IMPACT Mental Health

Digital Phenotyping

IMPACT-MH All Investigators Meeting

April 29, 2026

IMPACT Mental Health

Digital Phenotyping

Project Snapshots:

REACH

ARTEMIS

PPSN

COMPASS

PREDiCTOR

TRACC-MH

HALO

Working comparison what we're collecting,

What we're learning vs. *where we're stuck*

Highlight some bigger picture questions projects are facing in digital phenotyping space

IMPACT
Mental Health

Digital Phenotyping

- Consortium spans radically different populations, clinical targets, and methods
- Shares a common goal to build harmonizable, clinically translatable behavioral data
- Digital phenotyping :
  - **Dense, longitudinal, real-world behavioral signal**
  - Can be feasibly collected at scale
  - Provides **rich, actionable information** to inform assessment and care
  - Can define **new, mechanistic subtypes** that don’t emerge from cross-sectional data

REACH Study: Data Modalities Collected

Audio — Parent Interview (NLP)

Parents respond to 3 open-ended questions about their child:

1. What do you enjoy about your child?
2. Do you have any concerns about your child's behavior?
3. Has anyone in your life expressed concerns about your child's behavior?

NLP analyses will capture:

- **Sentiment & affect tone**
  Positive/negative/neutral emotional valence across narrative
- **Behavioral concern flags**
  Detection of language signaling developmental worries
- **Third-party concern mentions**
  Whether family, teachers, or clinicians raised concerns
- **Parental warmth markers**
  Language reflecting positive engagement and bonding
- **Linguistic complexity**
  Vocabulary richness, sentence structure, and coherence

Video — Mother-Child Drawing Task

A recorded interaction in which the child:

- Draws three shapes
- Writes their name on paper
- Interacts naturally with mom throughout

Dual capture approach:

- **Video recording**
  Dyadic interaction quality, turn-taking, responsiveness, affect
- **Static image capture**
  Drawing product scored for fine motor + developmental markers
- **Behavioral coding (planned)**
  Coding schemes for joint attention, maternal sensitivity
- **Computer vision (planned)**
  Automated analysis of drawing complexity and form

Sampling, Duration & Real-World Compliance

101
Median EMA continuous duration

75.6%
Median EMA completion rate

64–86%
IQR for EMA completion

>90%
Typical Ōura wear time

EMA Protocol

7-question EMA push notifications delivered daily throughout pregnancy and postpartum.

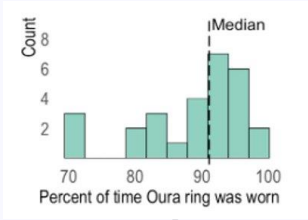
Pilot: n = 30 pregnant women; median 101 continuous days (IQR: 67.5 – 143.5 days).

Median completion: 75.61% of all pushes (IQR: 64.76 – 85.50%) — strong adherence in a real-world remote sample.

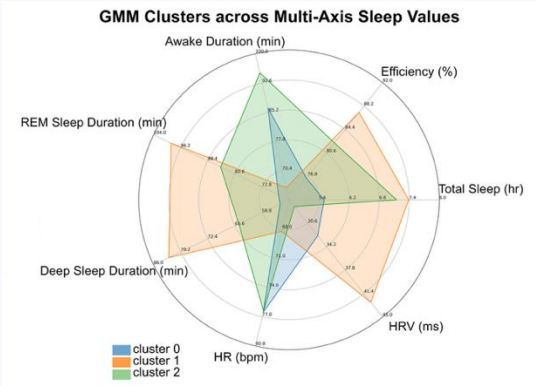
Ōura Ring Wearable

Continuous passive physiology: sleep stages, HRV, respiration, heart rate, activity.

