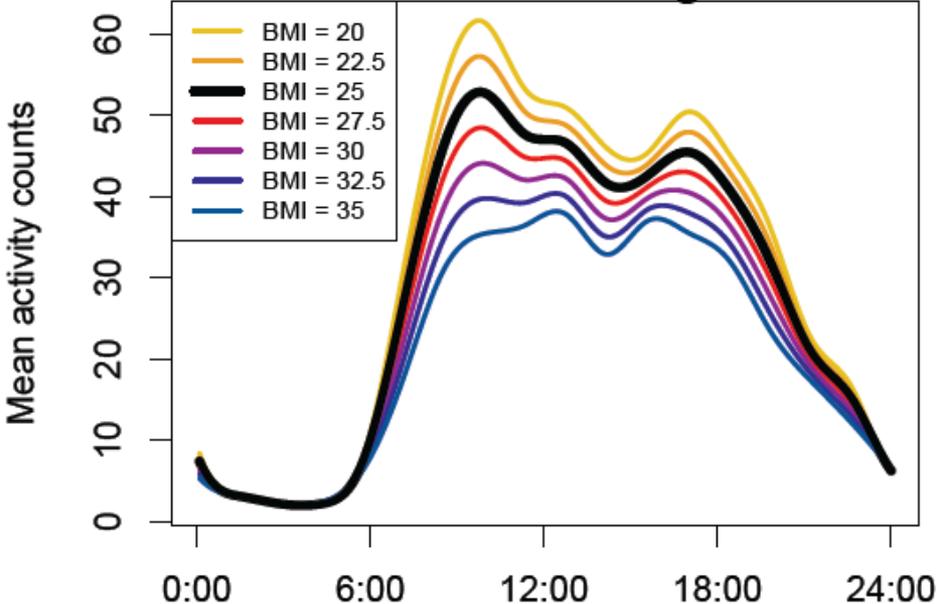$$Y_{ij}(t) = age_i\, \beta(t) + BMI_i\, \gamma(t) + W_{ij}(t)$$
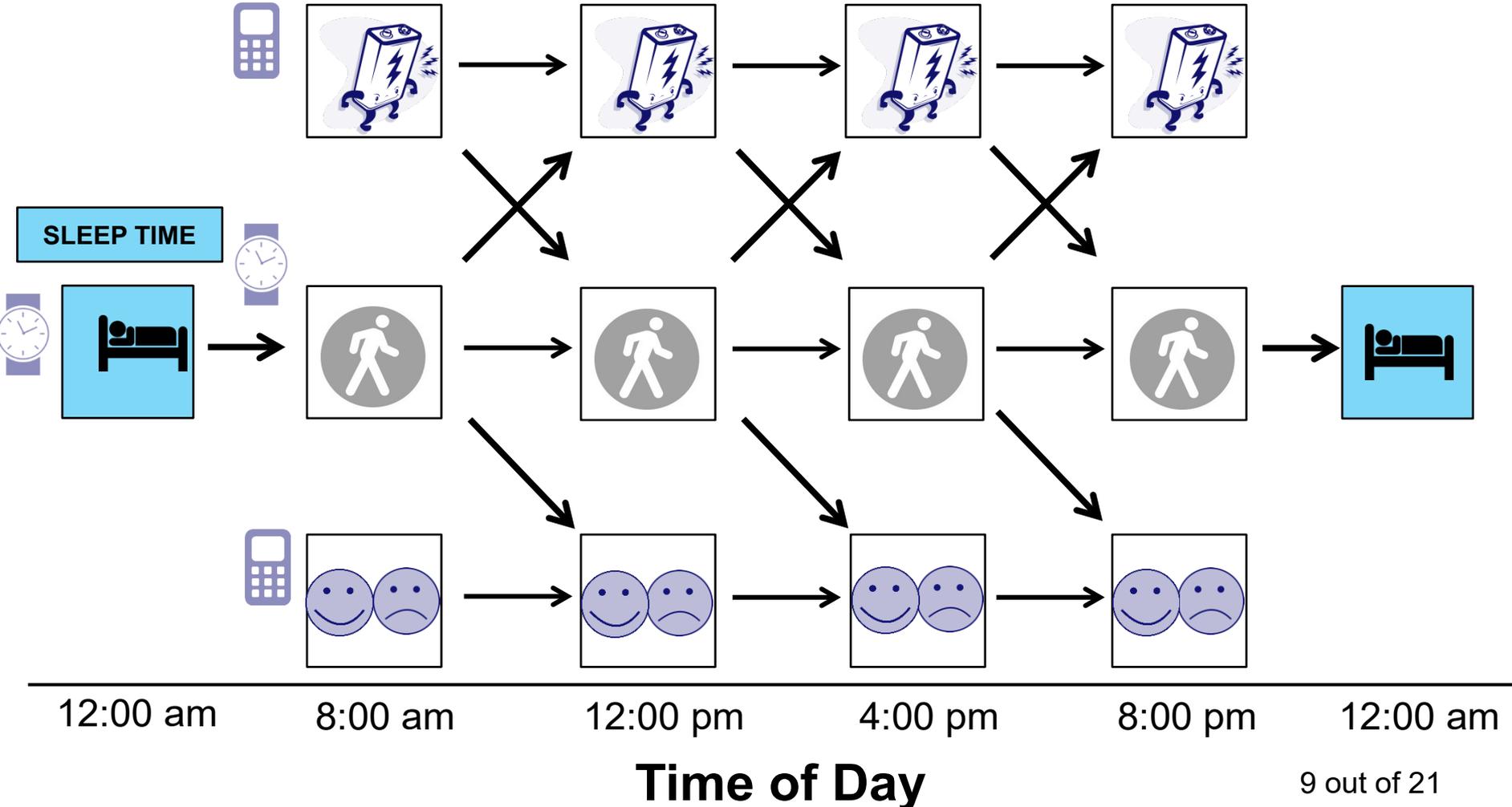
Structured function-on-scalar regression

Actigraphy & Environmental Momentary Assessment (EMA)

SLEEP TIME

12:00 am   8:00 am   12:00 pm   4:00 pm   8:00 pm   12:00 am

Time of Day

Granger causality

- Consider two time series: $X(t), Y(t)$

- We say that $X(t)$ Granger causes $Y(t)$ if the past of $X(t)$ predicts current values of $Y(t)$ even after accounting for the past of $Y(t)$

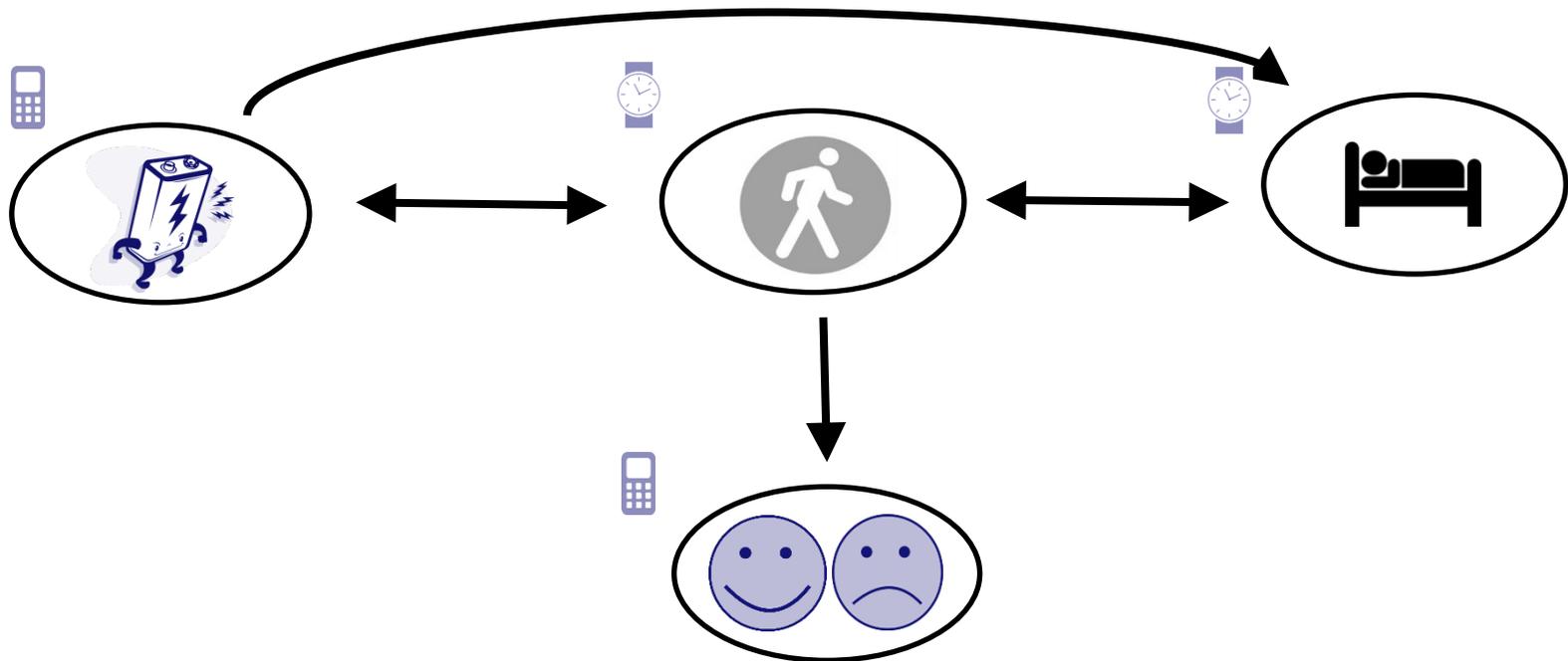- Examples of Granger causal model/test

$$Y(t) = \beta_0 + \theta_1 Y(t-1) + \delta_1 X(t-1) + \epsilon_t$$

$$Y(t) = \beta_0 + \sum_{l=1}^{L} \theta_l Y(t-L) + \sum_{k=1}^{K} \delta_k X(t-k) + \epsilon_t$$

- Models require local stationarity (θ_l and δ_k do not depend on t)

- Granger causality models
- Uni- and bi-directional associations
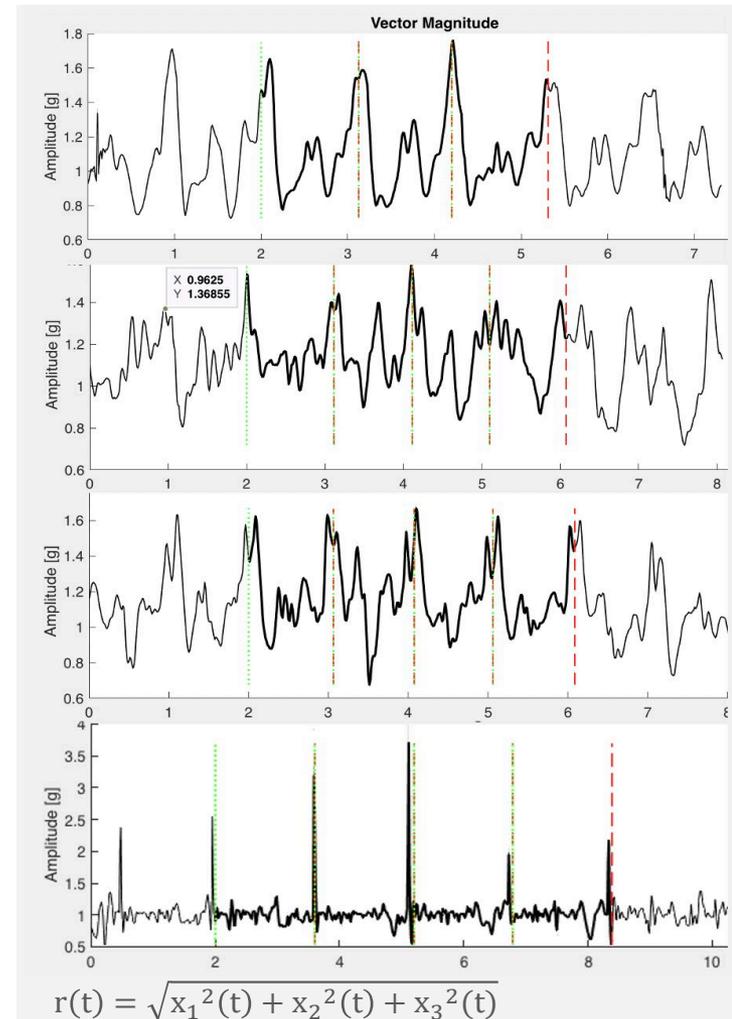- EMA: context, explore additional PA biomarkers

Scientific problem: walking strides segmentation from raw accelerometry data

Scientific problem

- Detailed walking characteristics have become increasingly important in health studies

  - Distance covered and speed in a 6-minute 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

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

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.
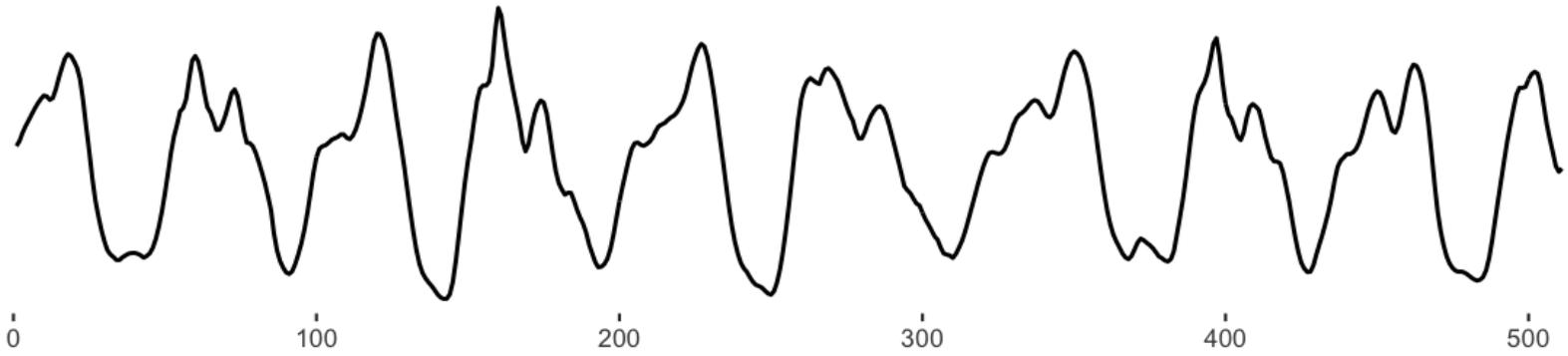
Adaptive Empirical Pattern Transformation (ADEPT)

