Philipps Universität Marburg

Associations between ecological momentary assessment and passive sensor data in a large student sample

Björn Siepe[1] **Rayyan Tutunji**[2] **Carlotta L. Rieble**[2] **Ricarda K. K. Proppert**[2] **Eiko I. Fried**[2]

May 28, 2025 – Society for Ambulatory Assessment

[1]Department of Psychology, University of Marburg [2]Department of Clinical Psychology, Leiden University

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Measurement **challenges**:

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Measurement **challenges**:

Many small sample studies

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Measurement **challenges**:

Many small sample studies

Unclear relationship between self-report and sensors

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Measurement **challenges**:

Many small sample studies

Unclear relationship between self-report and sensors

Strong focus on prediction

Digital Phenotyping in Mental Health Research

- Many **promises** of wearable sensor data: Scalable forms of assessment, high temporal resolution, more “objective”(?) measurements

Measurement **challenges**:

Many small sample studies	Unclear relationship between self-report and sensors	Strong focus on prediction	Proprietary algorithms

This Project

- Investigate overlap between wearable sensor metrics and self-report measures

This Project

- Investigate overlap between wearable sensor metrics and self-report measures
- Focus on three **transdiagnostic markers**:
    - **Stress** (transdiagnostic marker)

This Project

- Investigate overlap between wearable sensor metrics and self-report measures
- Focus on three **transdiagnostic markers**:
    - **Stress** (transdiagnostic marker)
    - **Sleep** (common in DSM-5, robust contributor to psychopathology)

This Project

- Investigate overlap between wearable sensor metrics and self-report measures
- Focus on three **transdiagnostic markers**:
    - **Stress** (transdiagnostic marker)
    - **Sleep** (common in DSM-5, robust contributor to psychopathology)
    - **Tiredness** (indicator of sleep problems, related to depression and other disorders)

WARN-D Study

Recruitment

- Four cohorts

WARN-D Study

Recruitment

- Four cohorts
- 500 students each

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments
- 18-21 items (more on Sundays)

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments
- 18-21 items (more on Sundays)
- Many items: 1-7 Likert scale

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands > 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments
- 18-21 items (more on Sundays)
- Many items: 1-7 Likert scale

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands $>$ 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments
- 18-21 items (more on Sundays)
- Many items: 1-7 Likert scale

Passive Sensing

- Smartwatch: Garmin vivosmart 3

WARN-D Study

Recruitment

- Four cohorts
- 500 students each
- Population: Students of higher education in The Netherlands > 18 years

Data Collection

- Baseline assessment
- Multiple follow-ups

EMA Protocol

- 3-month EMA phase
- 4× daily assessments
- 18-21 items (more on Sundays)
- Many items: 1-7 Likert scale

Passive Sensing

- Smartwatch: Garmin vivosmart 3
- Continuous physiological monitoring

Sensor-EMA Features

Self-report variable	Sensor variable

Sensor-EMA Features

Self-report variable	**Sensor variable**
Self-reported stress *"I feel stressed right now"*	Mean sensor stress Based on HR, HRV, & activity; scored 0-100 (0-25: resting, 26-50: low, 51-75: medium, 76-100: high stress).

Sensor-EMA Features

Self-report variable	Sensor variable
Self-reported stress *"I feel stressed right now"*	Mean sensor stress Based on HR, HRV, & activity; scored 0-100 (0-25: resting, 26-50: low, 51-75: medium, 76-100: high stress).
Self-reported tiredness *"I feel tired right now"*	Mean body battery Based on HRV and activity; scored 0-100 (0-25: charging, 26-100: tiring)

Sensor-EMA Features

Self-report variable	Sensor variable
Self-reported stress *"I feel stressed right now"*	Mean sensor stress Based on HR, HRV, & activity; scored 0-100 (0-25: resting, 26-50: low, 51-75: medium, 76-100: high stress).
Self-reported tiredness *"I feel tired right now"*	Mean body battery Based on HRV and activity; scored 0-100 (0-25: charging, 26-100: tiring)
Self-reported sleep quality *"Last night, I slept well"*	Sensor total sleep duration Total sleep duration in hours