Ōura Ring Wear-Time Distributio



The median for wearing Ōura was >90% of study days (median shown by dashed line)



What Worked · What Didn't · What the Data Show

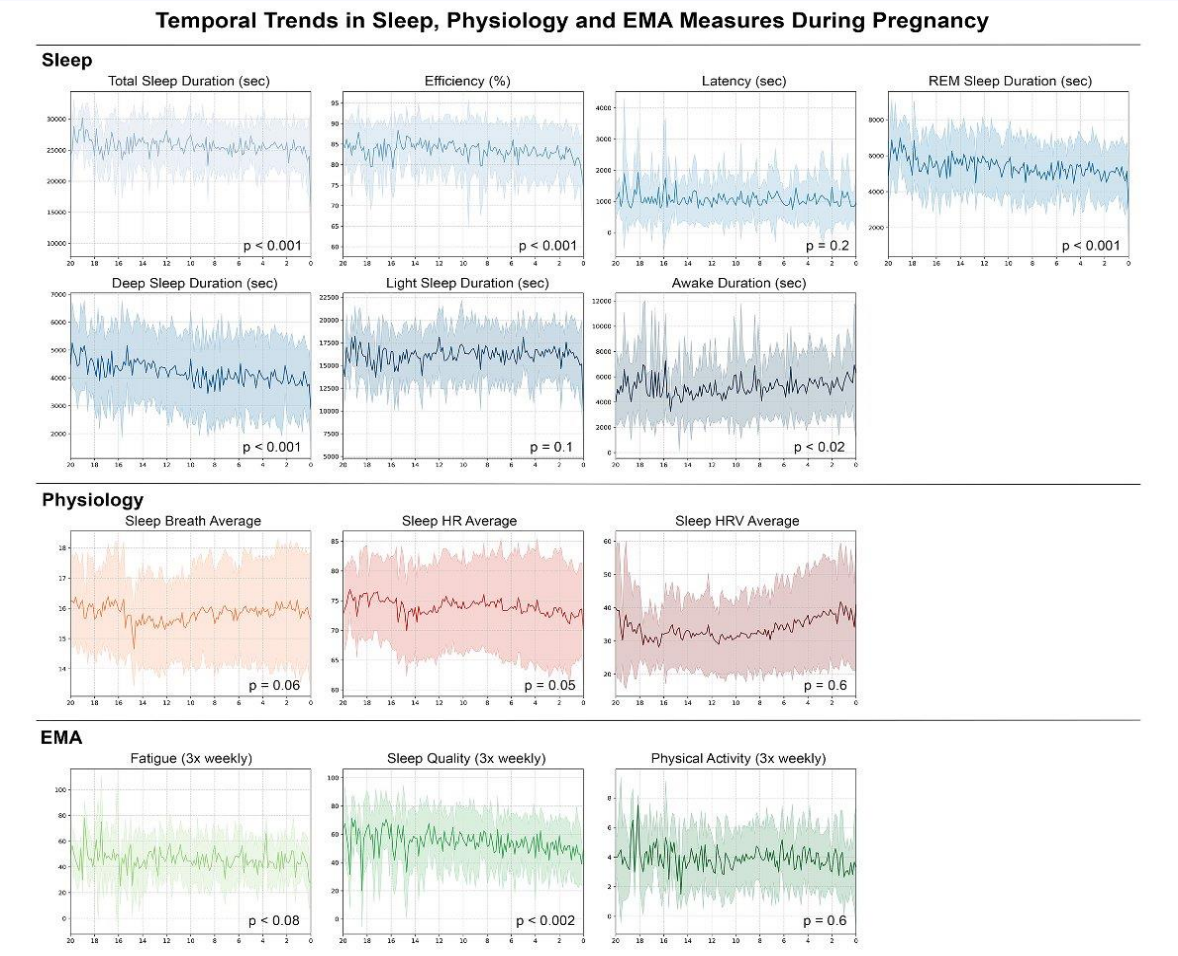
✅ What Worked

High passive compliance: Ōura adherence >90% for most participants with no active effort required - wearable data collection is feasible at scale.
EMA sustained engagement: 75%+ push completion over a median of 101 days demonstrates that brief daily check-ins are tolerated during pregnancy.
Meaningful signal detected: Longitudinal trajectories show significant changes in sleep efficiency, HRV, and self-reported sleep quality as delivery approaches.