- Use template, maximize local similarity (e.g., correlation)

- Multiple distinct templates can be used simultaneously

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

Template: ADEPT

Accelerometry
vector magnitude

ADEPT: applications

Applications

- **Karas et al. (2021):** higher free-living cadence associated with higher quality of life score
- **Urbanek et al. (2022):** higher free-living cadence associated with lower fall rates in older individuals (while clinic-based mobility measures were not)
- **Rubin et al. (2022):** Integrating ADEPT for smartphone-collected data for semi-supervised experiments; 6MWT cadence for preoperative risk stratification
- **Qiao et al. (2022):** novel waking cadence-based markers to characterize performance during a 400 m walk
- **Catallini (2020):** In MS thesis, ADEPT used for segmentation of neuronal activity traces from time series of calcium imaging

Resources

- "adept" R package (CRAN)
- "pyadept" Python package (actigraph/pyadept)
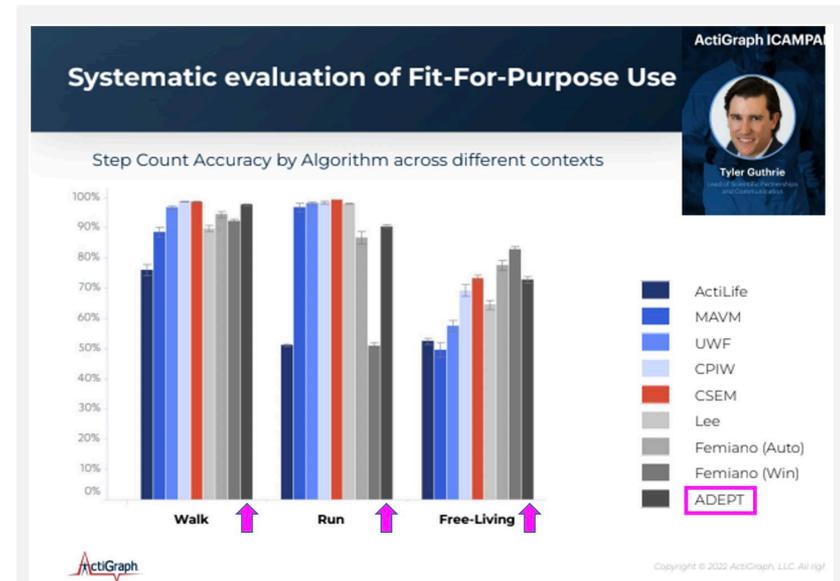
ICAMPAM 2022. Copyright @ 2022 ActiGraph, LLC. All rights reserved. Figure used thanks to courtesy of Christine Guo (ActiGraph, LCC).

Karas et al. (2021): ADEPT for free-living cadence estimation

Materials and methods

- Subsecond-level accelerometry data collected continuously for 4 weeks in 45 participants (30 arthritis and 15 healthy control)
- Estimated the daily free-living walking cadence for each participant
- Quantified cadence association with SF-36 quality of life (QoL) measures

Main results

- Free-living walking cadence was significantly associated with the Role physical score reported via SF-36 after adjusting for age, gender, weight and height

Summary of demographic information (rows 1-6) and SF-36 survey component scores (rows 7-16).

		All participants $n = 45$	Sub-group: arthritis patients $n = 30$	Sub-group: healthy volunteers $n = 15$
1	Gender: female	31 (68.9%)	21 (70.0%)	10 (66.7%)
2	Gender: male	14 (31.1%)	9 (30.0%)	5 (33.3%)
3	Age	51.9 (12.5)	53.9 (11.3)	47.9 (14.1)
4	Height [cm]	166.9 (7.7)	165.1 (7.6)	170.5 (6.7)
5	Weight [kg]	73.4 (13.8)	76.6 (13.4)	67.1 (12.7)
6	BMI	26.4 (5.0)	28.1 (4.9)	22.9 (3.0)
7	Physical function	64.8 (30.2)	49.2 (24.8)	96.0 (5.2)
8	Role physical	67.4 (29.4)	53.9 (26.1)	94.6 (10.5)
9	Bodily pain	61.5 (26.7)	48.3 (21.5)	87.8 (13.1)
10	General health	60.9 (24.8)	49.6 (21.6)	83.5 (11.8)
11	Vitality	53.0 (21.6)	44.7 (21.0)	69.6 (10.4)
12	Social functioning	78.1 (22.4)	69.8 (22.4)	94.6 (9.7)
13	Role emotional	82.0 (19.0)	76.2 (19.6)	93.6 (11.0)
14	General mental health	77.1 (13.3)	74.2 (14.7)	82.8 (7.4)
15	Physical Component Summary	44.0 (12.3)	37.5 (9.7)	57.1 (2.6)
16	Mental Component Summary	51.8 (7.1)	50.8 (7.9)	53.9 (4.7)
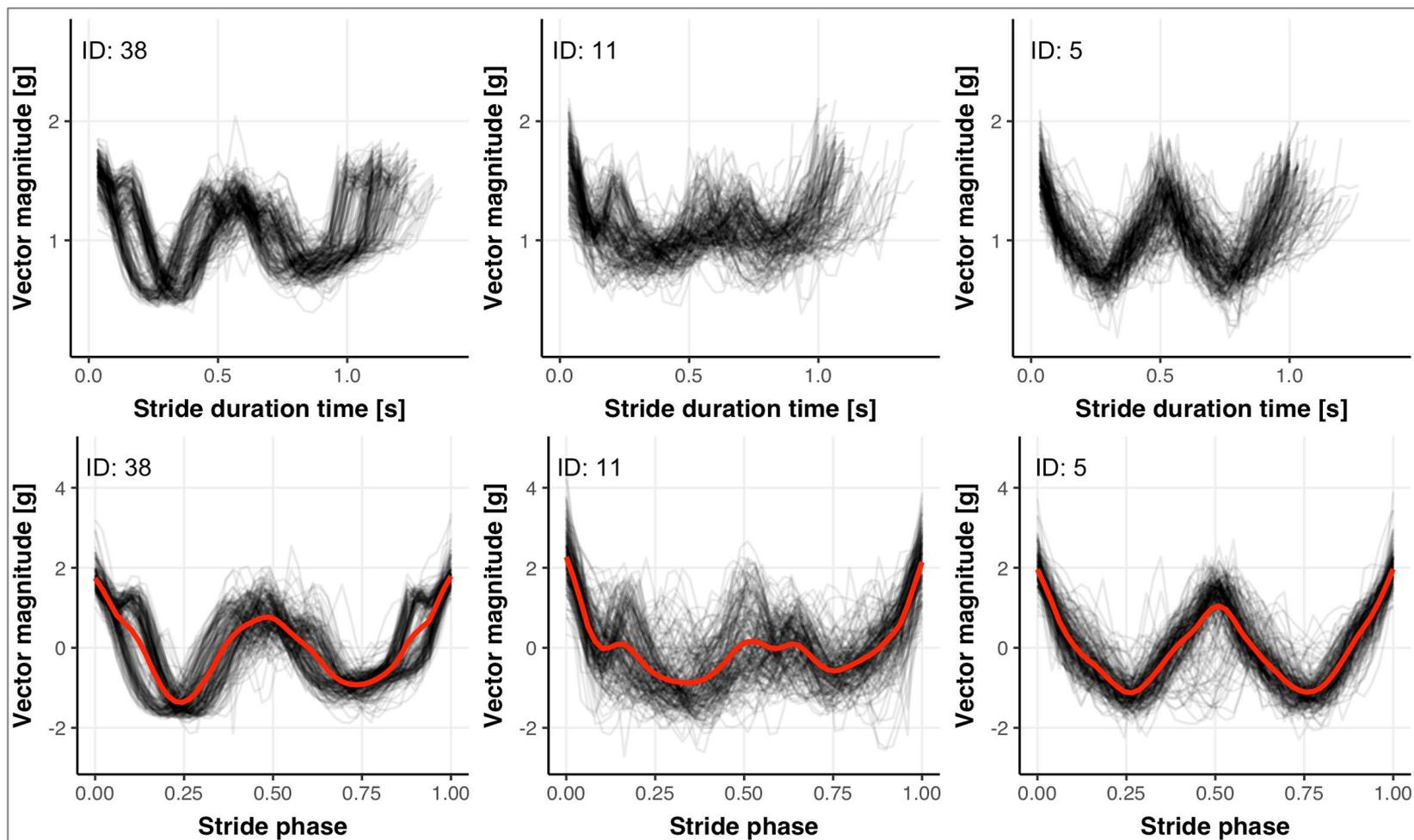
A sample of 200 estimated strides for three study participants. Plots in the first row display the vector magnitude (y-axis) as a function of clock time (x-axis). Plots in the second row display the same vector magnitude as a function of the time standardized to the [0, 1] interval. The red line is the point-wise mean of the standardized vector magnitude signals.
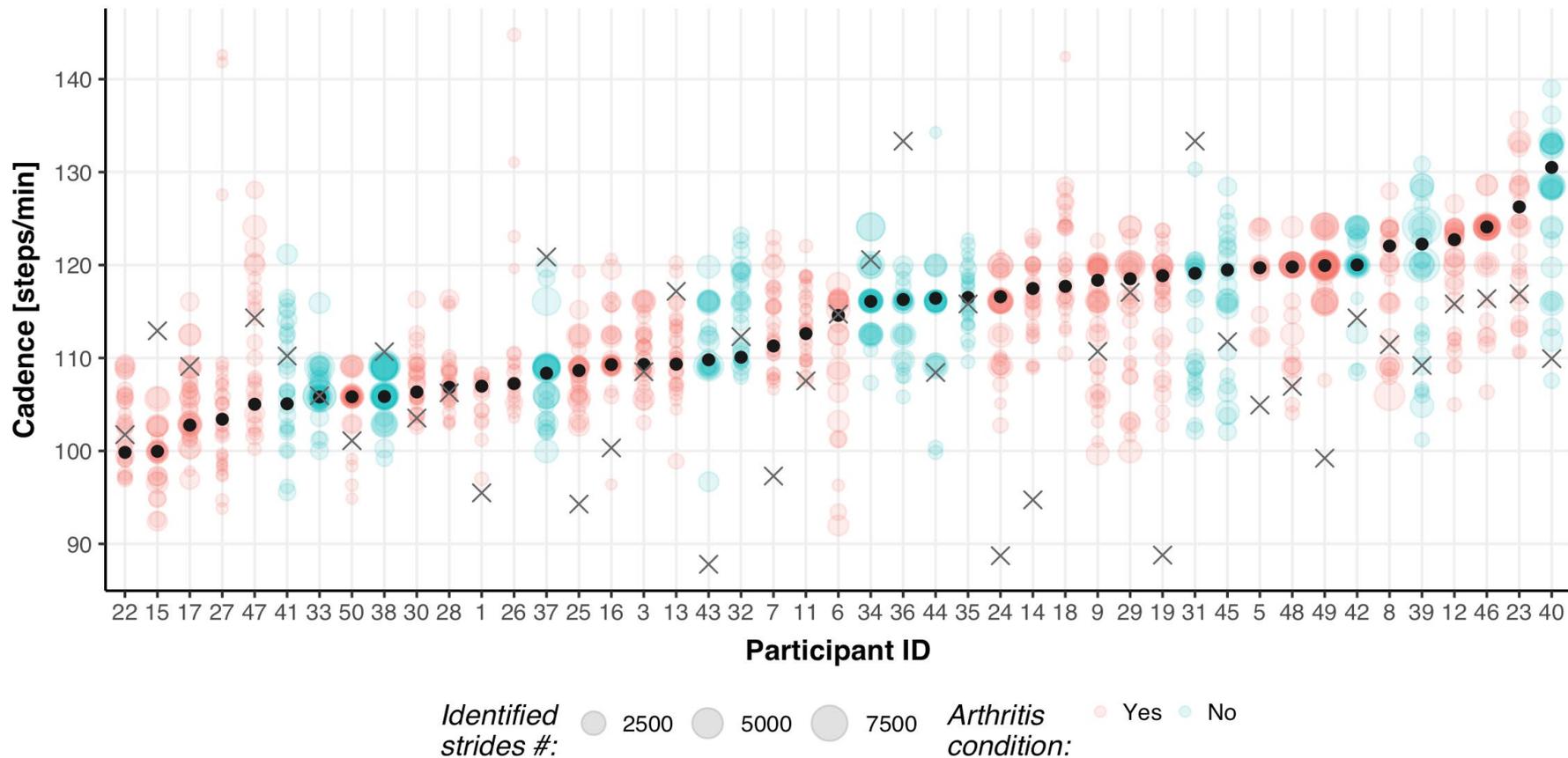
Daily cadence estimates (y-axis) **by study participant** (x-axis). Red denotes arthritis participants and blue denotes healthy volunteers. The size of a point: number of strides identified for a participant on a particular day. Black solid points: person-specific mode of daily cadence across all days. Black "x" signs: cadence estimated from supervised walking. Individuals are ordered according to their typical daily cadence.
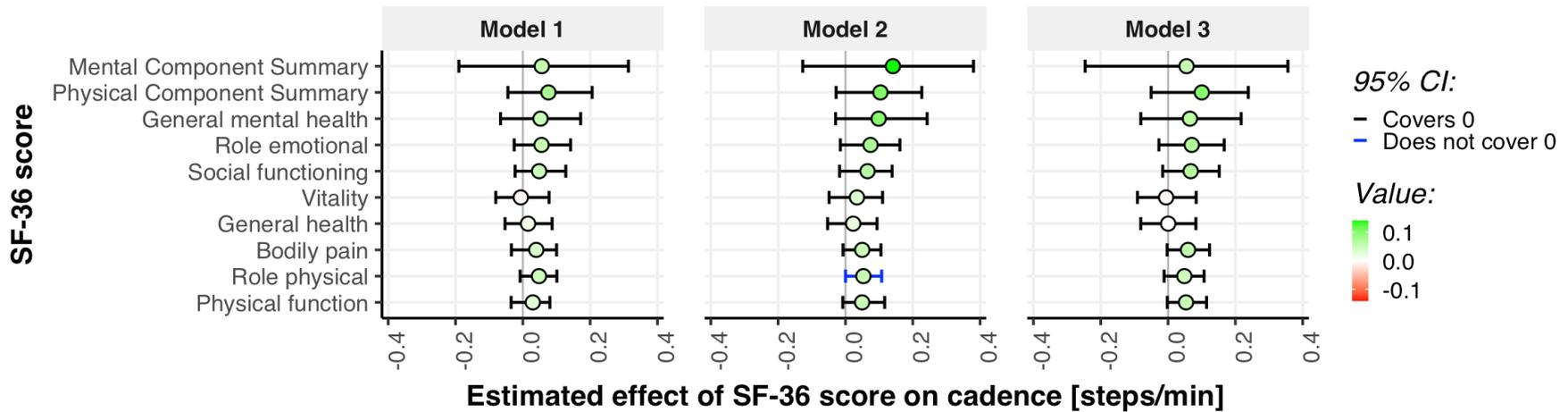
Linear mixed model estimates (x-axis) and corresponding 95% confidence intervals for the **association between SF-36 scores and free-living cadence (steps per minute).** The point color corresponds to the value of a model coefficient point estimate (green: positive, white: close to zero, red: negative). The confidence interval color is blue if the confidence interval does not cover 0 and is black otherwise.

