Objective Assessment of Depression

Image credit:
Black dog institute

a part of the
MIT-MGH
Grand Challenge
Depression Study

Asma Ghandeharioun, MIT Media Lab, Affective Computing group

Depression: Highly Prevalent and Disabling

- Depression is the leading cause of ill health and disability worldwide.
- According to the latest estimates from WHO, more than 300 million people are now living with depression, an increase of more than 18% between 2005 and 2015.

Major Depressive Disorder (MDD): Clinical Diagnosis

- $\geq$ 5 of 9 core depressive symptoms
- $\geq$ 2 weeks
- One of the 5 symptoms must be:
  - **Depressed Mood***
  - **Marked loss of interest or pleasure***
- Other symptoms: sleep, energy, appetite, concentration, guilt, suicidality, psycho-motor agitation or retardation

Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, American Psychiatric Association 2013

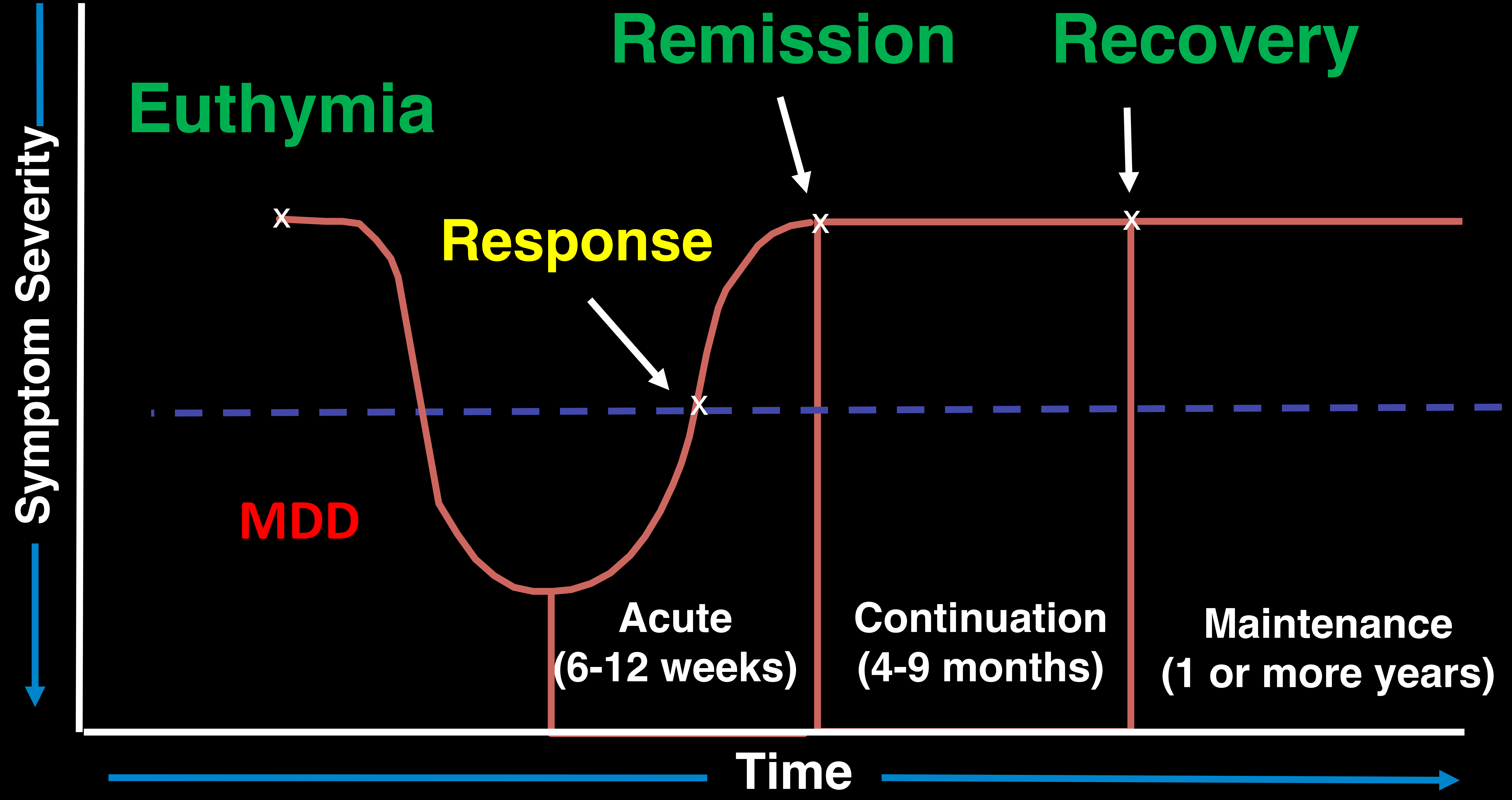
Major Depressive Disorder (MDD): Clinical Diagnosis

- $\geq$ 5 of 9 core depressive symptoms
- $\geq$ 2 weeks
- One of the 5 symptoms must be:
  - **Depressed Mood***
  - **Marked loss of interest or pleasure***
- Other symptoms: sleep, energy, appetite, concentration, guilt, suicidality, psycho-motor agitation or retardation

This approach was invented in 1960s!

Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, American Psychiatric Association 2013

The Idealized Phases of Treatment of MDD



Adapted from Kupfer DJ. *J Clin Psychiatry* 1991;52 (Suppl):28–34.

The Clinical Reality …

The Clinical Reality … is different!

The Clinical Reality … is different!

Low rates of remission:

- Only ~ 30% with 1 treatment trial
- Only ~ 67% with 4 sequential treatment trials
- ~ 50% relapse during the first year after remission

Even in optimal clinical settings, patients with MDD are monitored cross-sectionally with clinical interviews and symptom rating scales:

- Inadequate for subtyping of MDD, a condition marked by considerable heterogeneity
- Inadequate for capturing variability of time
- Inadequate for prediction of response or relapse

Paradigm Shift: Hypotheses

Longitudinal, multi-modal physiological measures in depression evaluation and treatment will:

- Improve evaluation and diagnosis including meaningful depression subtyping
- Enhance prediction/early detection of response
- Improve prediction/early detection of relapse
- Inform more effective "personalized" treatments

Objective Assessment of Depression

Can we predict HDRS from passive sensor data?