⚠️ Challenges / Didn't Work

First-trimester data gap: Limited enrollment in early pregnancy constrained analyses; first trimester remains underrepresented in coverage.
Later postpartum drop-off: Data density declines sharply beyond 3 months postpartum - ongoing engagement is harder to sustain after the newborn period.
Fatigue & activity not captured: Self-reported fatigue and physical activity showed no significant longitudinal trend - these constructs may need richer measurement.

Temporal Signal: Sleep, Physiology & EMA



Group-level trajectories by weeks before delivery. Key measures show significant longitudinal change ($p<0.001$ for sleep duration, efficiency, and REM).

ARTEMIS

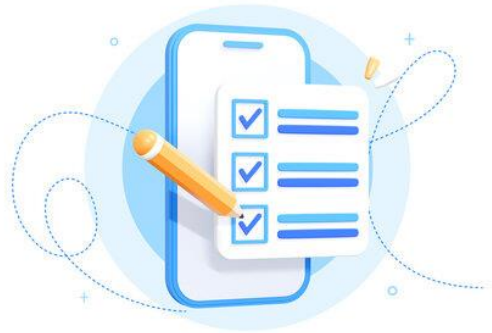
Analyses to Reveal Trajectories and Early Markers of Imminent Shifts in Suicidal States

PIs: Melanie Bozzay, Ph.D., Scott Langenecker, Ph.D., Jay Fournier, Ph.D., Ivy Tso, Ph.D., & Jessica Turner, Ph.D.

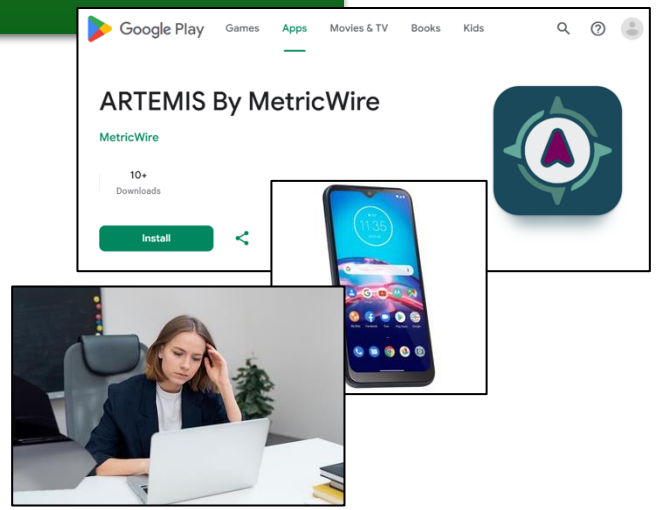
Study design

Screening Survey (5 min)

Baseline Survey (45-60 min)



Study Session (1-2 hours)



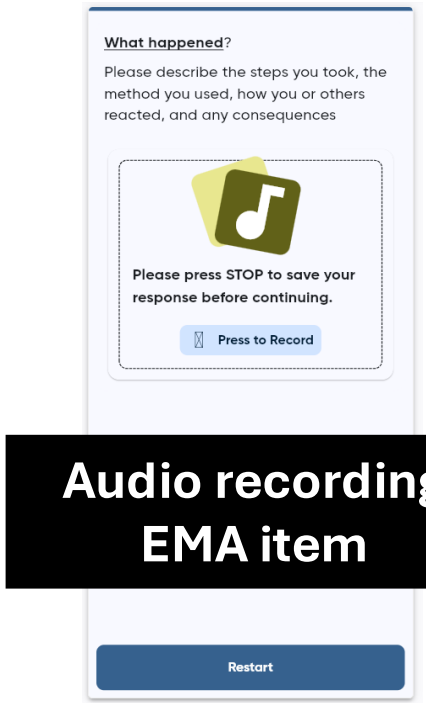
Longitudinal Battery

- Weekly Surveys & Tasks (6 weeks; 20 min – 1 hr.)
- Ecological Momentary Assessment (4 weeks; 5x/day)