Note: All self-report variables measured on 1-7 Likert scale ("not at all" to "very much")

Methods

Multilevel models in `nlme` with maximum likelihood estimation:

Multilevel models in `nlme` with maximum likelihood estimation:

$$\underset{\text{Sensor outcome}}{y_{it}} = \overset{\text{Fixed intercept}}{\beta_0} + \overset{\text{WP effect}}{\beta_1} (x_{it} - \bar{x}_i) + \underset{\text{BP effect}}{\beta_2} \bar{x}_i + \overset{\text{Time-of-day}}{\beta_3} d_{it} + \underset{\text{Random intercept}}{u_{0i}} + \overset{\text{Random slope}}{u_{1i}} (x_{it} - \bar{x}_i) + \underset{\text{Error}}{e_{it}}$$

Additional Analyses

Analysis	Stress	Tiredness	Sleep
Interaction with Age, Gender, Depression, Cohort	✓	✓	✓
Change of Residual Correlation Structure	✓	✓	✓
Informed “Binning” of Outcome	✓	✓	✗
Alternative Outcome Operationalization	✓	✓	✗
Different Lags & Aggregations	✓	✓	✗

Main Results: EMA-Sensor Associations

Main Results: EMA-Sensor Associations

	Stress	Tiredness	Sleep
WP association	0.49	-1.55	0.33
95% CI	(0.35, 0.63)	(-1.68, -1.42)	(0.31, 0.35)
R^2 (conditional)	0.22	0.42	0.28
RMSE	21.66	20.46	1.33

Main Results: EMA-Sensor Associations

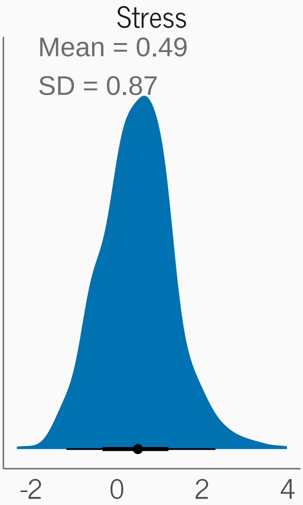
	Stress	Tiredness	Sleep
WP association	0.49	-1.55	0.33
95% CI	(0.35, 0.63)	(-1.68, -1.42)	(0.31, 0.35)
R^2 (conditional)	0.22	0.42	0.28
RMSE	21.66	20.46	1.33

Additional Findings

Cohort Differences: Stress weaker in cohorts 2 & 4 (summer); Tiredness & Sleep stronger in cohorts 2 & 4

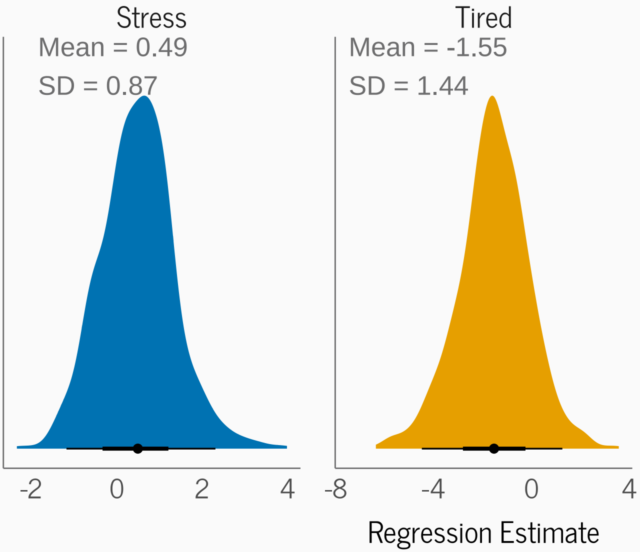
Demographics: No significant effects of age, gender, or depression on associations