Literature

- Heravi AS, Etzkorn LH, Urbanek JK, Crainiceanu CM, Punjabi NM, Ashikaga H, Brown TT, Budoff MJ, D'Souza G, Magnani JW, Palella FJ Jr, Berger RD, Wu KC, Post WS. HIV Infection Is Associated With Variability in Ventricular Repolarization: The Multicenter AIDS Cohort Study (MACS). Circulation. 2020 Jan 21;141(3):176-187.
- Leroux, A, Di, J, Smirnova, E, et al. Organizing and analyzing the activity data in NHANES. To appear in Statistics in Biosciences, 2019
- Karas, M., Bai, J., Strączkiewicz, M. *et al.* Accelerometry Data in Health Research: Challenges and Opportunities. *Stat Biosci* 11, 210–237
- Karas et al. (2019). Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. Biostatistics; 22(2): 331-347.
- 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, 42(6).
- Smirnova, E, Leroux, A, et al. The predictive performance of objective physical activity measures derived from accelerometry data for five-year all-cause mortality in NHANES, Unpublished manuscript
- Bai J, Goldsmith J, Caffo B, Glass TA, Crainiceanu CM. Movelets: A dictionary of movement. Electron J Stat. 2012;6:559-578
- Merikangas KR, Swendsen J, Hickie IB, et al. Real-time Mobile Monitoring of the Dynamic Associations Among Motor Activity, Energy, Mood, and Sleep in Adults With Bipolar Disorder. *JAMA Psychiatry*. 2019;76(2):190-198. doi:10.1001/jamapsychiatry.2018.3546
- He B, Bai J, Zipunnikov VV, et al. Predicting human movement with multiple accelerometers using movelets. *Med Sci Sports Exerc*. 2014;46(9):1859-66
- Goldsmith J, Zipunnikov V, Schrack J. Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics. 2015;71(2):344-53.
- Xiao L, Zipunnikov V, Ruppert D, Crainiceanu C. Fast Covariance Estimation for High-dimensional Functional Data. Stat Comput. 2014;26(1):409-421.
- Bai J, Sun Y, Schrack JA, Crainiceanu CM, Wang MC. A two-stage model for wearable device data. Biometrics. 2017;74(2):744-752.

SUPLEMENTARY SLIDES

Getting the organized NHANES accelerometry data

- NHANES data package (*rnhanesdata*):

    https://github.com/andrew-leroux/rnhanesdata

- Installing the rnhanesdata

    devtools::install_github(andrew-leroux/rnhanesdata)

Clinical trials with continuous wearable monitors

- Ideas stay the same: measurements are more complex
- Baseline balance (e.g., is activity the same in Tx/C?)
- Is treatment associated with outcome?
- Additional questions:
  - Timing of maximum mean Tx response
  - Time-varying effect of Tx
  - Time-varying quantile Tx effect
  - Subgroup time-varying analysis

High dimensional bi/tri-variate smoothing (BLSA)

$$Y_{ij}(t) = \mu(t, x_i) + U_i(t, x_i) + V_{ij}(t, x_i) + \varepsilon_{ij}(t)$$

- Subjects : 773 (394 females, 379 males): i
- Average number of days/subjects : 7 : j
- Daily profile : 1440 minutes : t
- Age : between 31 and 96 : x
- Data set : 5478 by 1440

- Requires:
  - fast new smoothers (Luo Xiao's penalty)
  - leave-one-subject-out CV (one-time data pass)

(a): Female activity surface (log counts)
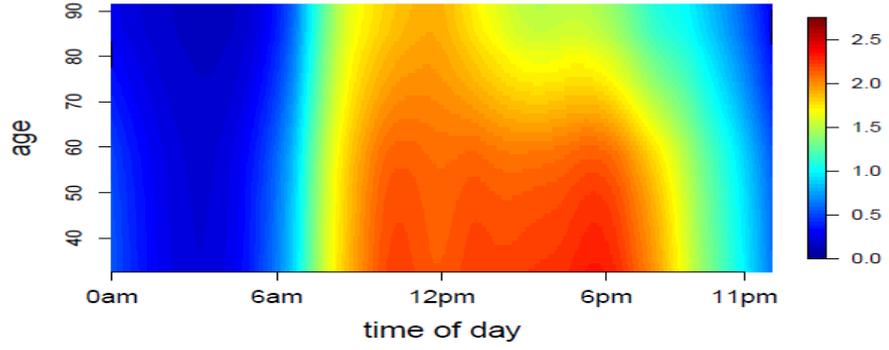
(b): Male activity surface (log counts)
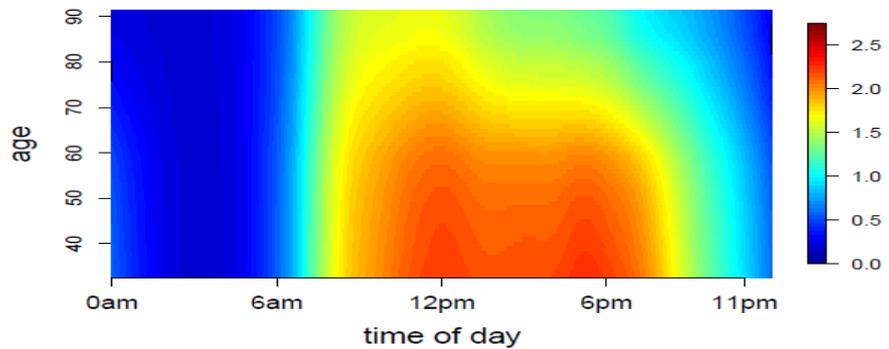
(c): Activity surface (log counts): male − female
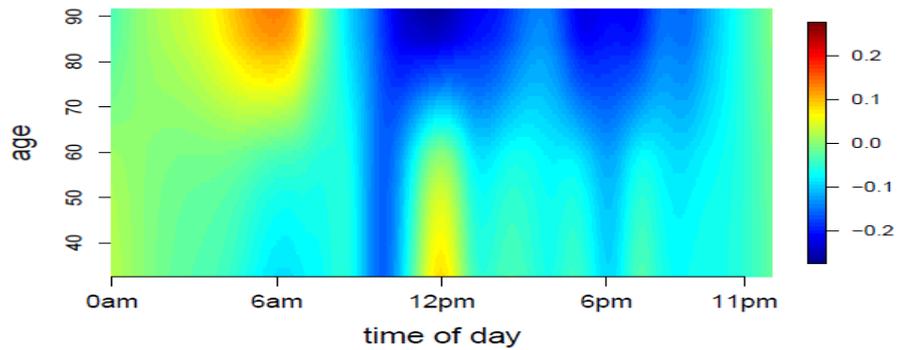
(d): Female
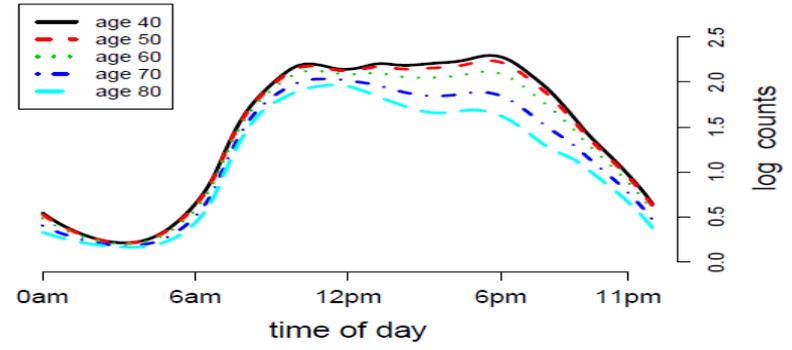
(e): Male
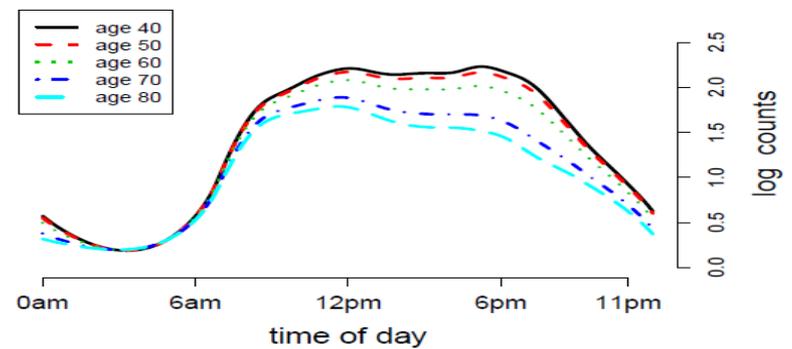
(f): Activity curves (log counts): male − female
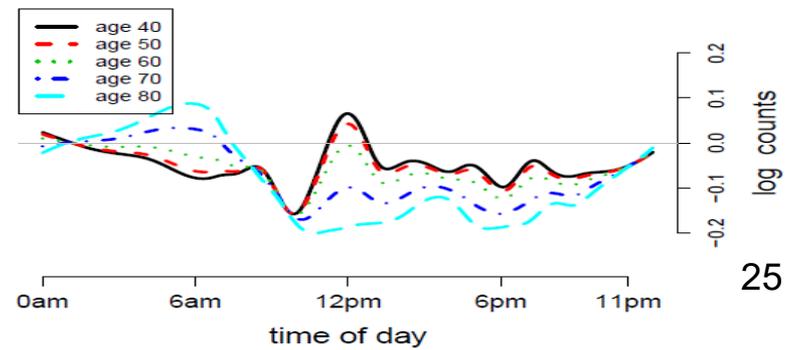
Do not trust your calorie counter

The Importance of Personalization in Counting Calories
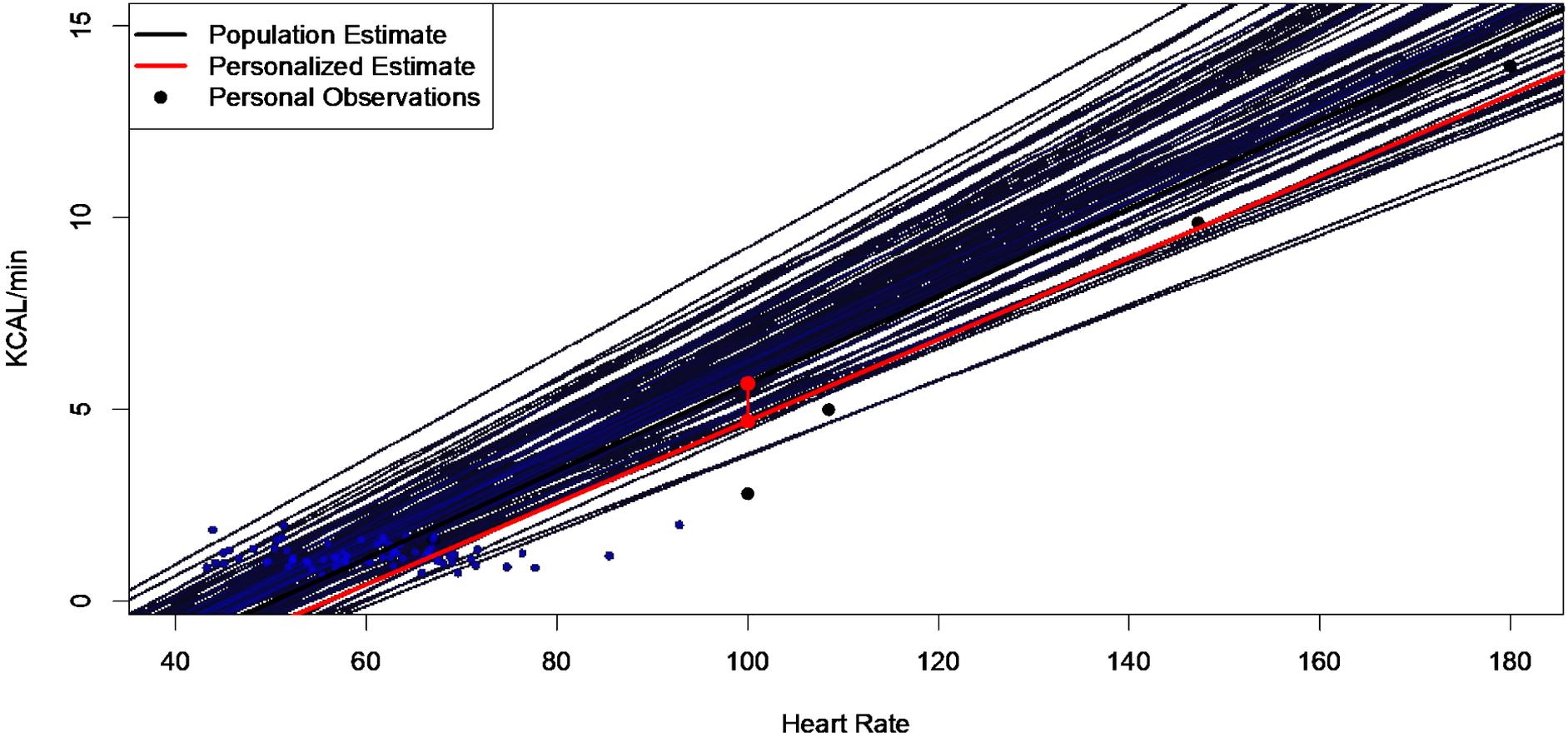
500 Calories	750 Calories	1000 Calories
43%	23%	9%