ARTEMIS

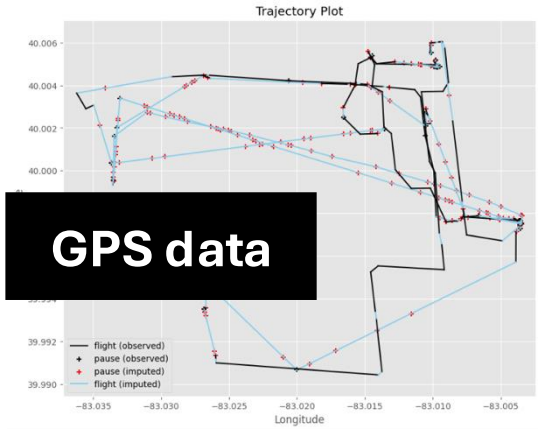
Ecological momentary assessment & digital phenotyping

- ARTEMIS mobile app published on app store
- Includes a personalized crisis response plan
- EMA assesses situational context, life stressors, emotional experiences, coping, & suicide risk
- Passive sensing assesses GPS, mobility, call/text message patterns, app use, ambient sound, light

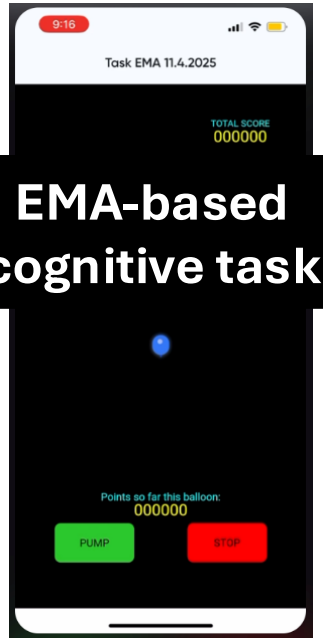


Audio recording EMA item

App use



GPS data



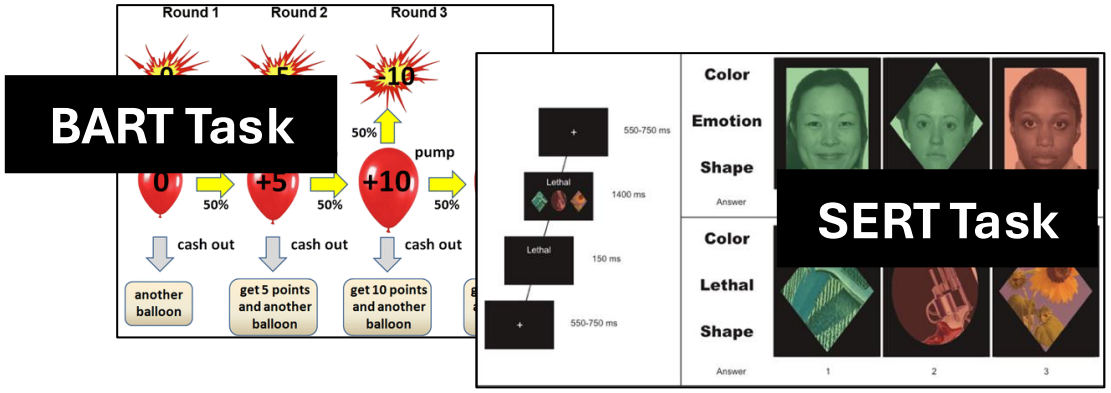
EMA-based cognitive task

ARTEMIS

Weekly measures & tasks

Cognitive Task Battery

- Different tasks each week, some tasks repeated
- Measuring cognitive domains and indices of motivational systems implicated in risk



Weekly Measures (mimic clinic)

- Suicide risk
- Physical and social functioning
- Sleep, pain
- Psychiatric symptoms (depression, anxiety)
- Treatment utilization & medication use
- Drug/alcohol timeline followback
- Risk/protective factors (hopelessness, coping, meaning in life)

Pediatric Precision Sleep Network (PPSN)