Provide an objective diagnosis of depression

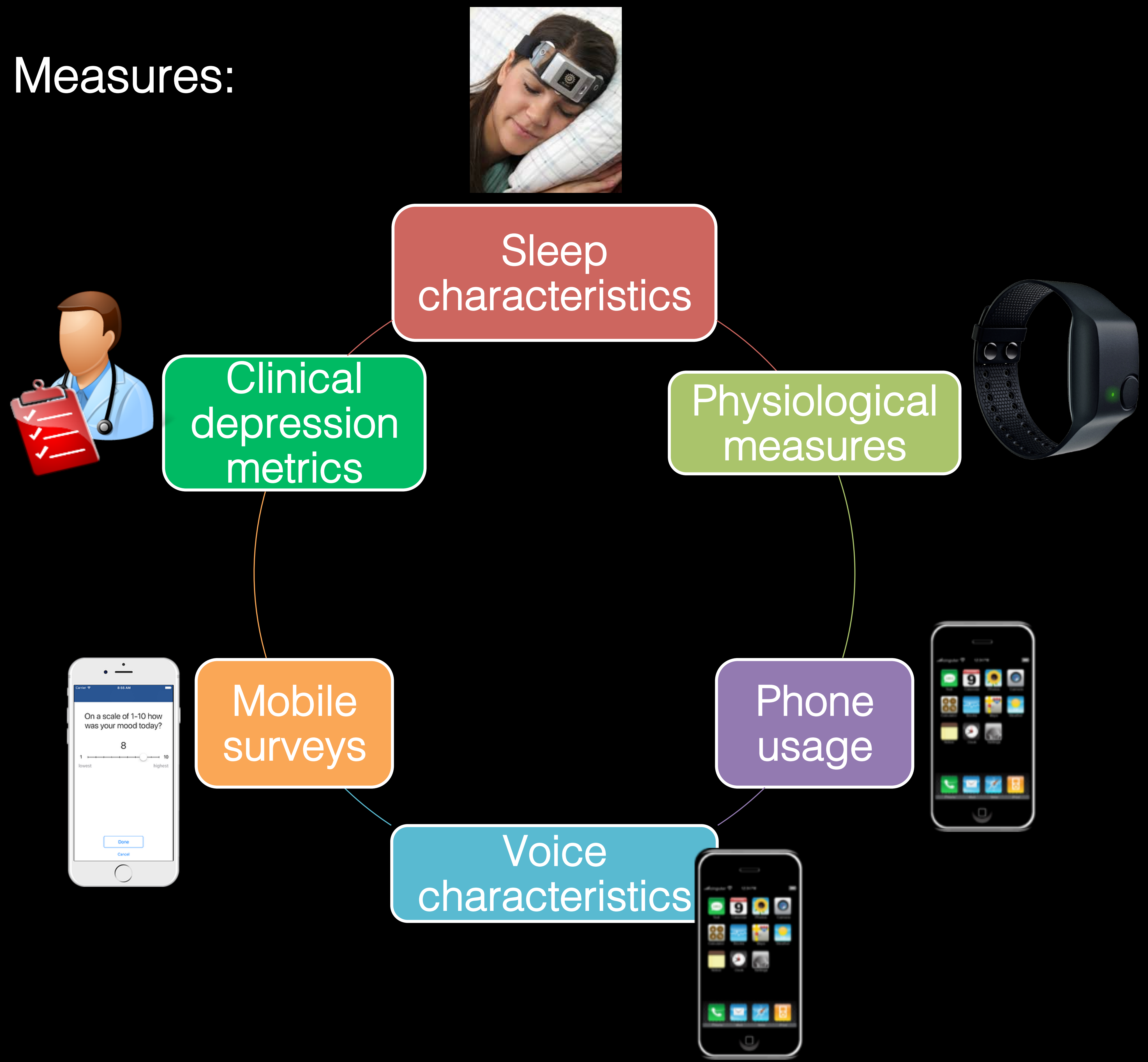
Predict early treatment response

Anticipate signs of relapse

Objective Assessment of Depression

Study participants:
50 people with Moderate Depression monitored 24/7 for 8 weeks
Currently 4 HC and 12 MDD have finished the protocol

Measures:



Collaborators:

Rosalind Picard ScD
Szymon Fedor PhD

David Mischoulon MD PhD
Paola Pedrelli PHD
Esther Howe BA

Past members:

Jonathan Alpert MD PhD
Dawn Ionescu MD
Lisa Sangermano BA
Chelsea Dale BA

Objective Assessment of Depression

Clinical depression measures:

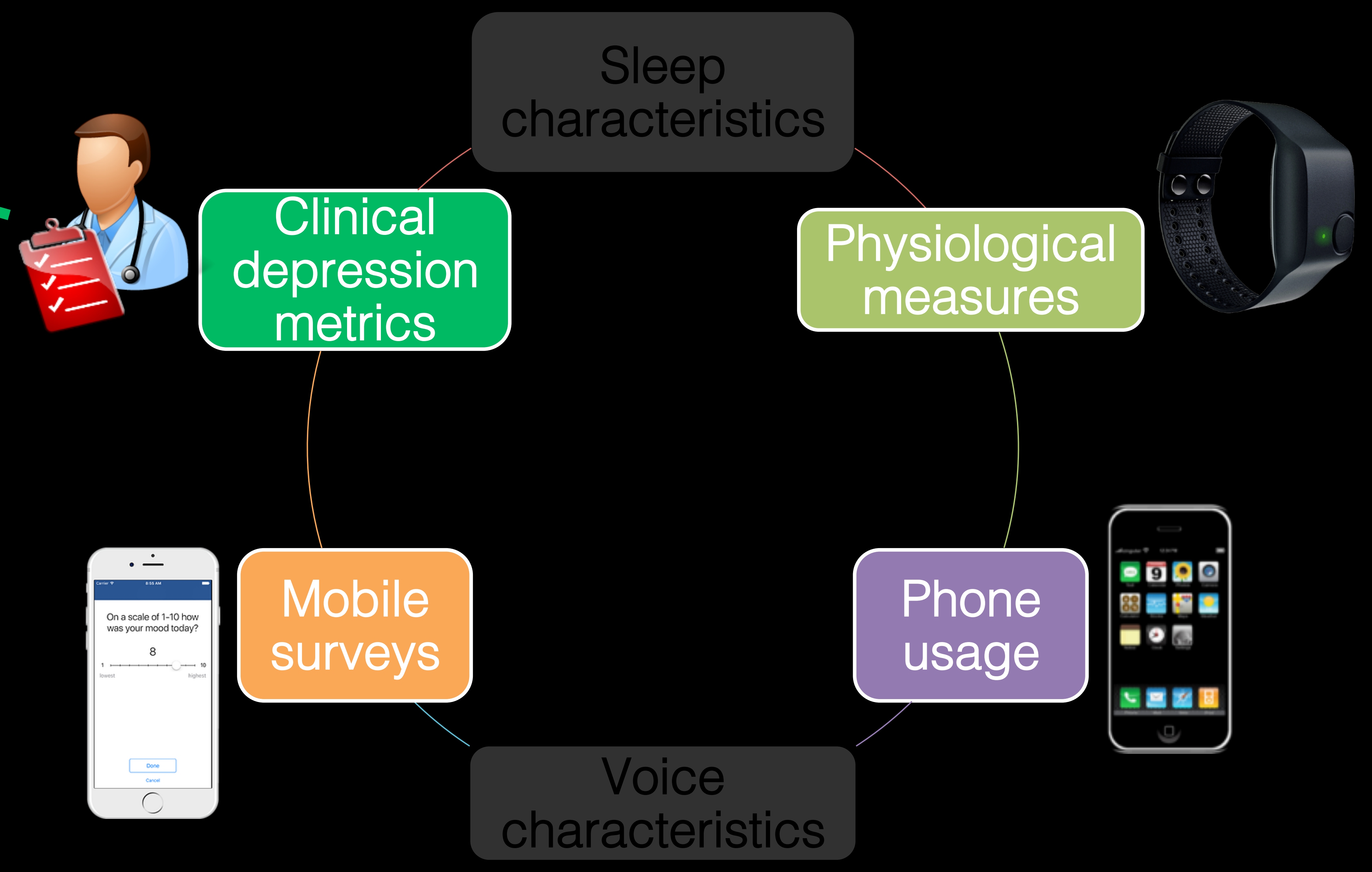
- Assessment visit at clinic
- Biweekly

Captures:

- Hamilton Depression Rating Scale (HDRS)

Dataset size: 59

Measures:



Objective Assessment of Depression

Mobile surveys:

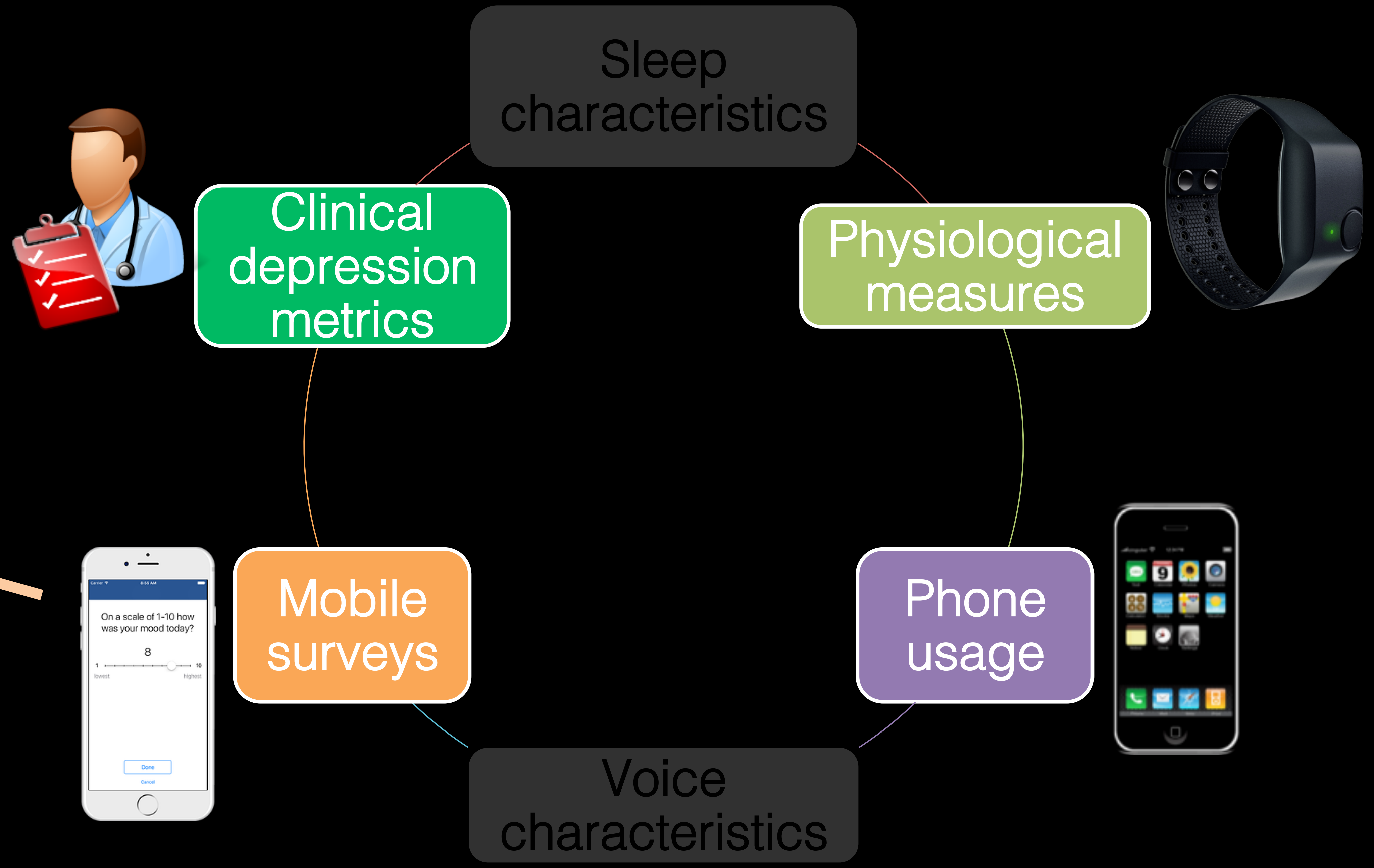
- Movisens on Android
- 4 times a day

Captures:

- Health condition
- Sleep
- Mood
- Stress
- Anxiety
- Alcohol/Caffeine/Drug consumption
- Social interactions

Dataset size: 503

Measures:



Objective Assessment of Depression

Phone usage:

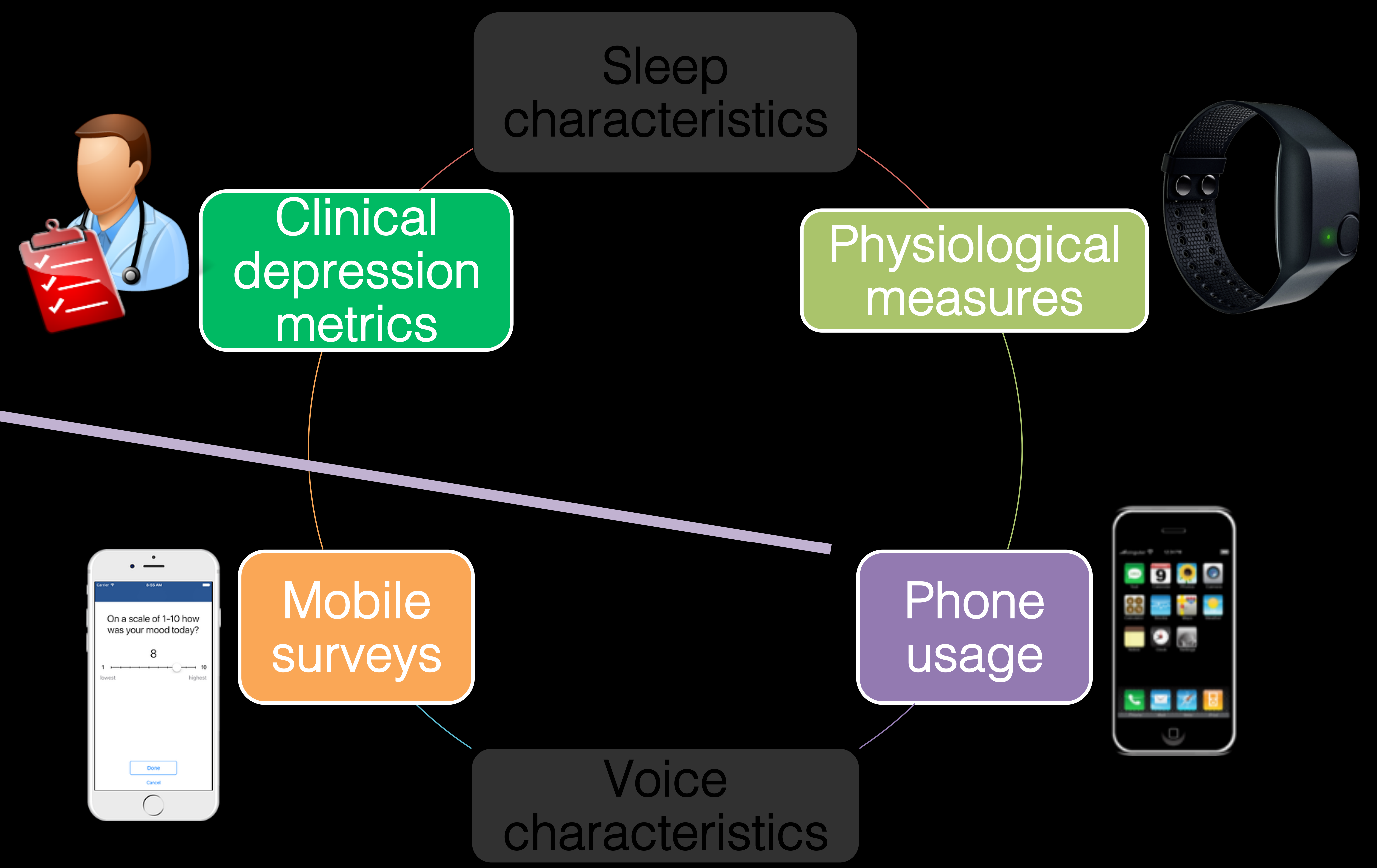
- Movisens on Android
- 24/7 measurement

Captures:

- Calls
- Text messages
- App usage
- Display on/off behavior
- Location

Dataset size: 605

Measures:



Objective Assessment of Depression

Physiological data:

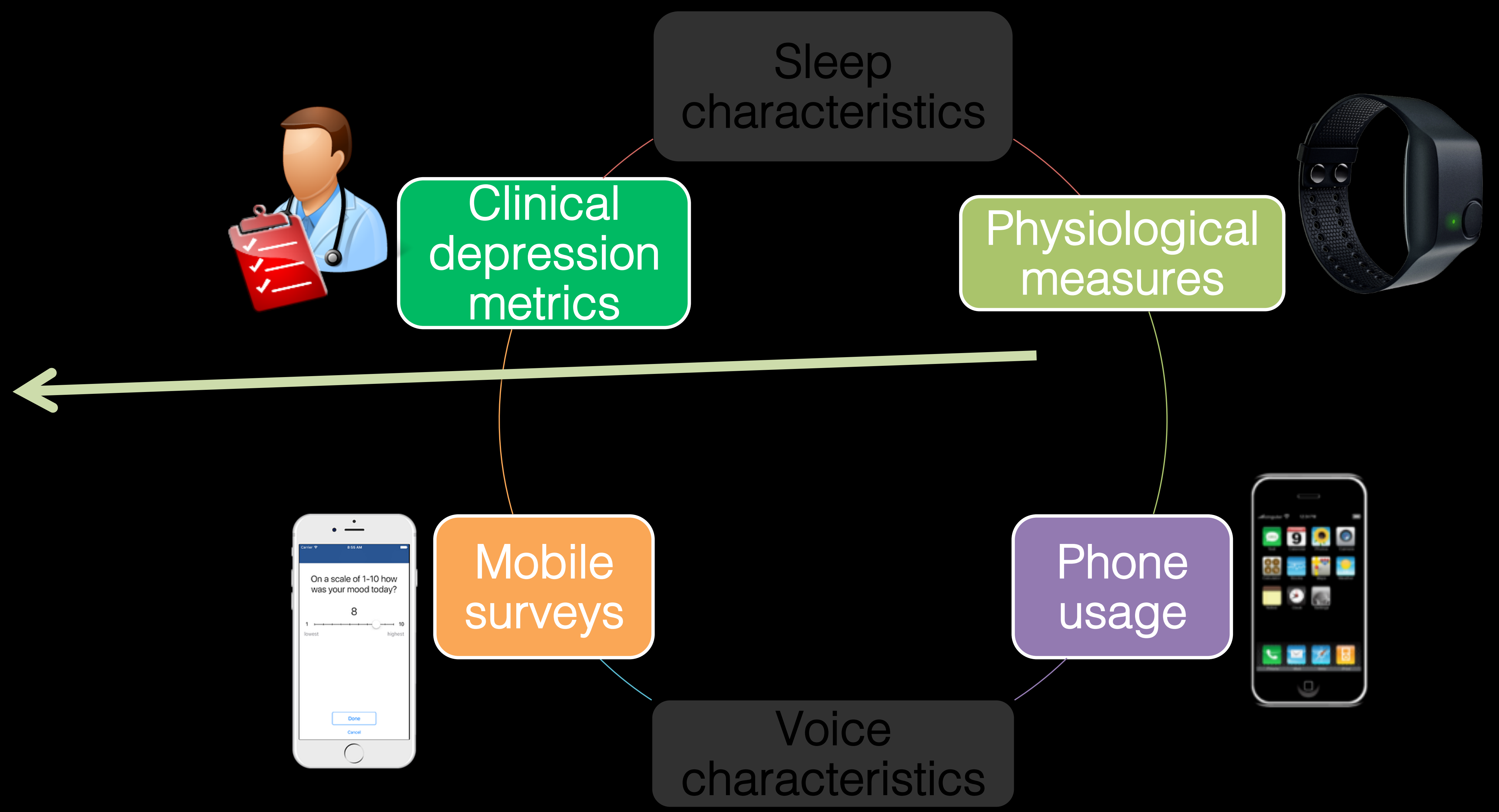
- An Empatica E4 sensor on each wrist
- 24/7 measurement

Captures:

- Electrodermal activity (EDA)
- Skin temperature
- 3-axis accelerometer data

Dataset size: 540

Measures:



Feature Architecture

Physiology

- EDA: mean, median, sd, SCR, asymmetry, SCL and SCR through convex optimization approach, recording time fraction
- Motion: mean, median, sd, recording time fraction
- Sleep: duration, onset, max uninterrupted, number of wake-ups, wake-up time, recording time fraction

Passive Phone Data

- Call: number of incoming/outgoing/missed calls, duration (mean, median, sd), incoming/outgoing ratio (number and duration)
- SMS: number of incoming/outgoing messages, ratio of incoming/outgoing
- Location: mean, median, sd of location for latitude, longitude, and overall
- Display on/off: mean, median, sd (duration, number)
- App usage: categorized the apps, mean, median, sd of duration and number of times each category was used

Interactive Surveys

- Preprocessing survey questions
- Mood: total PA/NA, weekly average, weighted weekly average, NA/PA ratio (daily, weekly, average weekly), sd of mood (weekly, overall)

Clinical Measures

- HDRS
  - Normal: 0-7
  - Mild: 8-13
  - Moderate: 14-18
  - Severe: 19-22
  - Very severe: >22

Feature Transformation and Selection

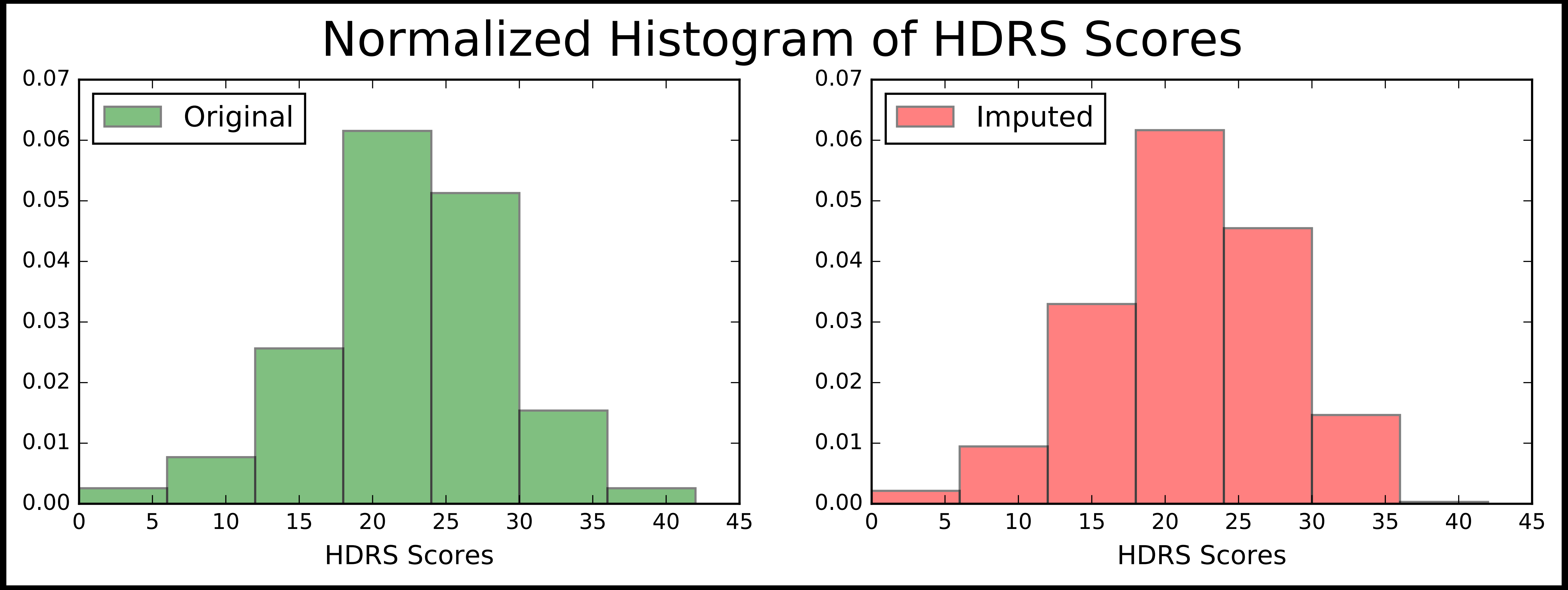
- Over 700 features for ~500 data points
- To avoid over fitting:
  - Regularization tricks
  - Transformations while balancing between the number of selected features and explaining enough variance of the data
    - PCA
    - Kernel PCA with radial-basis function kernel
    - Truncated SVD