*Considerations
for Analyses Contributing to
Clinical Validation of
Biometric Monitoring
Technologies*

/////////

/// Bohdana Ratitch ///

ASA
New Jersey Chapter and Bayer
9th ANNUAL WORKSHOP

/// October 7th, 2022 ///

Acknowledgements

DiMe Research Project under the leadership of Yasaman Damestani

Digital Biomarkers

Viewpoint

Digit Biomark 2022;6:83–97
DOI: 10.1159/000525897

Considerations for Analyzing and Interpreting Data from Biometric Monitoring Technologies in Clinical Trials

Bohdana Ratitch[a]   Isaac R. Rodriguez-Chavez[b]   Abhishek Dabral[c]

Adriano Fontanari[d]   Julio Vega[e]   Francesco Onorati[f]

Benjamin Vandendriessche[g]   Stuart Morton[h]   Yasaman Damestani[i]

Outline

// Digital clinical measures – What, Why, How?

// V3 validation framework

// Elements of clinical validation -

      questions to be answered using data and analyses

// Summary

Digital Clinical Measures – What Are They?

Digital Clinical Measures
are clinical assessments or outcomes derived from **sensor-based** Digital Health Technologies capturing **physiological or behavioral data**

Digital Health Technology (DHT): A system that uses computing platforms, connectivity, software, and/or sensors for healthcare and related uses.

Biometric Monitoring Technologies (BioMeTs): Connected digital medicine products that process data captured by mobile **sensors**, using **algorithms** to generate measures of behavioral and/or physiological function.

Sensor: A transducer that converts a physical, biological, or chemical parameter into an electrical signal.

Digital Clinical Measures – How Can We Use Them?

Digital Clinical Measures

are clinical assessments or outcomes derived from **sensor-based** Digital Health Technologies capturing **physiological or behavioral data**

Early identification of pharmacological activity

Diagnostic biomarker or a stratification factor

Clinical measure for market differentiation

Quality of Life endpoint demonstrating value to patients

Regulatory-accepted endpoint for labeling

Digital Clinical Measures – How Do We Get There?

Digital Clinical Measures

are clinical assessments or outcomes derived from **sensor-based** Digital Health Technologies capturing **physiological or behavioral data**

Measure a Meaningful Aspect of Health

Validate rigorously

Ensure usability and accessibility

Achieve scientific & regulatory acceptance

Fit-for-purpose in clinical research, care, and decision-making

What Do we Measure?

Based on Fig 1 from DOI: 10.1159/000525897
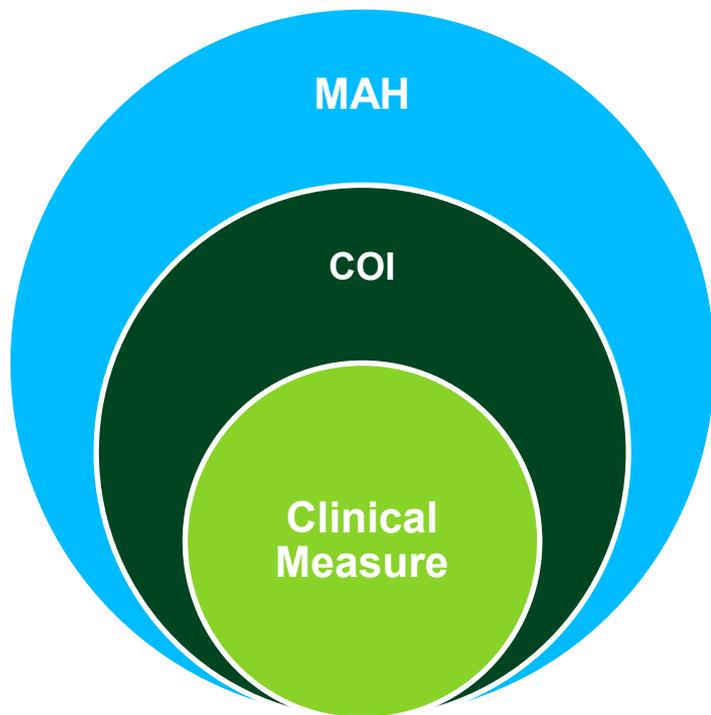
Meaningful Aspect of Health (MAH): Aspect of a disease that a patient does not want to become worse, or wants to improve, or wants to prevent.

Concept of Interest (COI): Simplified or narrowed element that can be practically measured.

Clinical Measure (Outcome): Specific assessment that accurately represents the COI.

Endpoint: Precisely defined, statistically analyzed health-related variable to demonstrate a clinical benefit of an experimental medical intervention.

What Do we Measure? – Example in Patients with Heart Failure

MAH

COI

Clinical Measure
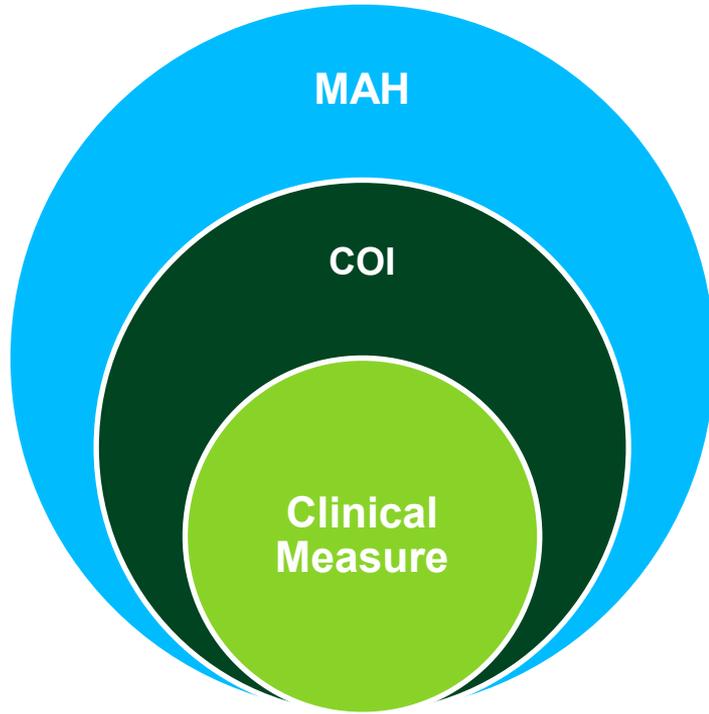
"I would like to be able to do everyday living activities without feeling exhausted"

Exercise capacity

Number of minutes of moderate or vigorous activity (MVPA) per day

Based on Fig 1 from DOI: 10.1159/000525897

How Do we Measure? - Example

MAH

COI

Clinical Measure

BioMeT

Based on Fig 1 from DOI: 10.1159/000525897

Algorithm

Raw signal — Three axes of accelerometer (x,y,z) sampled at 10Hz from a wrist-worn device

Feature, granular — Activity Count: a 1-minute aggregate from the pre-processed measurements of the three axes

Classification, granular — Classification of each minute as light, moderate, or vigorous activity based on Activity Count thresholds

Outcome — Number of minutes of moderate or vigorous activity (MVPA) per day

Endpoint — Change in average daily number of minutes of MVPA from baseline to Month 4

How Do We Validate?

Jennifer C. Goldsack[1], Andrea Coravos[1,2,3], Jessie P. Bakker[1,4], Brinnae Bent[5], Ariel V. Dowling[6], Cheryl Fitzer-Attas[7], Alan Godfrey[8], Job G. Godino[9], Ninad Gujar[10,11], Elena Izmailova[1,12], Christine Manta[1,2], Barry Peterson[13], Benjamin Vandendriessche[14,15], William A. Wood[16], Ke Will Wang[5] and Jessilyn Dunn[5,17]

Verification evaluates sample-level sensor outputs

Analytical validation evaluates the performance of an algorithm to convert sensor outputs into physiological metrics (outcome of interest) using a defined data capture protocol **in a specific subject population**

Clinical validation evaluates whether the clinical measure acceptably identifies, measures, or predicts a meaningful clinical, biological, physical, functional state, or experience, **in the stated context of use and specified population**

Design Specifications & Modular Prototyping

VERIFICATION

ANALYTICAL VALIDATION

CLINICAL VALIDATION

Clinical Utility

Source: DiMe - The Playbook

How Do We Validate? - Example

MAH

COI

Clinical Measure (Outcome)

BioMeT

Verification	Raw signal	Three axes of accelerometer (x,y,z) sampled at 10Hz from a wrist-worn device
Analytical Validation	Feature, granular	Activity Count: a 1-minute aggregate from the pre-processed measurements of the three axes
	Classification, granular	Classification of each minute as light, moderate, or vigorous activity based on Activity Count thresholds
Clinical Validation	Outcome	Number of minutes of moderate or vigorous activity (MVPA) per day
	Endpoint	Change in average daily number of minutes of MVPA from baseline to Month 4

Algorithm

Based on Fig 1 from DOI: 10.1159/000525897

Clinical Validation - Stages

Exploratory:
Feasibility and Fine Tuning

- Suggest a precise definition of a clinical measure
  - Qualitative input from patients, caregivers, and investigators
  - Quantitative, data-driven
- Select / fine-tune digital parameters
- Assess BioMeT maturity
- Evaluate feasibility

Confirmatory:
Demonstrating Measurement Properties

- Validate a precisely defined digital clinical measure
- Validate when measured using a specific type of BioMeT
- Validate in the target population

Exploratory

Confirmatory

Traceable

Transparent

Clinical Validation – Demonstrating Measurement Properties

Variability of repeated measurements under the same conditions

Associations between a novel clinical measure and other clinical measures or occurrence of clinical events

Reliability

Associations with Other Clinical Measures

Clinical Validation

Minimal Clinically Important Difference

Responsiveness

The smallest change in a clinical measure identified as important and indicating a change in clinical management

A.k.a. sensitivity to change

"Change" can be:
- within or across individuals
- concurrently or over time
- in relation to other measures

Based on Fig 2 from DOI: 10.1159/000525897

Clinical Validation – Demonstrating Measurement Properties

Variability of repeated measurements under the same conditions

Associations between a novel clinical measure and other clinical measures or occurrence of clinical events

Reliability

Associations with Other Clinical Measures

Clinical Validation

Minimal Clinically Important Difference

Responsiveness

The smallest change in a clinical measure identified as important and indicating a change in clinical management

A.k.a. sensitivity to change

"Change" can be:
- within or across individuals
- concurrently or over time
- in relation to other measures

Depend on context, including the type of clinical measure

Based on Fig 2 from DOI: 10.1159/000525897

Types of Clinical Measures?

Clinical Outcome Assessment (COA)

An outcome that describes or reflects how an individual feels, functions, or survives

- Clinician reported outcome (ClinPRO)
- Observer reported outcome (ObsRO)
- Patient reported outcome (PRO)
- Performance outcome (PerfO)

Biomarker

A defined characteristic that is measured as *an indicator* of normal biological process, pathogenic process, or biological response to an exposure or intervention

- Diagnostic
- Prognostic
- Susceptibility/ Risk
- Predictive
- Monitoring
- Safety
- Pharmacodynamic/ Response

electronic COA (eCOA) ⟵ **BioMeT** ⟶ **digital biomarker**

Key Analysis Objectives to Demonstrate Clinical Validity

Demonstrate that a digital measure can:	eCOA	Digital biomarker						
		PD	monitoring	safety	risk	prognostic	diagnostic	predictive
Reliably measure a COI within and across individuals in stable disease states within a range of environmental conditions	X	X	X		X	X	X	X
Differentiate between healthy individuals and those with disease	X	X	X	X			X	
Differentiate between concurrent disease/symptoms severity categories; correlate with other concurrent clinical outcomes	X	X	X	X			X	
Detect disease progression or a clinical event of interest	X	X	X	X			X	
Predict future outcomes (short or long term)			X	X	X	X		
Accurately detect functional states or activities of interest	X		X		X	X		X
Capture response to an intervention	X	X	X	X				
Predict response to an intervention								X

Table 1 from DOI: 10.1159/000525897

Reliably measure a COI within and across individuals in stable disease states within a range of environmental conditions

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X	X	X	X	X	X

// **Reliability**: Degree to which the results of repeated measurements under similar conditions can be replicated

// Prerequisite for other aspects of clinical validation

Reliably measure a COI within and across individuals in stable disease states within a range of environmental conditions

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X	X X	X	X

Reliability: Degree to which the results of repeated measurements under similar conditions can be replicated

Prerequisite for other aspects of clinical validation

Data and Study Design

// ≥2 measurements per individual during periods of stable disease

  // Clinical trials: repeated assessments during screening, treatment maintenance, follow-up

  // Separate studies without experimental treatment in individuals with stable disease

// Good representation of the target population (disease severity, phenotypes, etc.)