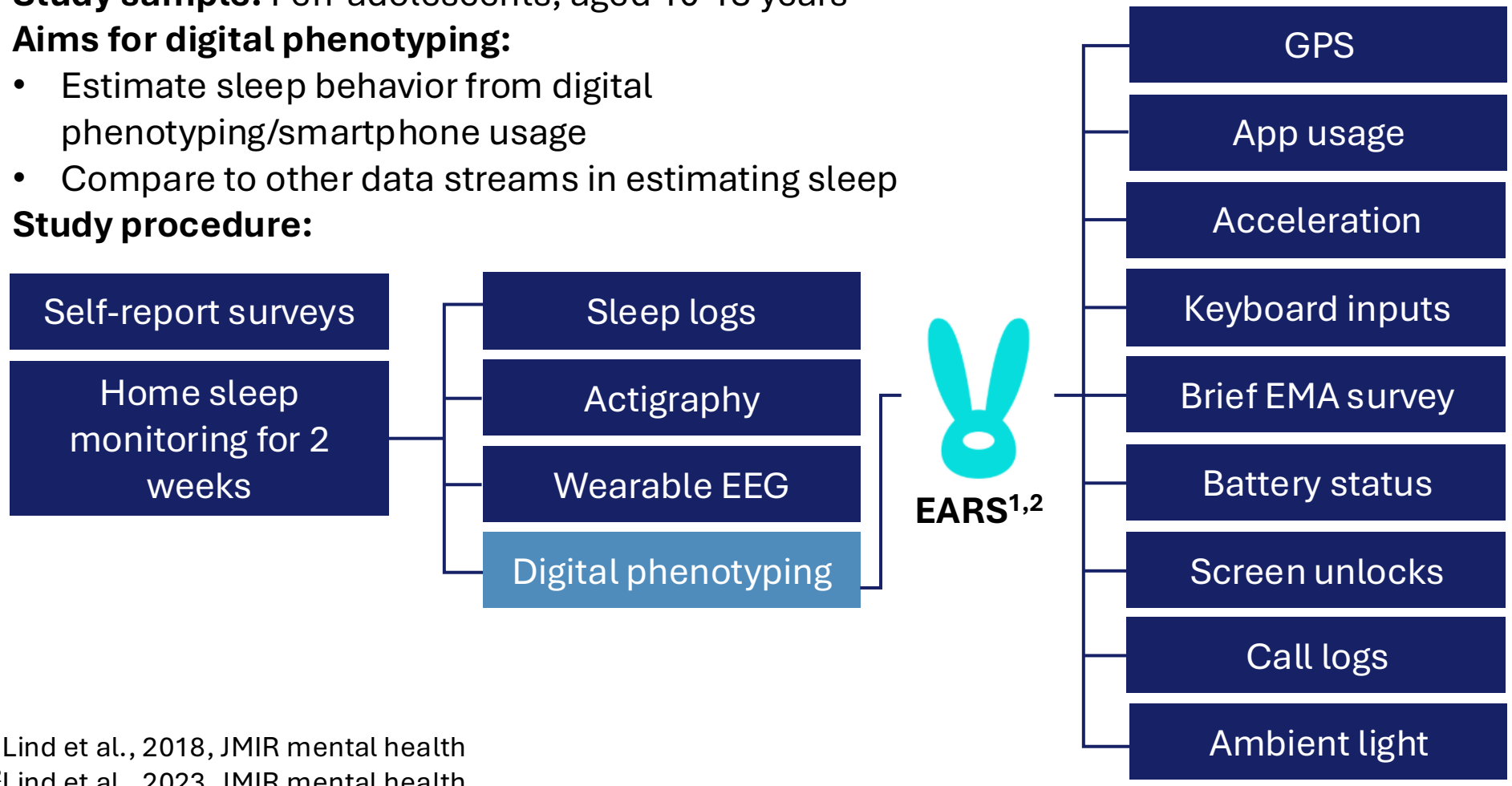
SleepLife
PEDIATRIC PRECISION SLEEP NETWORK

Study sample: Peri-adolescents, aged 10-13 years

Aims for digital phenotyping:

- Estimate sleep behavior from digital phenotyping/smartphone usage
- Compare to other data streams in estimating sleep

Study procedure:



[1]Lind et al., 2018, JMIR mental health
[2]Lind et al., 2023, JMIR mental health

Pediatric Precision Sleep Network (PPSN)