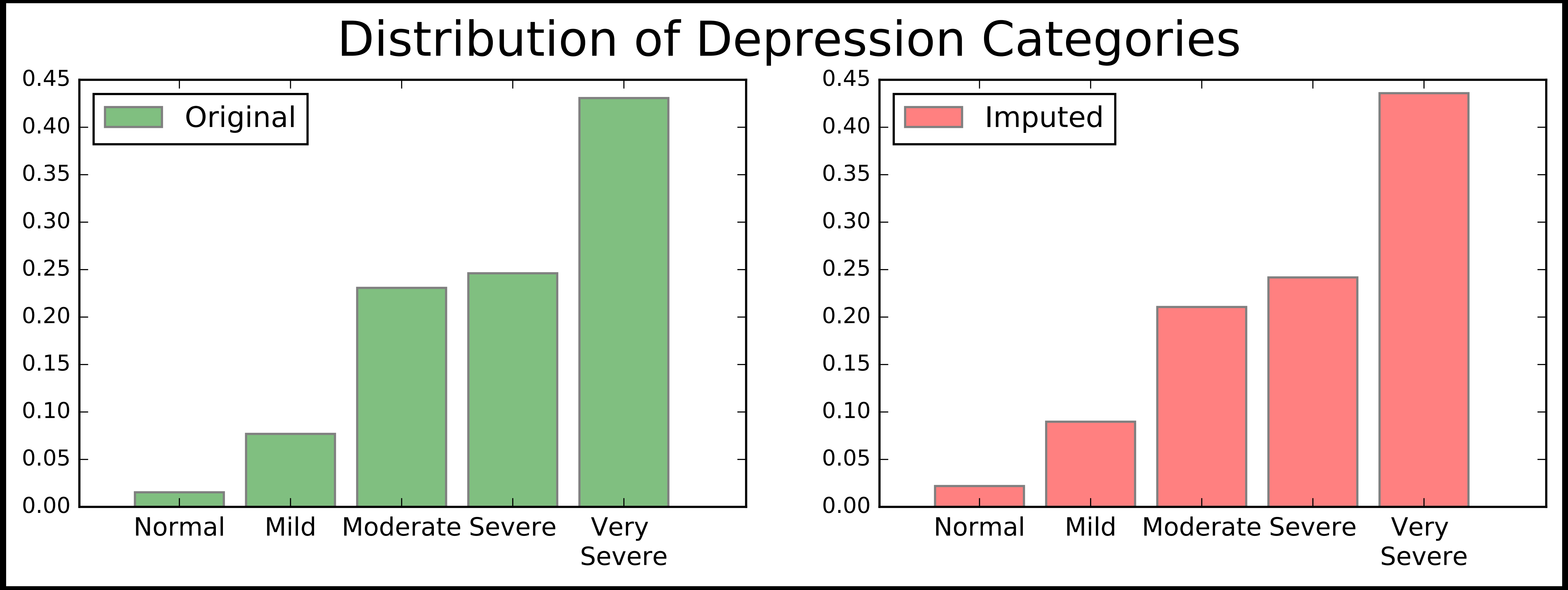
HDRS Imputation based on Survey Data

- Models
  - Regression: Ridge, Lasso, ElasticNet
  - Robust to outliers: Theil-sen, RANSAC, huber
  - Boosting (AdaBoost, Gaussian Boosting)
  - Random Forest
- Validation
  - 90% train, 10% test
  - Leave-one-out cross-validation on train

Model Info.	Name	Ridge (L2-Regularized Regression)
	Dataset	Mood Subset (PANAS)
RMSE	Validation	3.4
	Test	2.8
	Baseline 1 (Average)	6.8
	Baseline 2 (Median)	6.8

HDRS Imputation based on Survey Data





- Kolmogorov-Smirnov test for comparing distributions.
- We were not able to reject the null hypothesis of the data points coming from the same distribution.
- Continuous HDRS:

$$D_{original}(M = 21.5, SD = 6.4)$$
$$D_{imputed}(M = 21.2, SD = 6.3)$$
$$ks - statistic = 0.08, p = 0.83$$

- Discrete categories:

$$ks - statistic = 0.01, p = 1.00$$

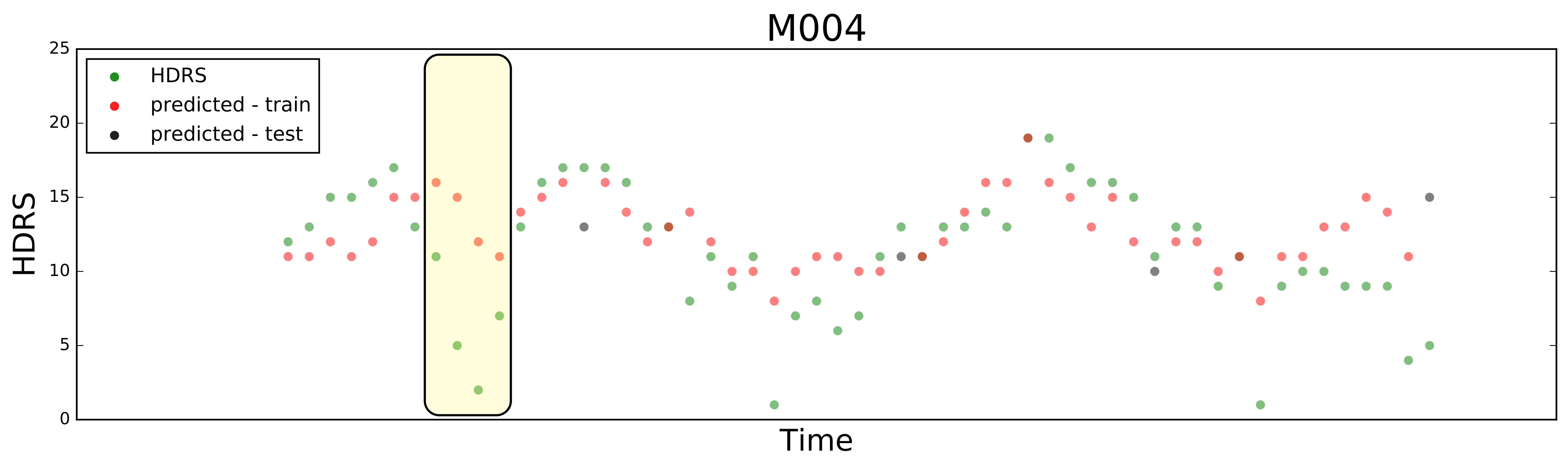
HDRS Prediction based on Sensor Data

- Models
  - Regression
  - Robust to outliers
  - Boosting
  - Random Forest
  - Gaussian Process
  - Customized ensemble method

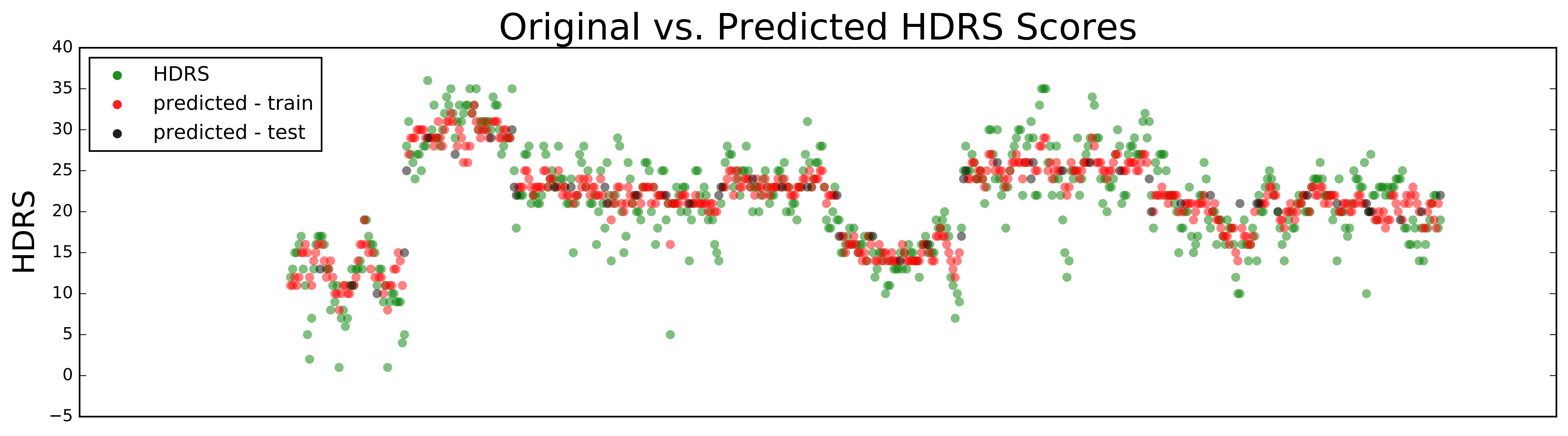
- Validation
  - 90% train, 10% test
  - Test set only selected from original HDRS values (no imputed)
  - Resembles deployment: no test data from the first two weeks
  - 10-fold cross-validation on train data