Analysis Methods

// Mixed effects models

// Intraclass correlation (ICC)

// Cohen's kappa

// Positive and negative agreement indices

// Estimation of measurement error

// Generalizability theory

// Latent variable modeling

Differentiate between healthy individuals and those with disease

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// E.g., differentiate between individuals with and without hypertension based on a measure derived from a continuous blood pressure monitoring device

Differentiate between healthy individuals and those with disease

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// E.g., differentiate between individuals with and without hypertension based on a measure derived from a continuous blood pressure monitoring device

Data and Study Design

// Measurements from individuals with and without a disease as determined by a reference diagnostic method

// Studies that can enroll participants without a disease but for whom diagnostic / monitoring procedure will be indicated in the future

Analysis Methods

// Classification methods

// Performance measures: sensitivity, specificity, precision, etc.

   // Attention to base rates of classified categories

// Validation of classification cut-off levels on independent data

Differentiate between concurrent disease/symptoms severity categories; correlate with other concurrent clinical outcomes

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// E.g., distinguish between New York Heart Association (NYHA) classes in patients with HF

// A weak association may sometimes be due to BioMeTs measuring different aspects of the MAH comparing to existing measures

Differentiate between concurrent disease/symptoms severity categories; correlate with other concurrent clinical outcomes

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// E.g., distinguish between New York Heart Association (NYHA) classes in patients with HF

// A weak association may sometimes be due to BioMeTs measuring different aspects of the MAH comparing to existing measures

Data and Study Design

// Concurrent assessments of digital measures and other related outcomes

// Good representation of the target population and disease states

// Data on potential confounders

Analysis Methods

// Regression and classification

// Distinct distributions of digital measurements across reference outcome/symptoms categories

// Significance of the digital measure effect on the reference outcome or vice versa

// Part of MCID determination (anchor analysis)

// Individual vs. group-level responsiveness

Detect disease progression or a clinical event of interest

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// Detect a concurrent and sometimes abrupt health status change

// E.g., detection of atrial fibrillation by Apple Watch based on heart rate monitoring

Detect disease progression or a clinical event of interest

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X		X	

// Detect a concurrent and sometimes abrupt health status change

// E.g., detection of atrial fibrillation by Apple Watch based on heart rate monitoring

Data and Study Design

// A study including participants who are monitored for an event of interest over a period of time

// Ideally, one group using a BioMeT and a control group not using it

// Good representation of the population to be monitored in the future

Analysis Methods

// Classification methods

// Careful evaluation of both false negative and false positive detection rates and their implications

Predict future outcomes (short- or long-term)

eCOA	Digital biomarker					
PD	monitoring	safety	risk	prognostic	diagnostic	predictive
	X		X	X	X	

// E.g., gait speed in patients with HIV as an early indicator of future decline in mobility

Predict future outcomes (short- or long-term)

eCOA	Digital biomarker						
PD	monitoring	safety	risk	prognostic	diagnostic	predictive	
	X		X	X	X		

// E.g., gait speed in patients with HIV as an early indicator of future decline in mobility

Data and Study Design

// A longitudinal clinical trial: digital clinical measure assessed at trial entry and possibly over time together with other clinical outcomes or events

// Can be embedded as secondary or exploratory objectives in interventional clinical trials

Analysis Methods

// Classification and regression models

// Evaluation of the significance of the biomarker effect in the model of the outcome of interest

// Predictive performance (discrimination and calibration)

// Adjustment for other relevant covariates

Detect functional states or activities of interest

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X		X		X	X		X

Examples:
// Sleep versus wake state
// Walking versus other types of locomotion
// Scratching behavior
// Hand movements associated with smoking or eating

Detect functional states or activities of interest

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X		X		X	X		X

Examples:
// Sleep versus wake state
// Walking versus other types of locomotion
// Scratching behavior
// Hand movements associated with smoking or eating

Data and Study Design

// Study with a reference method used to classify the functional states / activities of interest

// Conditions similar to those where BioMeT is intended to be used, e.g., lab versus daily living

Analysis Methods

// Feature engineering (exploratory stage)

// Classification methods

// Accuracy targets depending on context of use, including on individual and group level

Capture response to an intervention

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X			

// Sensitivity to intervention-induced changes – required for eCOAs or digital biomarkers planned to be used in clinical trials as endpoints

Capture response to an intervention

eCOA	Digital biomarker						
	PD	monitoring	safety	risk	prognostic	diagnostic	predictive
X	X	X		X			

// Sensitivity to intervention-induced changes – required for eCOAs or digital biomarkers (PD/response, monitoring, safety) planned to be used in clinical trials as endpoints

Data and Study Design

// Interventional studies with assessments of the digital and other clinical measures

// Well-established anchors, e.g., Patient Global Assessment of Severity (PGI-S) or Patient Global Assessment of Change (PGI-C)

Analysis Methods

// Assessment of the correlation between changes in the biomarker and changes in other clinical measurements from pre- to post-intervention

// MCID determination

Predict response to an intervention

eCOA	Digital biomarker						
PD	monitoring	safety	risk	prognostic	diagnostic	predictive
						X

// Predictive biomarkers - patient characteristics used to predict a future response to a treatment

// Combining multiple predictive biomarkers may capture the interplay between genomic, demographic, physiological, behavioral, and environmental factors

Predict response to an intervention

eCOA	Digital biomarker						
PD	monitoring	safety	risk	prognostic	diagnostic	predictive	
						X	

// Predictive biomarkers - patient characteristics used to predict a future response to a treatment

// Combining multiple predictive biomarkers may capture the interplay between genomic, demographic, physiological, behavioral, and environmental factors

Data and Study Design

// Interventional studies with candidate predictive biomarkers measured at baseline

// Ideally, separate studies for data-driven biomarker discovery and confirmation/estimation of treatment effect in biomarker-positive patients

Analysis Methods

// Principled data-driven identification of predictive biomarkers

// Addressing multiplicity and treatment effect estimation "optimism bias"

Biomarker as a Surrogate Endpoint

// Biomarkers as surrogate endpoints – based on mechanistic rationale and strong evidence that an effect on the surrogate endpoint predicts a specific clinical benefit

// Surrogate endpoints are used in cases where a clinical endpoint requires a very long follow-up or invasive assessment procedures

Biomarker as a Surrogate Endpoint

// Biomarkers as surrogate endpoints – based on mechanistic rationale and strong evidence that an effect on the surrogate endpoint predicts a specific clinical benefit

// Surrogate endpoints are used in cases where a clinical endpoint requires a very long follow-up or invasive assessment procedures

Data and Study Design

// Multiple clinical trials; may be a mix of randomized and observational studies, prospective and retrospective

// A combination of several biomarkers (composite biomarkers) may have stronger surrogate properties than any single surrogate parameter

Analysis Methods

// Analysis must show that:

// biomarker is prognostic with respect to the clinical outcome; and

// treatment effect on the surrogate endpoint reliably predicts treatment effect on the clinical outcome of interest

// Meta-analyses and model-based estimations of direct and indirect effects

Summary - What It Takes

Patient, Caregiver, Clinical Input

Robust Analytical Methodologies

Early Engagement with Regulators

Scientific Transparency

Pre-competitive Collaborations

Technologies with Demonstrated Verification and Analytical Validation

Data, Hardware, Software Standards

Reliability

Associations with Other Clinical Measures

Clinical Validation

Minimal Clinically Important Difference

Responsiveness

Fig 2 from DOI: 10.1159/000525897

Thank you!

Digital Health Technologies in Clinical Trials
Insights From Recent Applications

CV Damaraju, PhD

Janssen Research & Development, LLC

October 7, 2022

2022 ASA New Jersey Chapter/Bayer Statistics Workshop

Disclaimer

The opinions/views expressed in this presentation and on the following slides are solely of the speaker and do not represent those of the Janssen Research & Development, LLC (JRD, LLC).

Digital Health Technologies (DHTs) and Useful Terminologies

A DHT is "a system that uses computing platforms, connectivity, software, and/or sensors, for healthcare and other related uses."

Digital endpoint: A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question and derived from or includes a digital measurement.

Digital Biomarker (BM): A digital biomarker is an objective, quantifiable measure of physiology and/or behavior used as an indicator of biological, pathological process or response to an exposure or an intervention that is derived from a digital measure. The clinical meaning is established by a reliable relationship to an existing, validated endpoint.

Electronic Clinical Outcome Assessment (eCOA): An eCOA is a quantifiable measure used as a measure of how patients feel, function or survive that is derived from a digital measure

Randomized Clinical Trials with DHTs

Randomized Drug Trials

Drug   Placebo
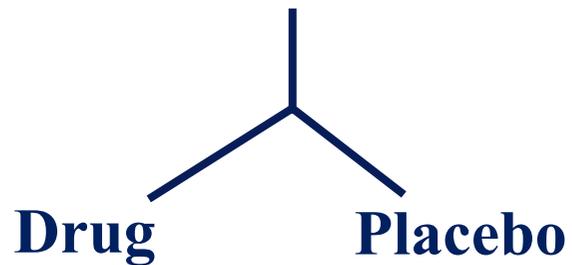
Endpoints

(clinical, DHT/ ePROs,
EMRs/Claims)

Efficacy

Safety

Randomized DHI* Trials

DHI*   No DHI*
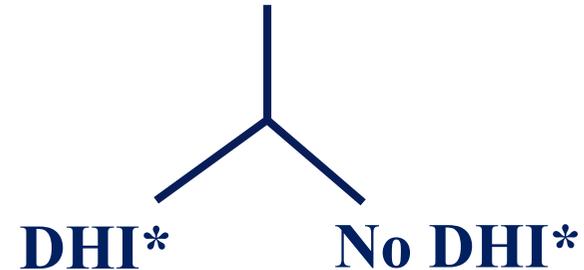
Endpoints

(clinical, DHT/ePROs,
EMRs/Claims)

Effectiveness

Safety

* The FDA defines Digital Health Interventions (DHIs) as the software or hardware used to improve the quality, access, efficacy, or efficiency of healthcare delivery (SaMD = software as medical device)

Common Objectives with DHTs

Exploratory

- Describe/examine (verify/validate) novel DHT algorithms
- Develop and investigate 'fit-for-purpose' DHTs
- Psychometric evaluation and interpretation of DHT endpoints

Causality

- Primary effect of intervention (Drug/DHIs/SaMDs)
- Adjunctive/secondary impact on study outcomes

Prediction

- Develop models (using DHT data) to predict clinical outcomes (eg, mortality, CV outcomes, readmissions/ER visits)
- Identify strong 'digital surrogates' for clinical outcomes

FDA Draft Guidance on use of DHTs

Digital Health Technologies for Remote Data Acquisition in Clinical Investigations

Guidance for Industry, Investigators, and Other Stakeholders

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Elizabeth Kunkoski, 301-796-6439; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Program Operations Staff at 301-796-5640.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Oncology Center of Excellence (OCE)

December 2021
Clinical/Medical

24217145dft.docx
21/20/2021

Section IV

A – Selection of DHT

C – Verification, Validation & Usability of DHTs

D – Clinical Endpoint evaluation using DHT data

E – Statistical Analysis

G – Record Protection & Retention

Considerations for Clinical Trials using DHTs

Statistical

Definition(s) and *estimand(s)* of DHT endpoint(s)

- Primary, secondary, exploratory

Data handling (raw/epoch/daily summary)

- Definition of a threshold/baseline (eg, "# valid days")
- Missingness mechanisms (MAR/MCAR/MNAR)
- Methods using MMRM, time series, ML

Study drug non-compliance and other intercurrent events

Meaningful change *versus* Anchor (correlation)

Heterogeneity in multiple assessments using different DHTs

Considerations for Clinical Trials using DHTs

Operational

Early regulatory input on the study design and selection of DHTs/endpoints

Robust data platform for multiple data sources
- DHTs, clinical, EMRs/Claims etc.

Site-less enrollment, eConsent, central randomization

Direct supply of study meds/DHTs, e-Visits

Recent RCTs with DHTs

mSToPs (n=2659): Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation
(see https://jamanetwork.com/journals/jama/fullarticle/2687353)

ACTIVATE (n=83): Wearable technology-based intervention for increasing moderate to vigorous physical activity in postmenopausal breast cancer survivors (see https://doi.org/10.1002/cncr.32143)

HEARTLINE (n=28000): Fully remote study to investigate if Early Atrial Fibrillation (AF) Diagnosis Reduces Risk of Events Like Stroke in the Real-World
(see https://clinicaltrials.gov/ct2/show/NCT04276441)

HEART WATCH (n=150): A randomized pragmatic trial among a population of patients with atrial fibrillation and/or atrial flutter undergoing cardioversion to determine the impact of the heart rate measurement, irregular rhythm notification, and ECG features of the Apple Watch on quality of life and healthcare utilization
(see https://bmjopen.bmj.com/content/bmjopen/11/12/e054550.full.pdf)

Canagliflozin: Impact on Symptoms, Physical Limitations and Quality of Life in Heart Failure (CHIEF-HF) Trial

Details of the CHIEF-HF study are available via multiple links to the American Heart Association (AHA) 2021 late breaking presentations as well as the primary publication (https://rdcu.be/cLprC)

Treatment Goals for Heart Failure

Principal Treatment Goals

To Make Patients
Live Longer

To Make Patients
Feel Better

Research Question:
Does 100 mg/d of canagliflozin improve the symptoms of patients with heart failure after 12 weeks of treatment?