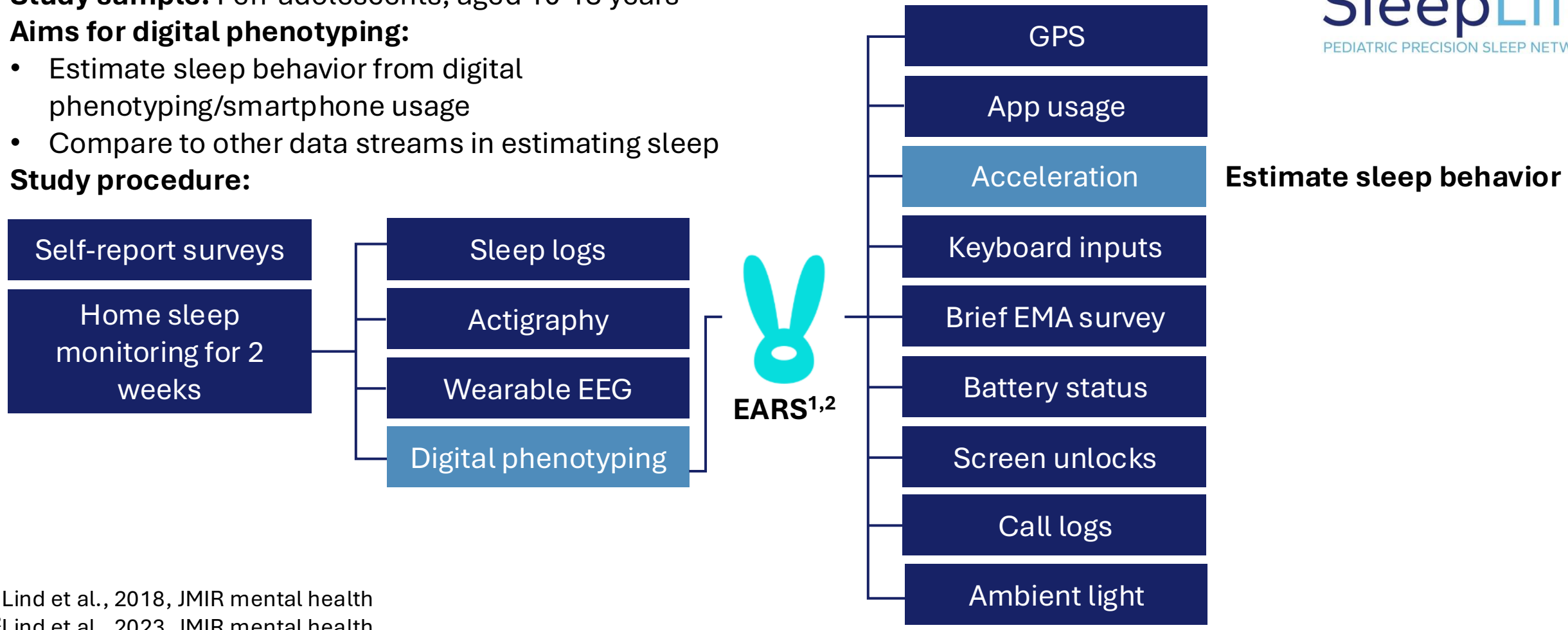
SleepLife
PEDIATRIC PRECISION SLEEP NETWORK

Study sample: Peri-adolescents, aged 10-13 years

Aims for digital phenotyping:

- Estimate sleep behavior from digital phenotyping/smartphone usage
- Compare to other data streams in estimating sleep

Study procedure:



[1]Lind et al., 2018, JMIR mental health
[2]Lind et al., 2023, JMIR mental health

Pediatric Precision Sleep Network (PPSN)

Data coverage:

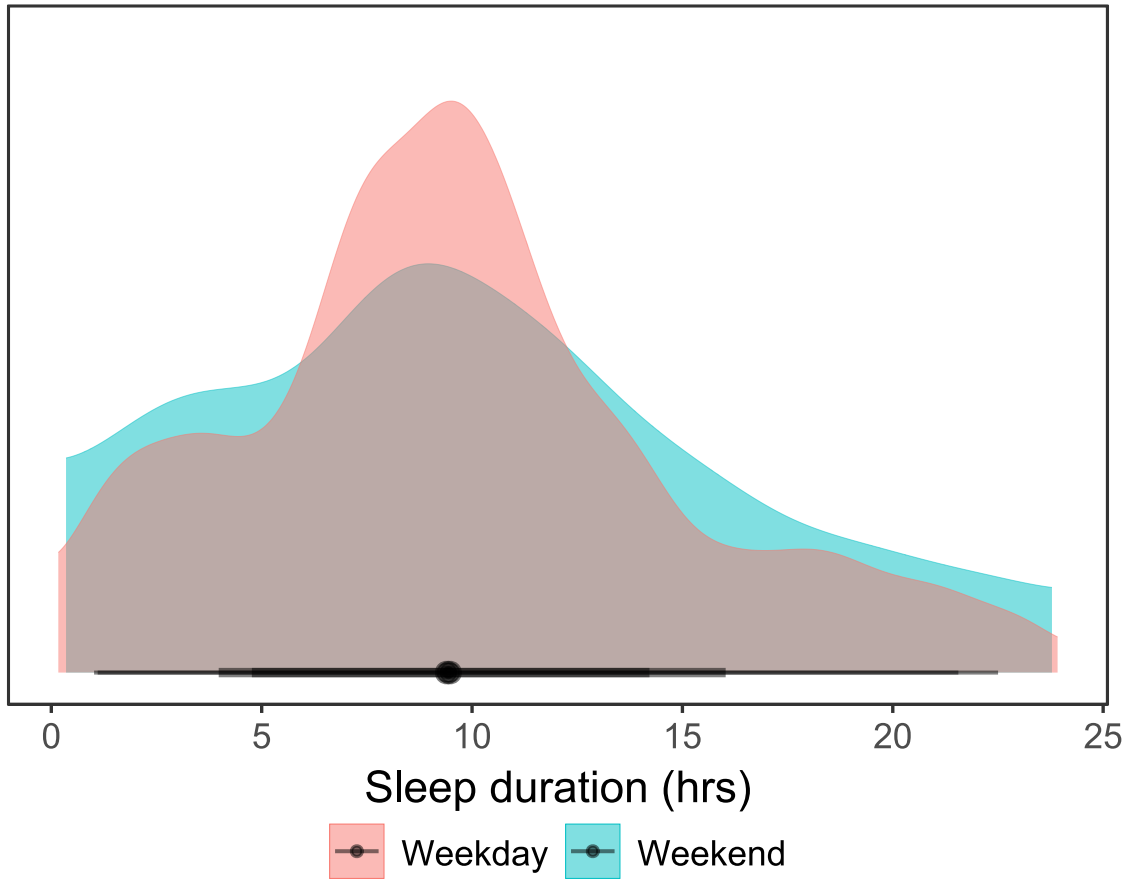
N completed Onboarding	N participating in EARS	N with at least 3 consecutive days of data
305	201 (65.9%)	168 (83.5%)

Sensor	N IDs	N with ≥3 consecutive days of data	Days of data (mean, SD)	Sampling rate
Acceleration	179	168 (93.9%)	12.1 (2.9)	50Hz (iOS); 10Hz (Android)
Brief EMA survey	167	113 (67.7%)	10.6 (3.7)	Once daily
Call logs	143	111 (77.6%)	9.9 (3.4)	Event-based
Battery state	147	134 (91.2%)	11.8 (3.0)	Event-based
GPS	173	166 (96%)	12.4 (2.8)	>100m of motion (iOS); every 15 mins (Android)
Screen unlocks	147	139 (94.6%)	12.4 (2.6)	Event-based
Keyboard inputs	142	107 (75.4%)	11.1 (3.6)	Event-based
Motion from OS	166	151 (90.9%)	12.5 (2.9)	Event-based
Light intensity	32	28 (87.5%)	12.2 (3.0)	Event-based
App usage	74	64 (86.4%)	13.6 (1.5)	Event-based

Pediatric Precision Sleep Network (PPSN)