HDRS Prediction based on Sensor Data

Model Type	Model	Parameters	Dataset	RMSE Validation	RMSE Test	Baseline Average	Baseline Median
Regression	Regression		Kernel PCA subset	5.2	4.9	7.1	7.1
Robust	Ransac	ms=0.3	Kernel PCA subset	5.0	4.9	7.1	7.1
Boosting	AdaBoost	n=50, lr=1	Subset data	5.5	4.6	7.1	7.1
Random Forest	-	n=15	Subset data	5.4	4.6	7.1	7.1
Gaussian Process	-	α=0.1, n=5	Kernel PCA subset	5.3	5.5	7.1	7.1
Overall Ensemble		k=1	selected by individual models	5.8	4.5	7.1	7.1



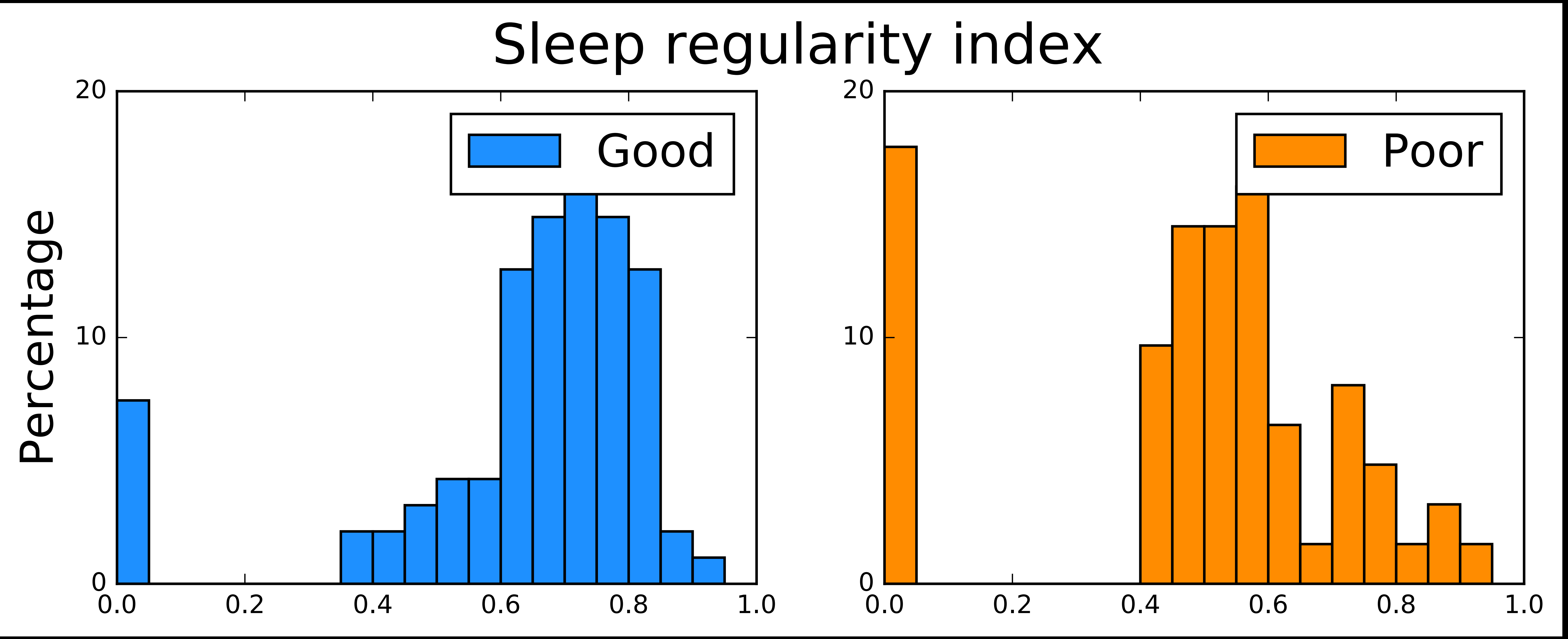
HDRS Prediction based on Sensor Data



Post-hoc Interpretability Analyses

Sleep regularity index:

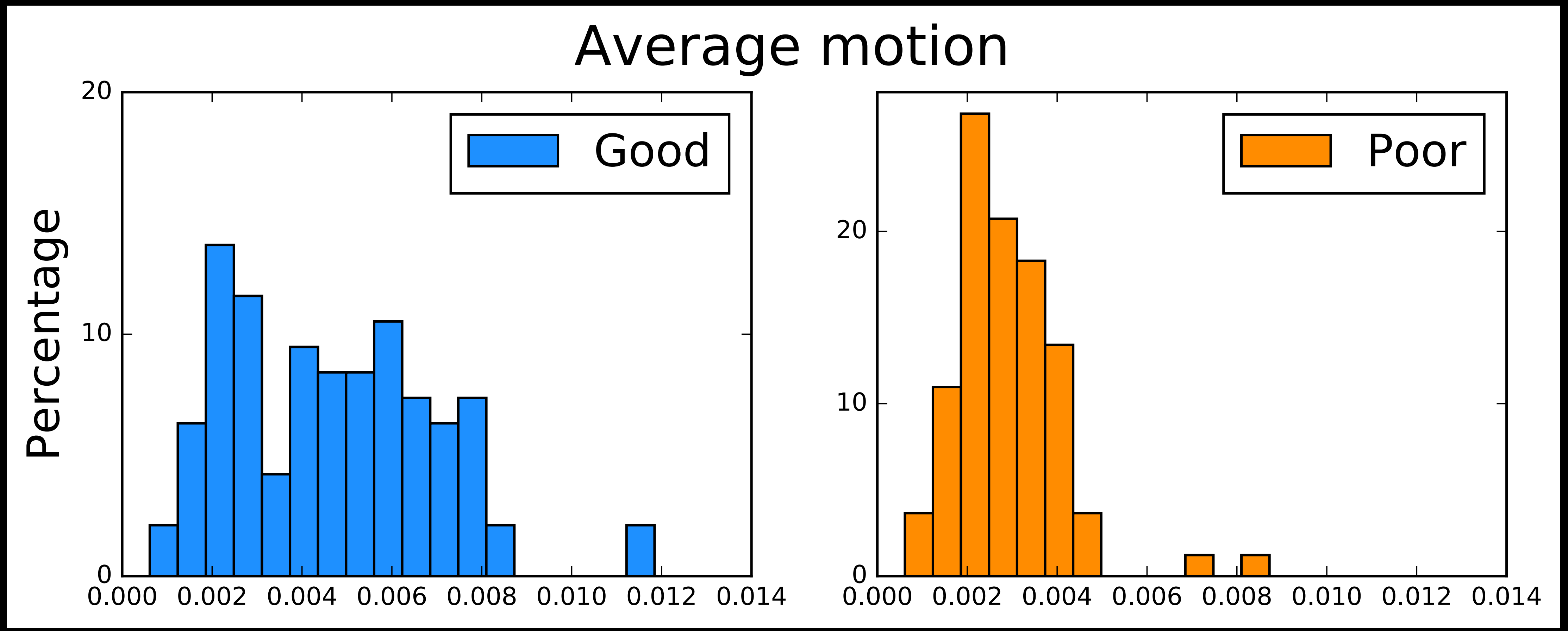
$$SRI = \frac{1 + \frac{1}{T-\tau}\int_0^{T-\tau} s(t)s(t+\tau)dt}{2}$$