Mortality

Quality of Life

Remote Conduct of the CHIEF-HF Study (N=476)

Planned ≥ 1 year *prior* to the pandemic (indication seeking study)

- Sponsor received early regulatory input on using a qualified COA as the primary
- Data on the primary collected via a mobile app and daily function through a Fitbit

Sites screened EMRs for potential patients (with HFrEF/HFpEF)

– Patients invited by email, portal, phone, or at visit

Patients went to website to learn of trial

– If interested, they opted in to be screened

If eligible, app downloaded on mobile & eConsent obtained by PI

Study Med

- 12-week

SAEs and c
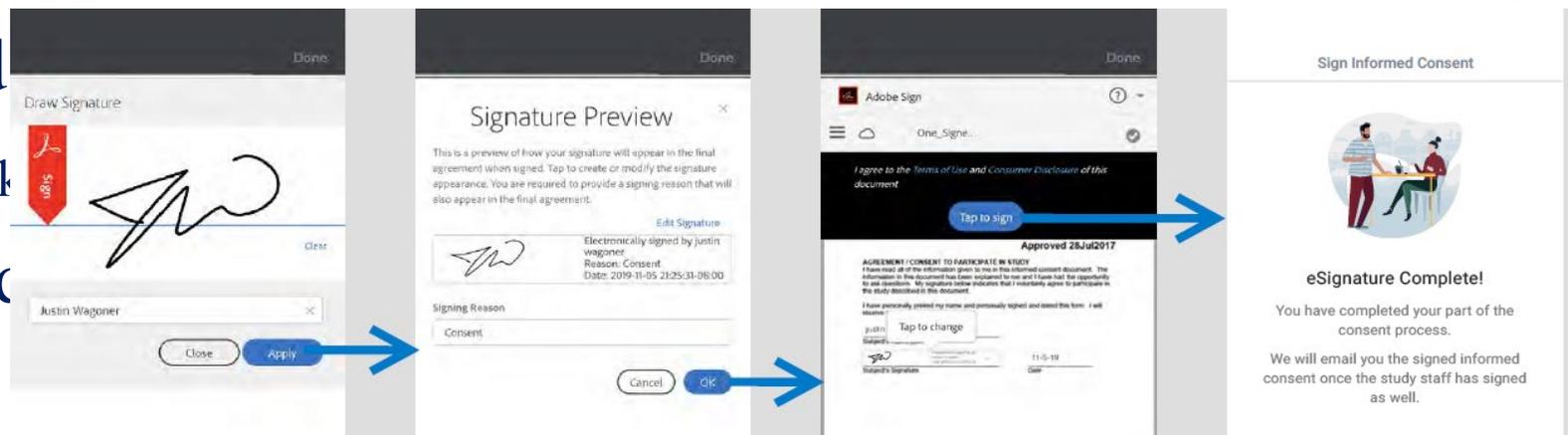
CHIEF-HF: A Fully Remote Randomized Trial

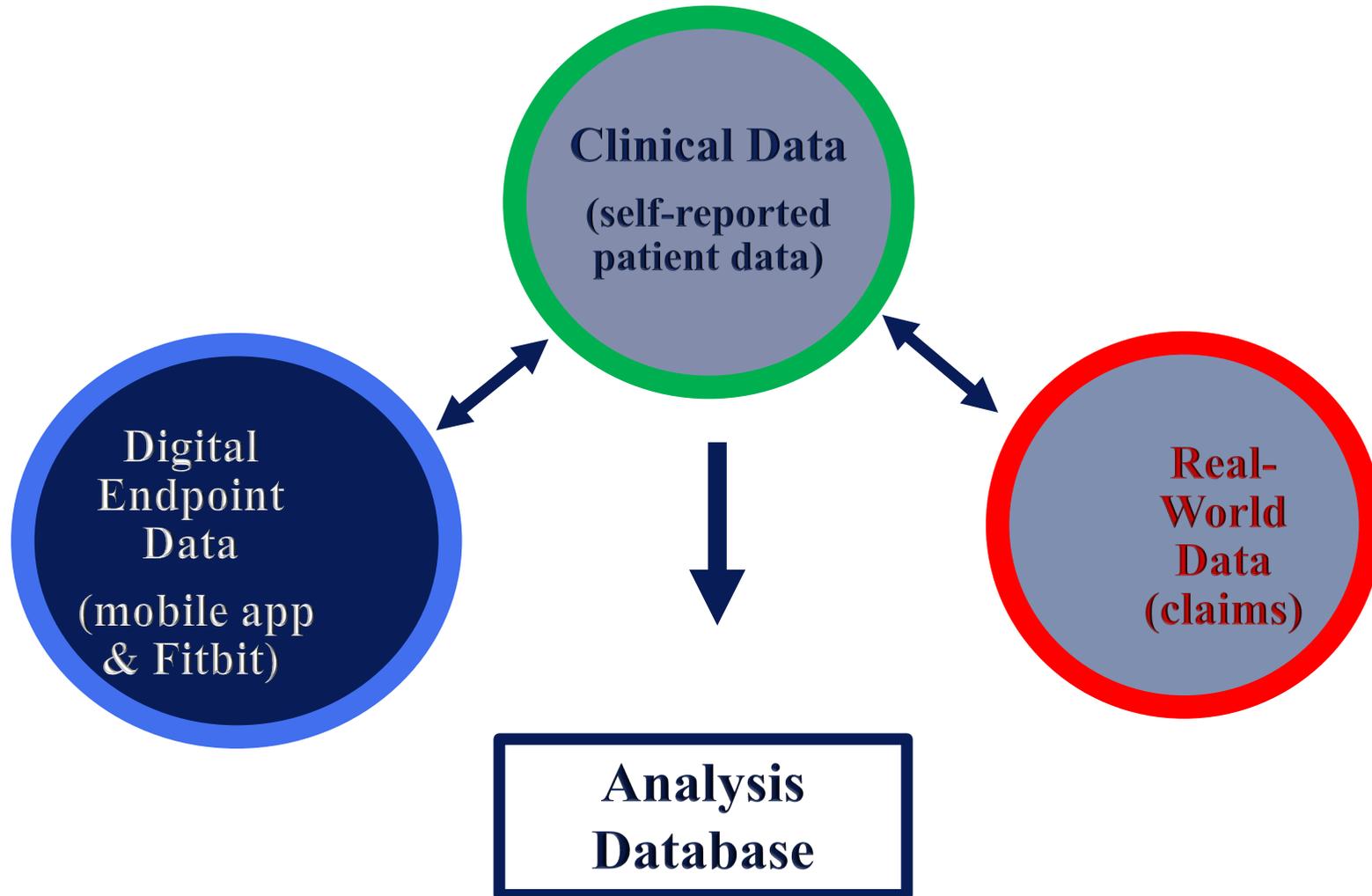Spertus, J.A., *et al.* Novel Trial Design: CHIEF-HF. *Circ Heart Fail* **14**, e007767 (2021)

The KC Cardiomyopathy Questionnaire

❖ 23/12 items that measure 5 clinically relevant domains

   » Physical Limitation

   » Symptoms

     Frequency   ⎤   Total

             ⎬ Symptom

     Severity   ⎦   Score

   Social Limitation

   Quality of Life

❖ Represents the *patient's* perspective of their HF

❖ Meaningful change well characterized

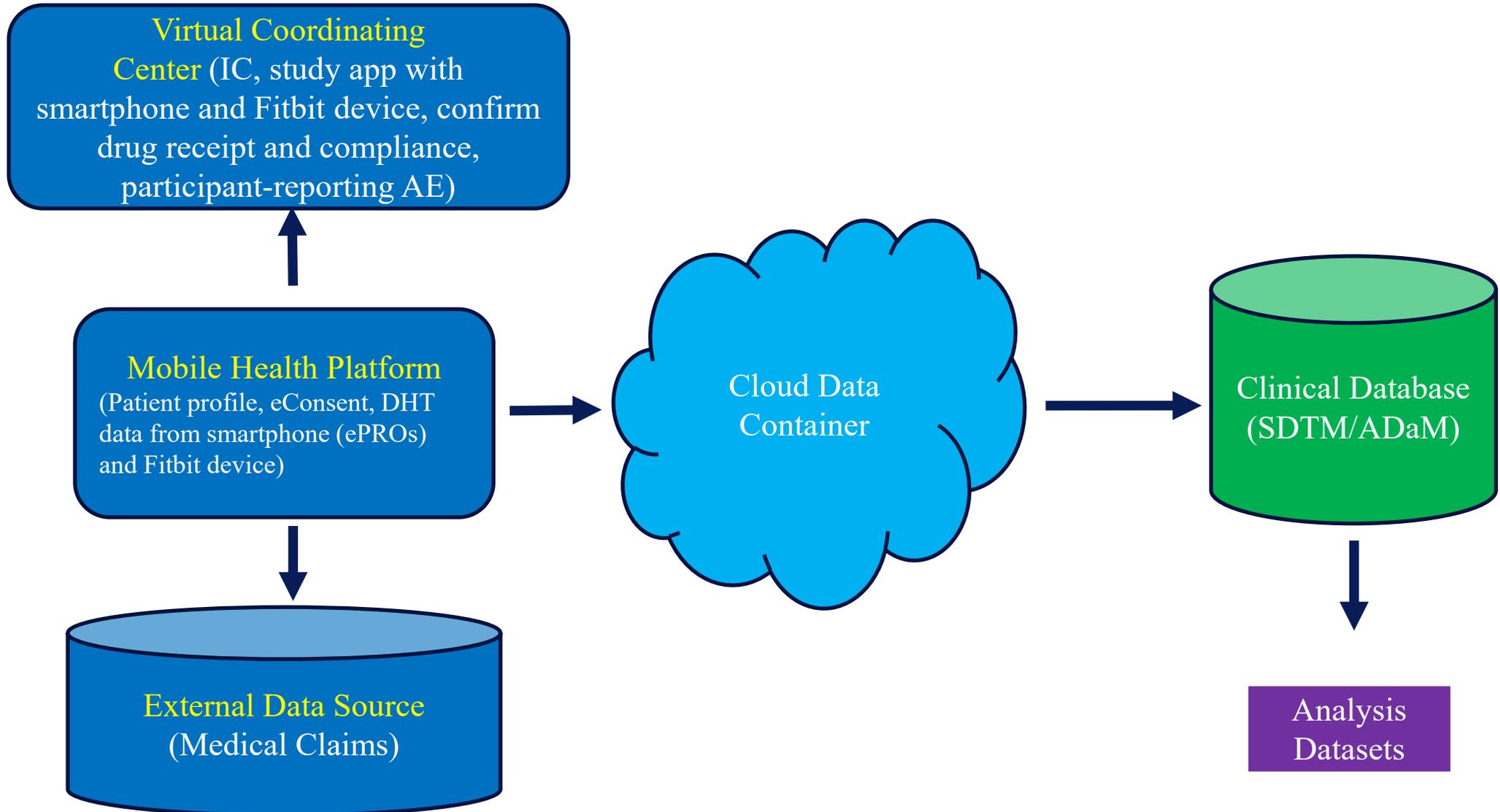❖ Qualified by FDA (CDRH & CDER) as a COA

❖ *Can be collected virtually!!*

Green et al, JACC 2000; 35:1245-55

CHIEF-HF Trial: Three Data Sources

Remote Data Flow/Digital Health Platform

Virtual Coordinating Center (IC, study app with smartphone and Fitbit device, confirm drug receipt and compliance, participant-reporting AE)

Mobile Health Platform (Patient profile, eConsent, DHT data from smartphone (ePROs) and Fitbit device)

External Data Source (Medical Claims)

Cloud Data Container

Clinical Database (SDTM/ADaM)

Analysis Datasets

Success of Study Execution

	Collected/Confirmed	Expected	Rate
Participant Eligibility			
	448	448	100%
Study Drug Delivery and Adherence			
Direct to patient drug delivered	448	448	100%
Self-reported medication compliance ≥80%	385	426	91%
Data Collection			
eDiary compliance ≥ 80%	426	448	95%
KCCQ at 2 weeks	444	448	99%
KCCQ at 4 weeks	431	438	98%
KCCQ at 6 weeks	418	429	97%
KCCQ at 12 weeks	414	422	98%
Fitbit data compliance ≥ 80%	422	448	94%

Kansas City Cardiomyopathy Questionnaire (KCCQ)

Results from the CHIEF-HF Trial

- Primary Endpoint (based on a validated *mobile* app)
  - 12-Week Change in KCCQ TSS from baseline
  - Tracked consistently well with PGIC (anchor variable)
  - Analysis for the primary endpoint used an MMRM approach

- Secondary Endpoints (based on a *Fitbit* device)
  - Change in 2-week <u>averages</u> for daily step counts & daily floors climbed
    - Daily step count/floors climbed set to 0 if less than 100 steps/day/ <1 floor/day
    - Values set to missing if Fitbit not used for >7 consecutive days
  - Bi-weekly averages analyzed using an MMRM approach
  - Missing data was minimal on these two endpoints
  - No significant change observed on both secondary endpoints