Individual Estimates

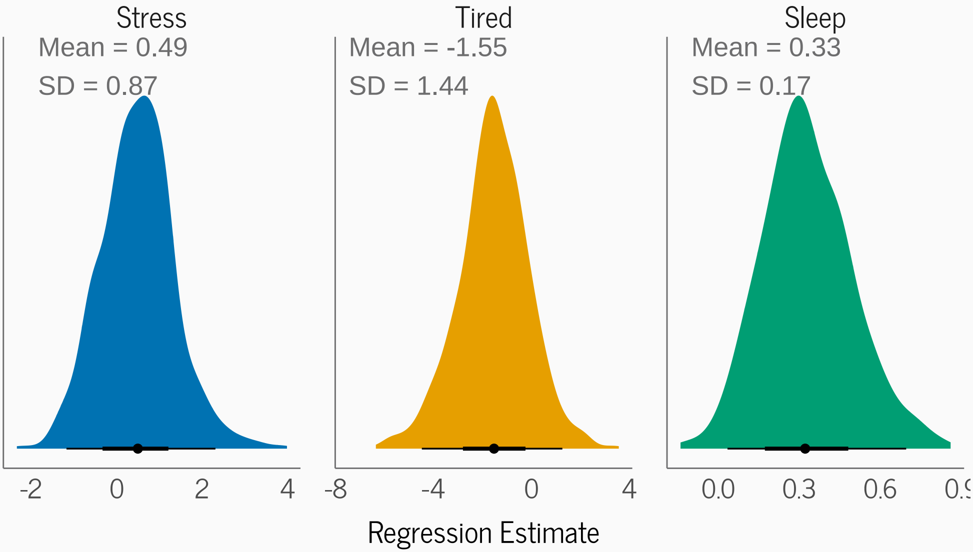


Regression Estimate

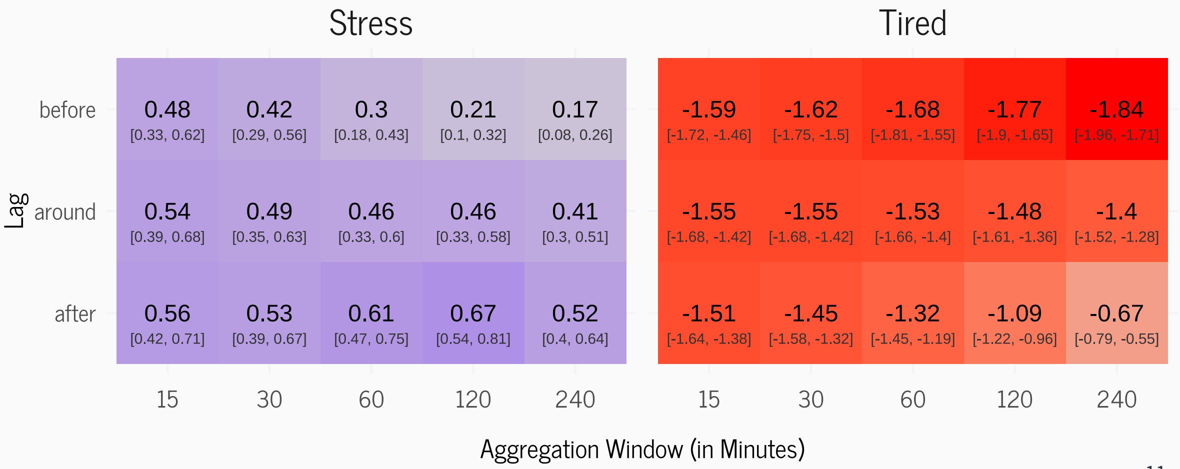
Individual Estimates



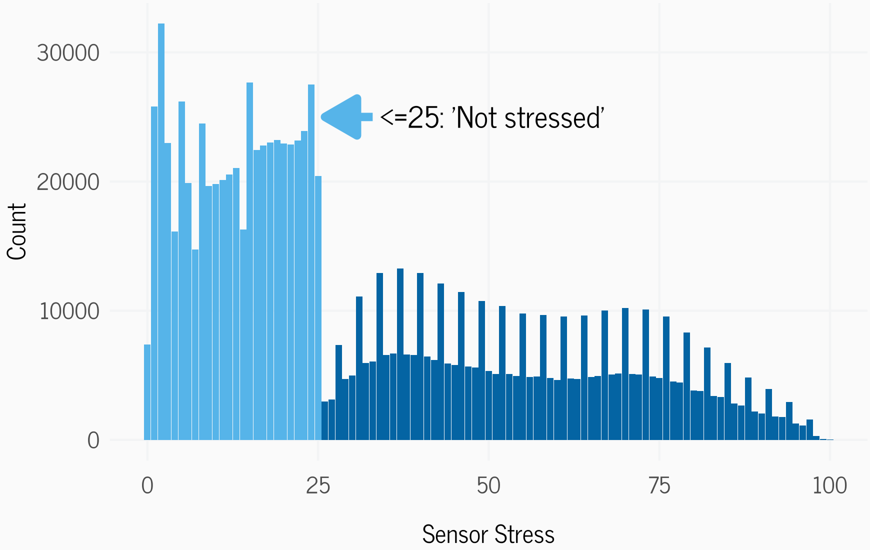
Individual Estimates



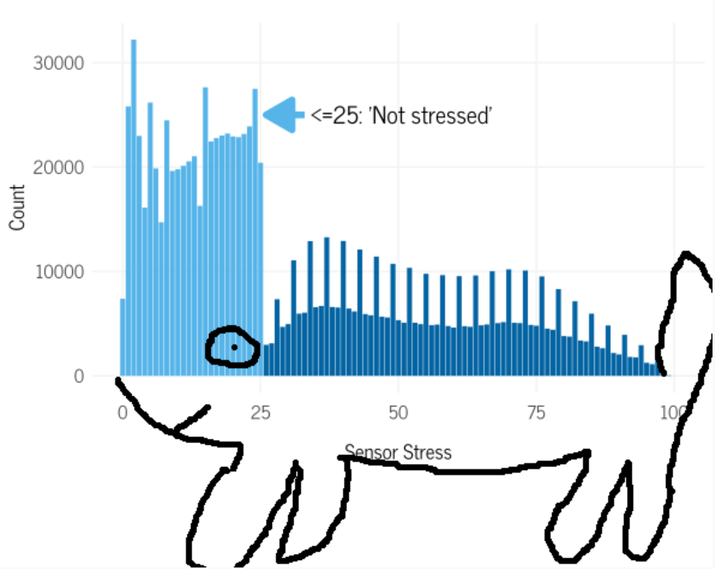
Sensitivity Analysis



Raw Stress Distribution



Stress Dinosaur



Key Takeaways

Findings:

- **Stress:** Weak to no associations

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Implications:

- Raw data could improve connection

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Implications:

- Raw data could improve connection
- Sensors could still hold predictive value

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Implications:

- Raw data could improve connection
- Sensors could still hold predictive value
- **Semantic gap** between sensors and self-report?

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Implications:

- Raw data could improve connection
- Sensors could still hold predictive value
- **Semantic gap** between sensors and self-report?

Key Takeaways

Findings:

- **Stress:** Weak to no associations
- **Sleep & Tiredness:** Somewhat stronger associations
- Overall R^2 and RMSE relatively weak

Implications:

- Raw data could improve connection
- Sensors could still hold predictive value
- **Semantic gap** between sensors and self-report?

Image credit: Advnture/Future

Thank you

WARN-D TEAM

Get In Touch

- bjoern.siepe@uni-marburg.de
- BlueSky: bsiepe
- https://bsiepe.github.io/

Paper & Slides

References i

Image Source

Garmin Watch:
https://www.advnture.com/news/
if-your-garmin-watch-is-giving-you-strange-stress-warning-dont-ignore-them