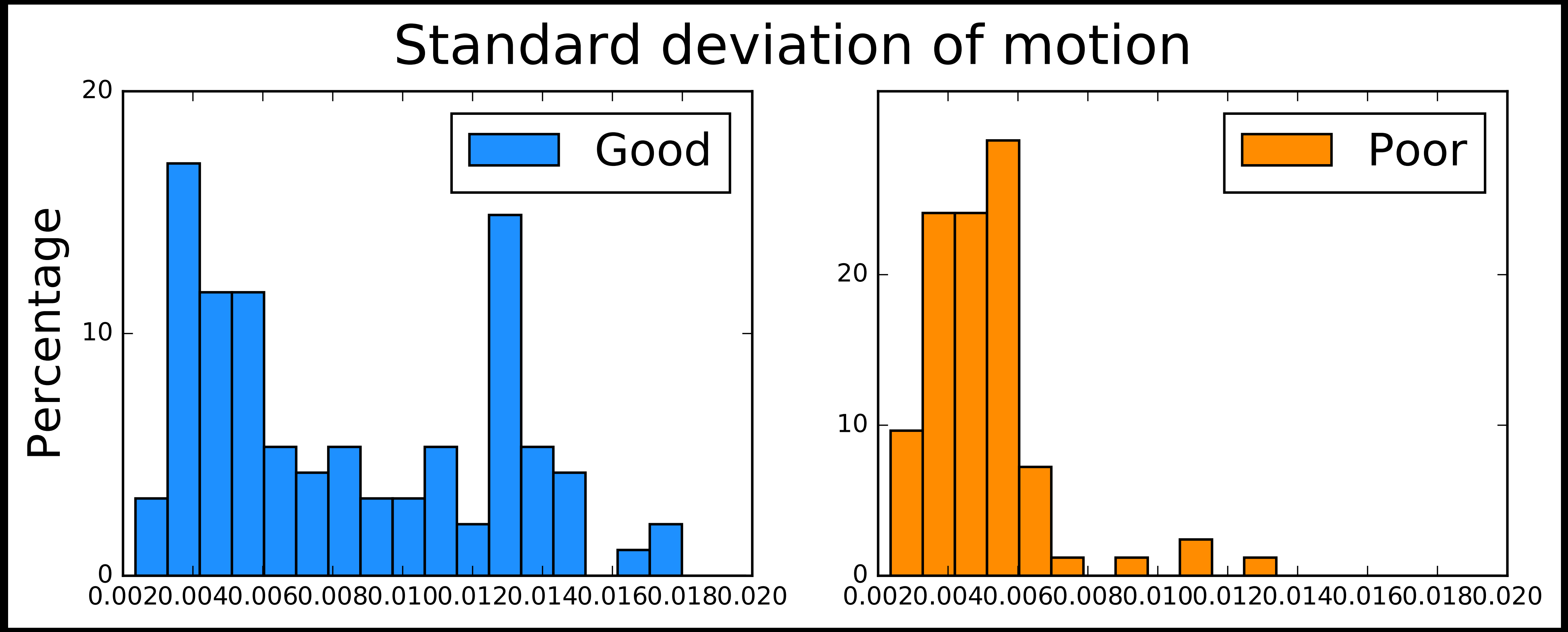


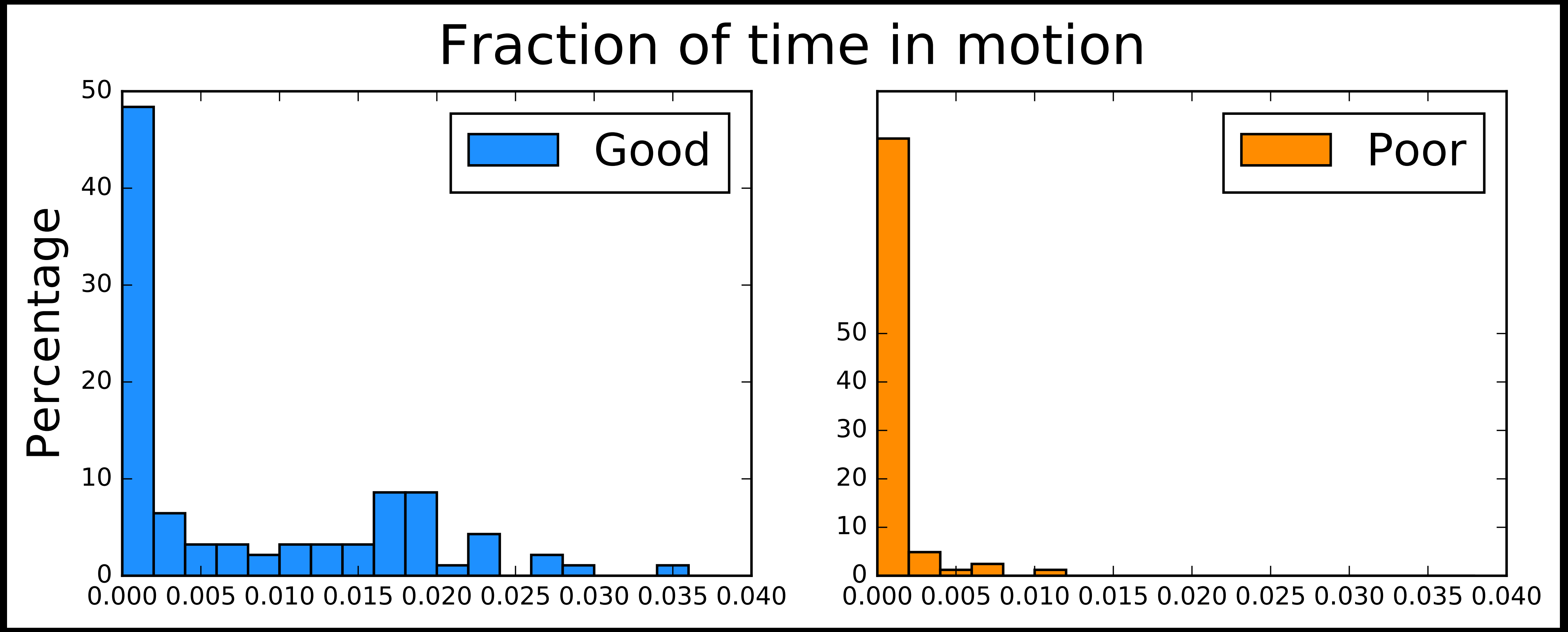
Post-hoc Interpretability Analyses

Motion:

$$VM = \sum_{t=0}^{N} VM_t + |R_{(z,t)} - M_z|$$



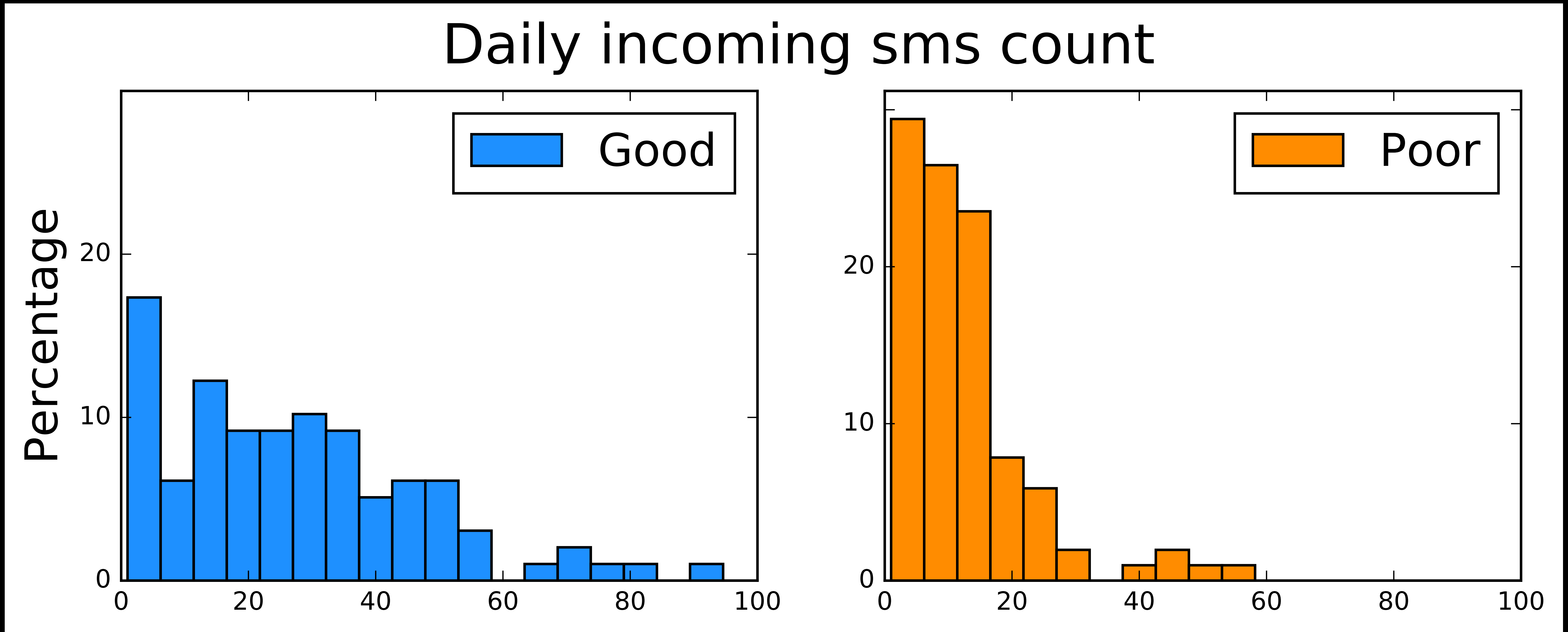




Post-hoc Interpretability Analyses

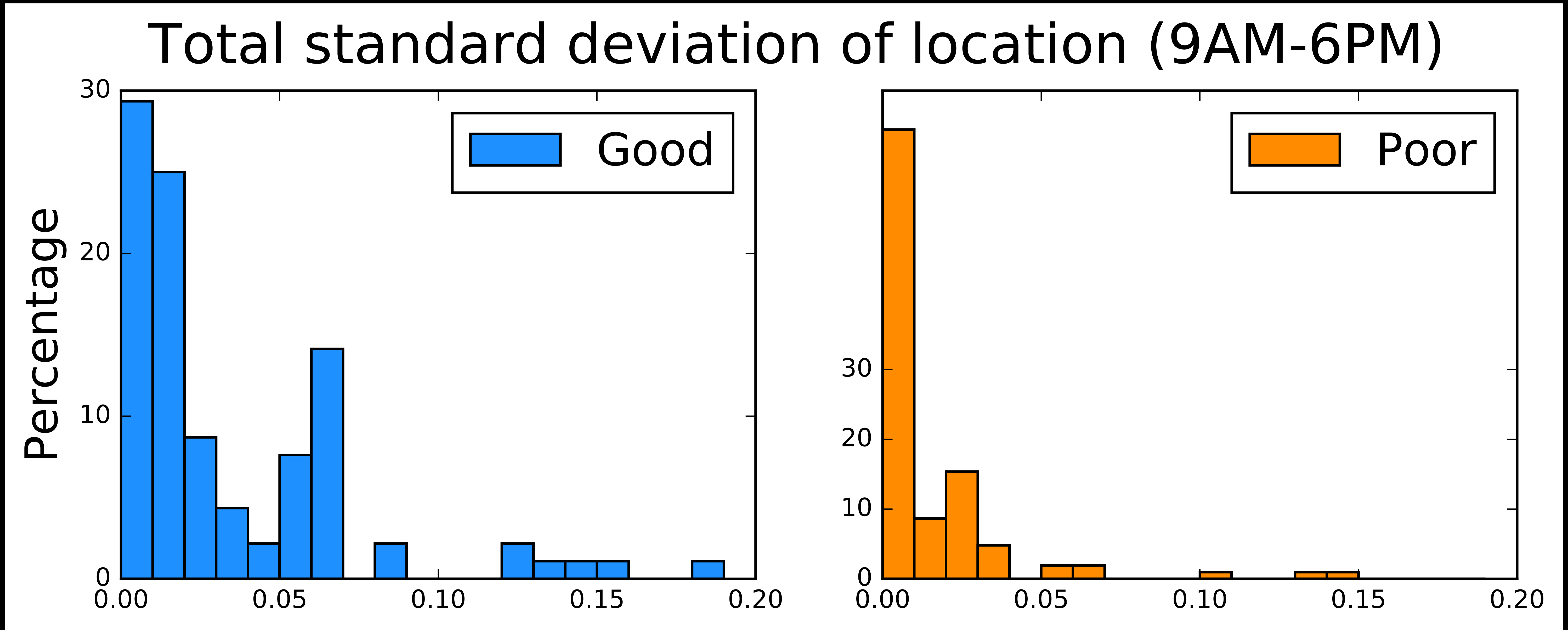
Communication

(incoming messages)



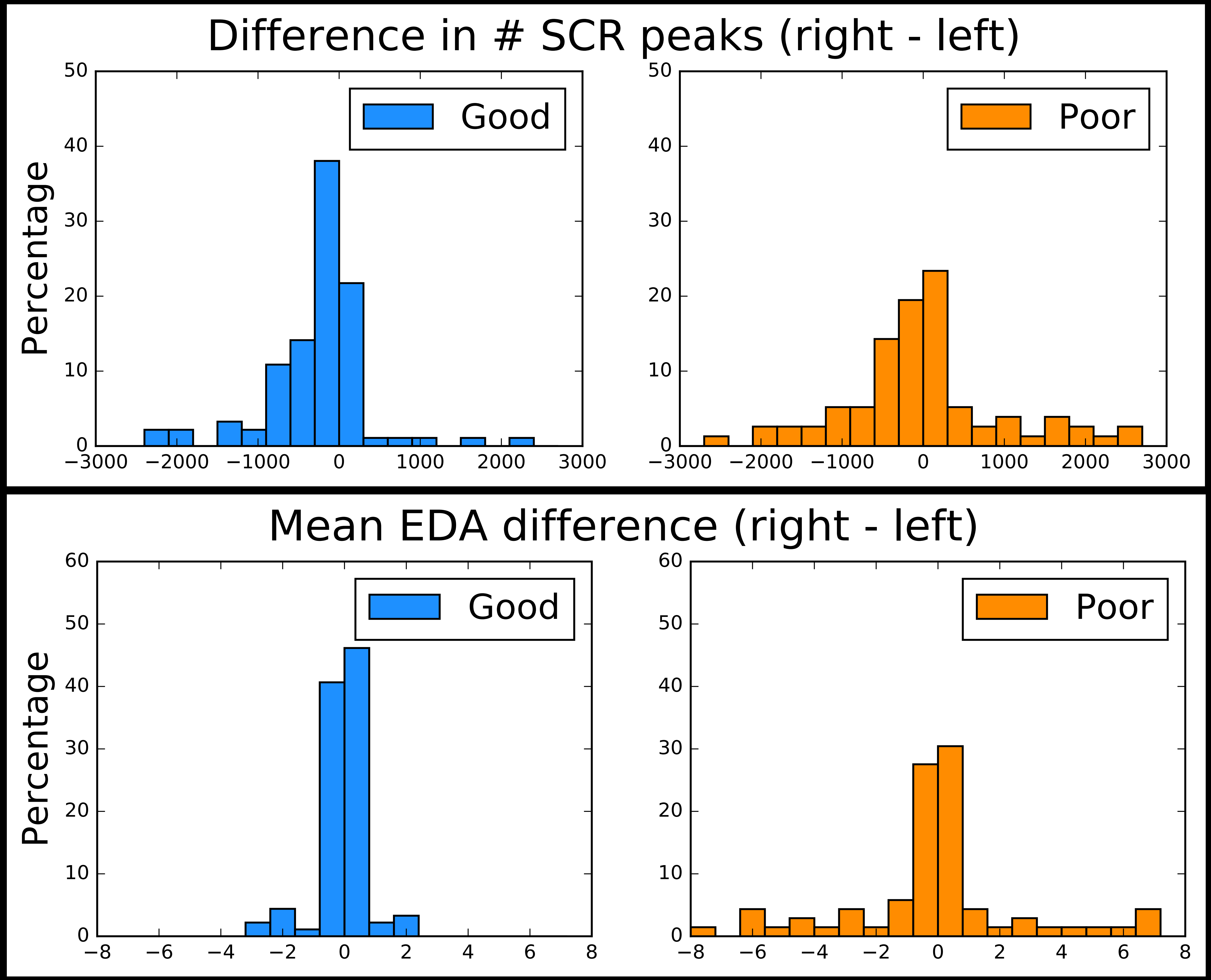
Post-hoc Interpretability Analyses

Location patterns



Post-hoc Interpretability Analyses

EDA asymmetry



Summary

- Generated several features
- Conducted several feature transformations
- Imputed missing HDRS values using survey data
- Predicted HDRS using passive data
- Analyzed the associations between mental health and several behavioral/physiological features