Primary Results – KCCQ Total Symptom Score
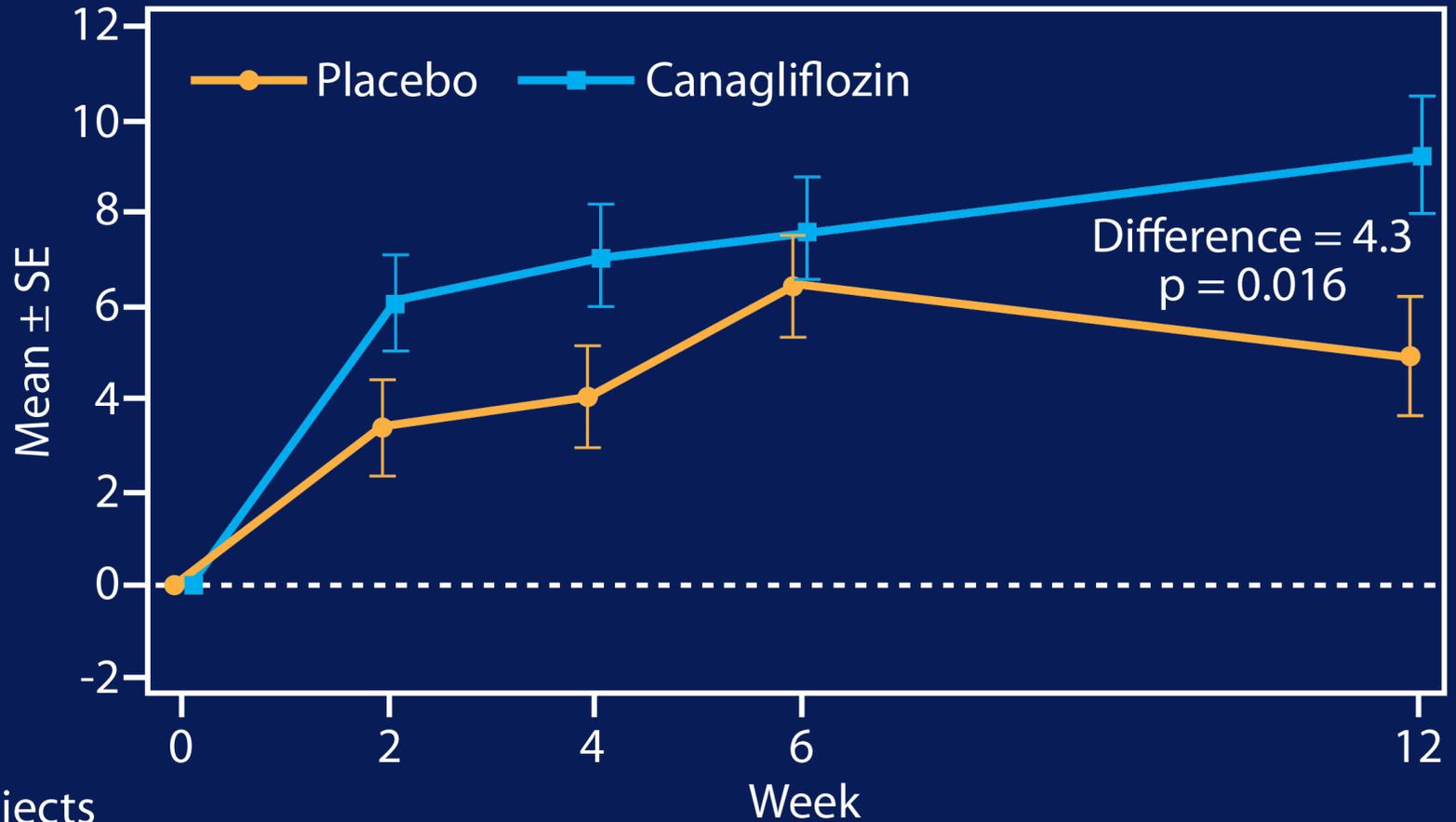
Number of Subjects

Placebo	226	222	215	209	206
Canagliflozin	221	221	215	208	207

Concluding Remarks

Fully remote clinical trials with DHTs support

    – Novel approaches to generate relevant evidence

    – Direct assessment of patient-centered outcomes

    – Early regulatory interactions on trial designs/endpoints

    (see https://www.fda.gov/media/155022)

DHTs with intense longitudinal data may require

    - Special analytical strategies, including for missing data

    - Definitions of the thresholds for "valid" intervals of data

    - Additional validation for association with clinical events

    (see Di et al. https://doi.org/10.1016/j.cct.2021.106661)

Thank you

7 October 2022

Causal Inference in Networks with Applications in Public Health,
ASA New Jersey Chapter, Bayer Statistics and Data Insights,
9th ANNUAL WORKSHOP.

Natallia V. Katenka, Ph.D., Associate Professor of Statistics, Department of Computer Science and Statistics, University of Rhode Island

Acknowledgements

Motivation and Background

- About 10% worldwide and 30% outside Africa of new HIV infections occurred because of injection drug use (Prejean et al., 2013; Lansky et al., 2014).

- Collecting data in network-based studies of hard to *reach populations* (e.g., People Who Inject Drugs (PWID)) is **non-trivial**.

- 2011 HIV outbreak among PWID in Athens demonstrated a large proportion of HIV sequences from newly diagnosed grouped into PWID-specific phylogenetic clusters → Effective interventions were urgently needed to prevent further transmission in Athens.

- *Transmission Reduction Intervention Project (TRIP)*: sociometric network-based study of injection drug users and their contacts in Athens, Greece from 2013 to 2015.

Motivation and Background

- PWID are embedded in social (HIV/HCV risk) networks and exert biological and social influence on the members of these networks (Hayes et al., 2000; Ghosh et al., 2017).

- In PWID networks, interventions often have **interference effects** (aka dissemination or spillover), which frequently depends on the network structure and intervention coverage levels.

- Evaluating causal effects in network-based studies is a **challenging** task.

Problem Statement

- Estimating causal effects in the presence of interference on a social network is challenged by complex *network structure*.

- Inteference effect could be stronger than individual effects and ignoring interference can lead to under-estimation the full impact of interventions (Buchanan et al., 2018).

- Interference sets could be deffined various ways (e.g., study clusters, communities, neighborhoods, nearest neighbors).

- Each individual's interference set may be overlapping with others due to network structure.

Motivating Study: Transmission Reduction Intervention Project (TRIP)

- Sociometric network-based study of injection drug users in Athens, Greece from 2013 to 2015 (following 2011 outbreak).
- Intervention: **community alerts** (inform of recent HIVs among close in network individuals).
  Outcome: **the HIV risk behavior at 6-month follow up**.

- Alerted
- Not alerted
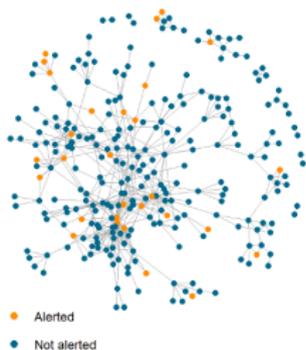
Baseline interview:

- 277 participants (25 alerted)
- 542 links

6-month follow up:

- 217 participants (25 alerted)
- 363 links

Exposure and Outcome in TRIP Study

Exposure

- **Community Alerts from Study Staff**

$$A_i = \begin{cases} 1, & \text{Received} \\ 0, & \text{Did not receive} \end{cases}$$

Outcome

- **Report of Injection HIV Risk Behavior at 6-month Visit**

$$Y_i = \begin{cases} 1, & \text{At least some} \\ 0, & \text{None reported} \end{cases}$$

Methodology

- Recent methodological developments have relaxed the no interference assumption and allowed for interference within clusters, known as **partial interference**

Nearest neighborhood IPW estimators

When partial interference assumption does not hold, we proposed twp inverse probability weighted (IPW) estimators where the interference set is the individual's *nearest neighbors* within the network.

1. IPW1 is a novel application for a sociometric network-based study setting (Liu et al., 2016).

2. IPW2 uses a generalized propensity score (Forastiere et al., 2016) with a weighted estimator instead of a stratified estimator.

Contribution

- Recent methodological developments have relaxed the no interference assumption and allowed for interference within clusters, known as **partial interference**

Closed-Form Variance

Using M-estimation, we derived closed-form variance estimators that allow for overlapping interference sets, often leading to a more statistically efficient estimator in network-based studies due to the use of additional information on connections between individuals.

Nearest Neighborhoods
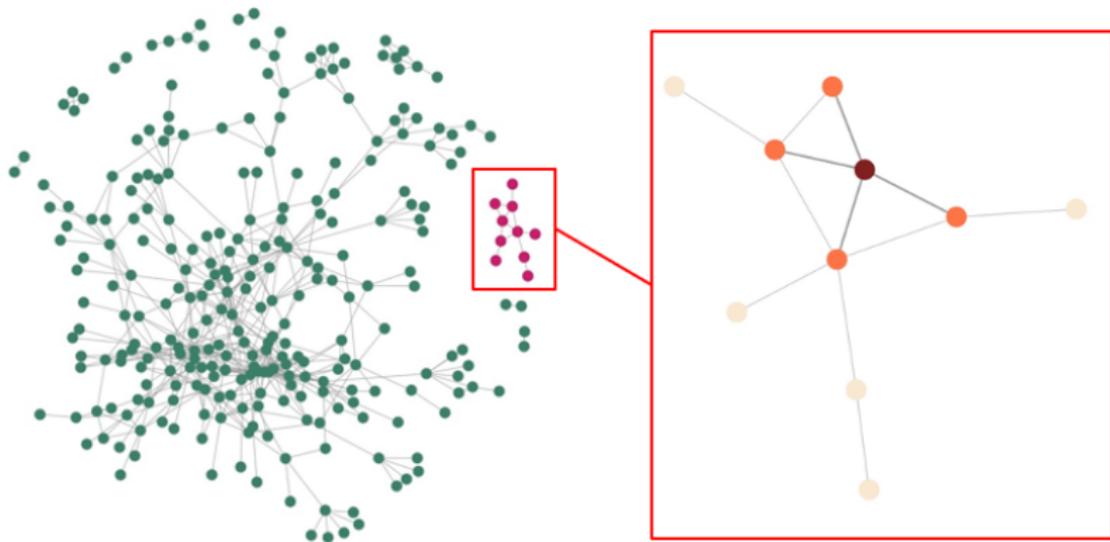
Notations

Let $i = 1, 2, \cdots, n$ denote each participant in the study.

y_i: denote the potential outcome, $y_i(a_i; a_{\mathcal{N}_i})$

Y_i: be the observed outcome, $Y_i(A_i; A_{\mathcal{N}_i})$

A_i: the self-selected binary treatment/exposure of participant i

Z_i: the vector of covariates for participant i

$\mathcal{N}_i$: the set of participants that share a link with i

d_i: $|\mathcal{N}_i|$, the degree of node i

$A_{\mathcal{N}_i}$: the vector of baseline exposures for participants in $\mathcal{N}_i$

$Z_{\mathcal{N}_i}$: the vector of baseline covariates for participants in $\mathcal{N}_i$

Estimands (1)

Under allocation strategy α, the probability of neighborhood of i is denoted by $\pi(a_{\mathcal{N}_i}; \alpha) = \alpha^{\sum a_{\mathcal{N}_i}} (1 - \alpha)^{d_i - \sum a_{\mathcal{N}_i}}$ and the probability of individual of i is $\pi(a_i; \alpha) = \alpha^{a_i} (1 - \alpha)^{1 - a_i}$.

The population average potential outcome is defined by

$$\bar{y}(a, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \sum_{a_{\mathcal{N}_i}} y_i(a_i = a, a_{\mathcal{N}_i}) \pi(a_{\mathcal{N}_i}; \alpha)$$

and the marginal population average potential outcome is

$$\bar{y}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \sum_{a_i, a_{\mathcal{N}_i}} y_i(a_i, a_{\mathcal{N}_i}) \pi(a_i, a_{\mathcal{N}_i}; \alpha)$$

Estimands (2)

Under allocation strategy α, the direct effect is

$$\overline{DE}(\alpha) = \bar{y}(1, \alpha) - \bar{y}(0, \alpha).$$

The spillover or indirect effect under allocation strategy $\alpha = (\alpha_0, \alpha_1)$ is

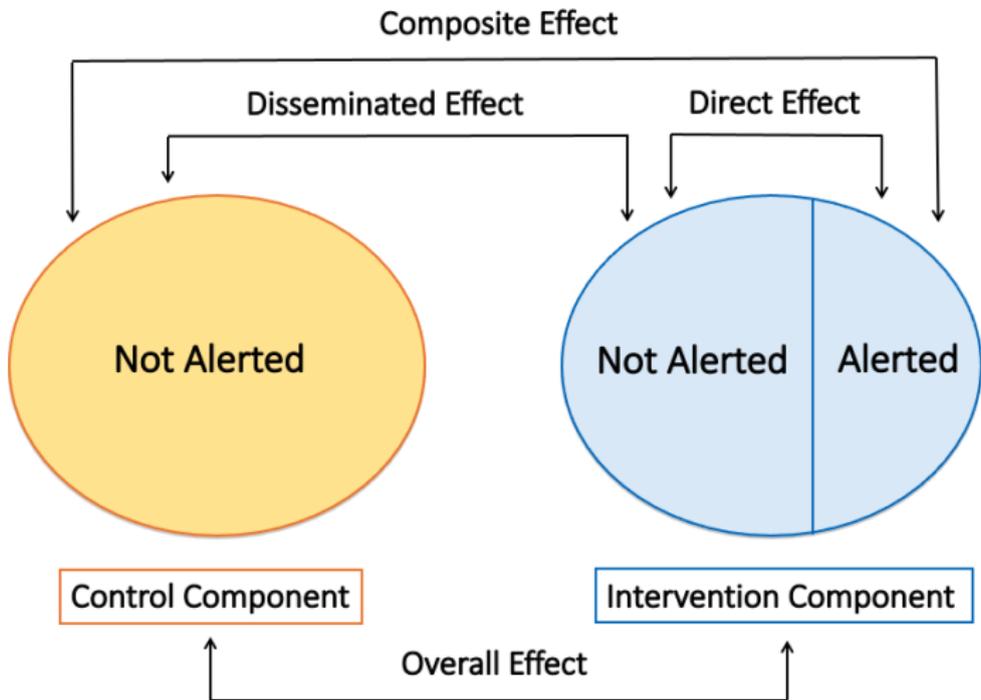$$\overline{IE}(\alpha) = \bar{y}(0, \alpha_1) - \bar{y}(0, \alpha_0).$$

The composite or total effect is

$$\overline{TE}(\alpha) = \bar{y}(1, \alpha_1) - \bar{y}(0, \alpha_0).$$

The overall effect is

$$\overline{OE}(\alpha) = \bar{y}(\alpha_1) - \bar{y}(\alpha_0).$$

Estimands

Assumptions (1)

- The potential outcomes only depends on the individual and their nearest neighbors $y_i | a_i, a_{\mathcal{N}_i}$.
- Conditional exchangeability for participants:

$$\Pr(A_i = a_i | Z_i = z_i) = \Pr(A_i = a_i | z_i, z_{\mathcal{N}_i}, y_1(\cdot), \ldots, y_n(\cdot))$$

- Conditional exchangeability for neighbors:

$$\Pr(A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}) = \Pr(A_i = a_i, A_{\mathcal{N}_i} = a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}, y_1(\cdot), \ldots, y_n(\cdot))$$

Assumptions (2)

- We assume the treatment positivity

$$Pr(a_i, a_{\mathcal{N}_i} | z_i, z_{\mathcal{N}_i}) > 0 \text{ for all } a_i, a_{\mathcal{N}_i}, z_i, \text{ and } z_{\mathcal{N}_i}.$$

- $C_i \perp\!\!\!\perp A_i | L_i$. i.e. C_i only depends on the baseline covariates.
- Stratified interference assumption
- Smaller groupings or neighborhoods for each individual can be identified in the observed network.

Nearest Neighborhood IPW1 Estimator

$$\widehat{Y}^{IPW_1}(a, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i}) I(A_i = a) \pi(A_{\mathcal{N}_i}; \alpha)}{f(A_i, A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})}.$$

The marginal IPW estimator is defined as

$$\widehat{Y}^{IPW_1}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i(A_i, A_{\mathcal{N}_i}) \pi(A_i, A_{\mathcal{N}_i}; \alpha)}{f(A_i, A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i})}.$$

The propensity score is defined as

$$f(A_i, A_{\mathcal{N}_i} | Z_i, Z_{\mathcal{N}_i}) = \int \prod_{j \in \mathcal{N}_i^*} p_j^{A_j} (1 - p_j)^{1-A_j} f(b_i, 0, \theta_s) db_i$$

where $\mathcal{N}_i^* = \mathcal{N}_i \cup \{i\}$ and $p_j = \text{logit}^{-1}(Z_i \cdot \theta_z + b_i)$.

Nearest Neighborhood IPW2 Estimator

- IPW2 estimator uses an individual and nearest neighbors propensity score (Forastiere et al., 2016).
- Potential outcomes of individual i depend on the total number of exposed neighbors, $s_i = \sum_{j \in \mathcal{N}_i} a_j$ ($S_i = \sum_{j \in \mathcal{N}_i} A_j$):

$$y(a_i, a_{\mathcal{N}_i}) = y(a_i, s_i)$$

Nearest Neighborhood IPW2 Estimator

The IPW2 estimator for treatment a with coverage α is defined as

$$\widehat{Y}^{IPW_2}(a; \alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i(A_i, S_i)I(A_i = a)\pi(S_i; \alpha)}{f_2(A_i, S_i | Z_i, Z_{\mathcal{N}_i})}, \tag{1}$$

and the IPW marginal estimator as

$$\widehat{Y}^{IPW_2}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i(A_i, S_i)\pi(A_i, S_i; \alpha)}{f_2(A_i, S_i | Z_i, Z_{\mathcal{N}_i})}. \tag{2}$$

The propensity score $f_2(A_i, S_i | Z_i, Z_{\mathcal{N}_i})$ is the joint probability distribution of individual treatment and nearest neighbors treatment given the covariates Z_i and $Z_{\mathcal{N}_i}$.

Variance Estimators

- The large sample variance estimators can be derived using M-estimation theory.

- We assume that the observed network can be expressed as the union of connected subnetworks, referred to as components of the network. Consider a social network with n participants and m components $\{C_1, C_2, \cdots, C_m\}$.

- We assume that the m components are a random sample from the infinite super-population of groups and the size of each component is bounded.

- To perform inference, we use m independent components (i.e., subnetworks), while we preserve the underlying connections comprising the network structure of each component.

Simulation

We generate a degree 4 regular network with m components using the following steps:

1. Generate a connected regular network with degree 4 where the number of nodes is sampled from the Poisson(10) distribution.
2. Repeat m times.
3. Constitute these m networks.

Note: 10, 50, 100, 200 components regular networks are generated for studying the impact of the number of components on the performance of IPWs.

Simulation

Given a network with $k = 1, 2, ..., m$ components, we generate the potential outcomes and the observed outcomes.

1. Assign the random effects $b_k \sim N(0, 0.5^2)$ to each component and the censoring random effects $\rho_k \sim N(0, 0.3^2)$
2. We generate the covariate $Z_i \sim \text{Bern}(0.5)$

Simulation

3. We then generate the potential outcomes

$$y_i(a_i, a_{\mathcal{N}_i}) \sim \text{Bern}(\text{logit}^{-1}(q)),$$

where $q = -1.75 + 0.5 \cdot a_i + \frac{\sum a_{\mathcal{N}_i}}{d_i} - 1.5 a_i \cdot \frac{\sum a_{\mathcal{N}_i}}{d_i} + 0.5 Z_i$

4. The treatments is generated as

$$A_i \sim \text{Bern}(\text{logit}^{-1}(0.7 - 1.4 \cdot Z_i + b_k))$$

where the participant i belongs to component k.

5. Based on the treatments, extract the corresponding observed outcomes from the potential outcomes.

Simulation Results

Figure 1: Absolute bias (left) of IPW_1 (top) and IPW_2 (bottom) estimator and corresponding Wald 95% CIs empirical coverage probability (right) for different number of components
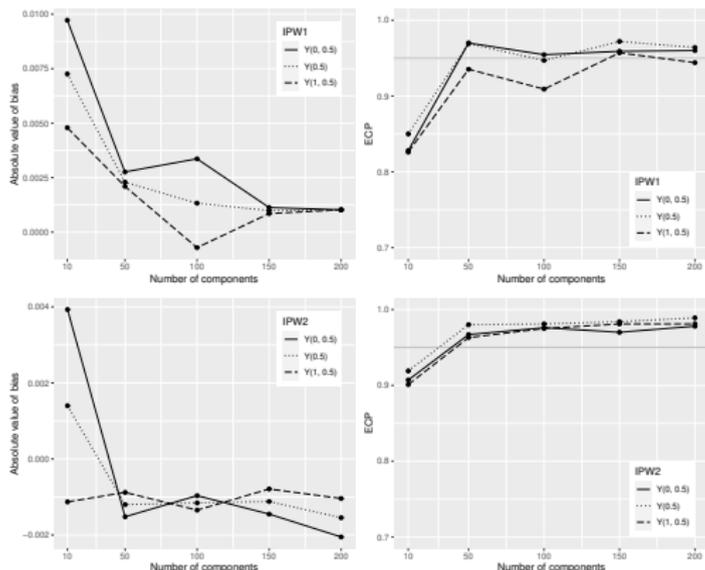
Sensitivity Analysis

Results from 1000 simulation dataset on a network with 100 components for
IPW_1 (left) and IPW_2 (right) for treated ($a = 1$), not treated ($a = 0$), and
marginal estimators under allocation strategies 25%,
50%, and 75% using exposure generating model $A_i = \text{Bern}(\text{expit}(0.7 - 1.4 \cdot Z_i))$.

	IPW1				IPW2			
	Bias	ESE	ASE	ECP	Bias	ESE	ASE	ECP
$\widehat{Y}(1, 0.25)$	0.0013	0.048	0.046	0.91	0.0149	0.039	0.036	0.86
$\widehat{Y}(1, 0.5)$	-0.0003	0.027	0.028	0.96	0.0016	0.022	0.024	0.96
$\widehat{Y}(1, 0.75)$	-0.0032	0.041	0.041	0.93	0.0093	0.035	0.033	0.88
$\widehat{Y}(0, 0.25)$	-0.0051	0.039	0.040	0.94	0.0093	0.033	0.031	0.90
$\widehat{Y}(0, 0.5)$	-0.0018	0.028	0.029	0.95	0.0018	0.023	0.025	0.97
$\widehat{Y}(0, 0.75)$	0.0004	0.053	0.051	0.92	0.0195	0.045	0.042	0.83
$\widehat{Y}(0.25)$	-0.0035	0.032	0.032	0.94	0.0107	0.027	0.026	0.88
$\widehat{Y}(0.5)$	-0.0010	0.021	0.021	0.97	0.0017	0.016	0.018	0.97
$\widehat{Y}(0.75)$	-0.0023	0.033	0.030	0.93	0.0118	0.029	0.028	0.86

Community Alerts and HIV Risk Behavior in TRIP at 6 months

- We apply IPW1 and IPW2 to estimate the causal effects of community alerts at baseline on report of risk behavior at the six-month visit.
- The community alerts intervention status of the index individual and their neighbors was defined with respect to the baseline visit date for the index person.
- The network structure in TRIP had 10 connected components (i.e. observed subnetworks) with 217 participants and 363 shared connections (average degree is 3.35) after excluding isolates.
- Among the 217 participants in TRIP, 25 participants (11.5%) have received a community alert about the increased risk for HIV acquisition in close proximity in their network.

Descriptive statistics of TRIP network (isolates removed)

Characteristics	Nodes	217
	Edges	363
	Components	10
	Average Degree (SD)	3.35 (2.75)
	Density	0.0155
	Transitivity	0.25
	Assortativity	0.24
Community alert	Exposed	25 (11.5%)
	Not Exposed	192 (88.5%)
HIV Status	Positive	113 (52.1%)
	Negative	104 (47.9$)
Date of first interview	Before ARISTOTLE ended	105 (48.4%)
	After ARISTOTLE ended	112 (51.5%)
Education	Primary School or less	63 (29%)
	High School (first 3 years)	69 (32%)
	High School (last 3 years)	53 (24%)
	Post High School	32 (15%)
Employment status	Employed	33 (15.2%)
	Unemployed; looking for work	54 (24.9%)
	Can't work; health reason	102 (47%)
	Other	28 (12.9%)
Shared injection equipment in last 6 months	Yes	58 (26.7%)
	No	159 (73.3%)
Outcome: sharing injection equipment at the 6-month visit	Yes	126 (58%)
	No	91 (42%)

Baseline covariates in TRIP study

- HIV test status
- The date of first interview
- Shared drug equipment in the past 6 months (needles, syringes)
- Education (primary school, high school, and post high school)
- Employment status (employed, unemployed/looking for a job, cannot work because of health reason, and others)

Community Alerts and HIV Risk Behavior in TRIP at 6 months
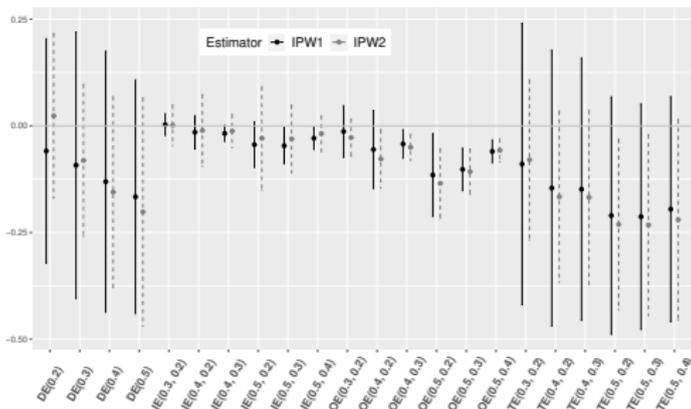
Figure 2: The risk difference estimates and the Wald 95% confidence intervals of direct, indirect, total, and overall effects under allocation strategies 25%, 50%, and 75%.

Conclusions

- The nearest neighborhood IPWs were demonstrated to be consistent and asymptotically normal.
- Closed-form estimator of the asymptotic variance was derived.
- In the simulation study, the both IPW1 and IPW2 estimators performed well when given a large number (> 100) of components in the network.
- When the exposure mechanism is misspecified, IPW2 had coverage below the nominal level while IPW1 still maintained proper coverage levels.
- When more baseline covariates were added into the exposure generating model, both estimators had coverage below the nominal level while IPW_2 had slightly higher coverage than IPW_1.
- In TRIP, we found that the community alerts are protective overall and evidence of spillover benefits.

Github R Code

We have R functions available to run the IPW Nearest Neighbor Estimators.

https://github.com/uri-ncipher/Nearest-Neighbor-estimators

Selected Reference

- EJ Tchetgen Tchetgen, and TJ Vander Wheele, *On Causal inference in the presence of interference*, *Statistical Methods in Medical Research*, 21(1):55-75, 2012
- Ashley L Buchanan, Sten H Vermund, Samuel R Friedman, and Donna Spiegelman (2018). Assessing individual and desseminated effects in network-randomized studied, *American Journal of Epidemiology*, 187(111), 2449-2459.
- TingFang Lee, A Buchanan, N Katenka, L Forastiere, M Halloran, S Friedman, and G Nikolopoulos, *Estimating Causal Effects of Non-Randomized HIV Prevention Interventions with Interference in Network-based Studies among People who Inject Drugs*, Annals in Applied Statistics (accepted).
- JG Ibrahim, MH Chen, SR Lipsitz, and AH Herring (2005) *Missing-Data Methods for Generalized Linear Models*, Journal of the American Statistical Association, 100:469, 332-346