SleepLife
PEDIATRIC PRECISION SLEEP NETWORK

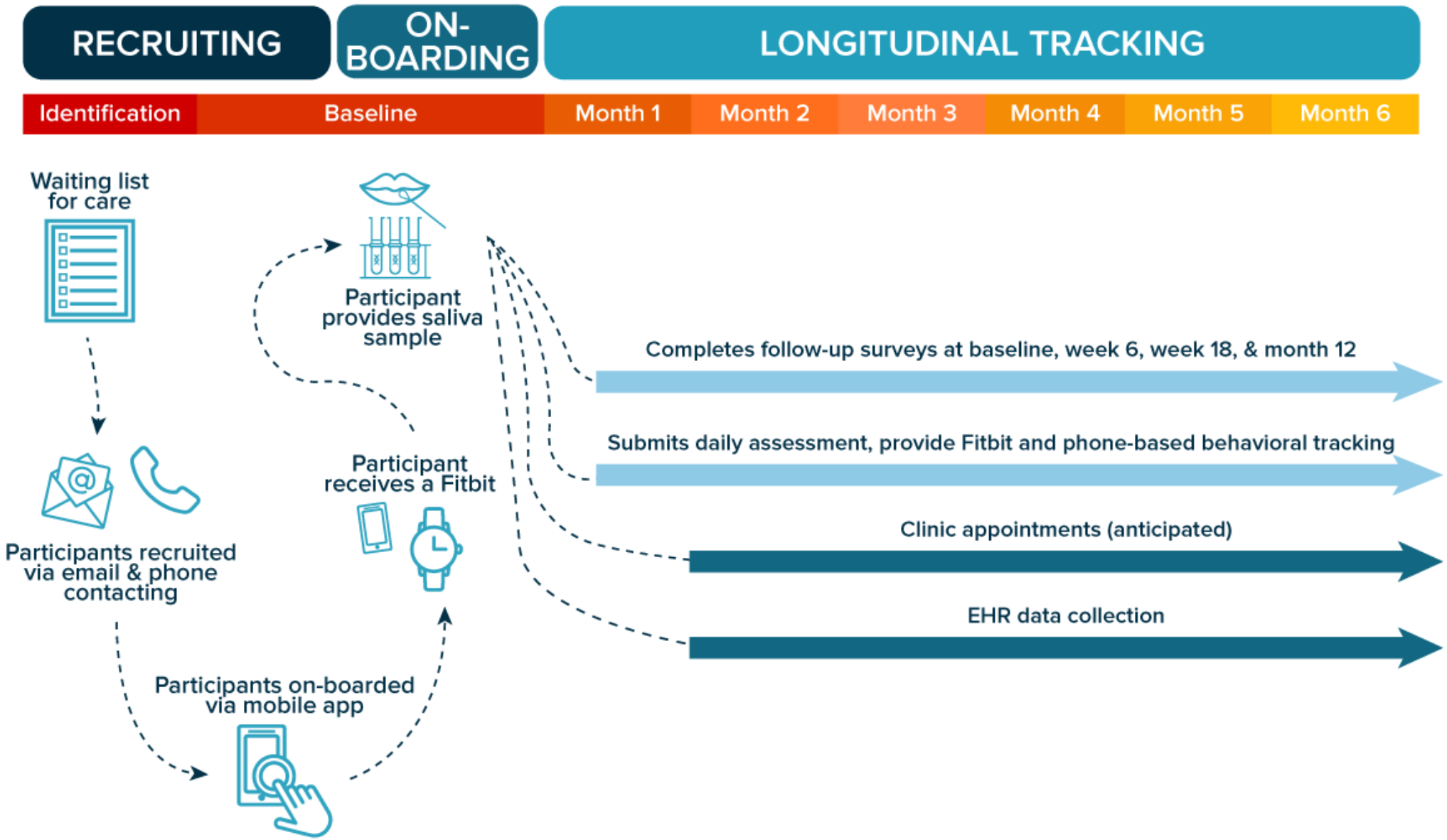
Success: Able to use accelerometer data to estimate sleep duration
Challenge: Adapting accelerometer data from phone into actigraphy processing pipeline (GGIR)



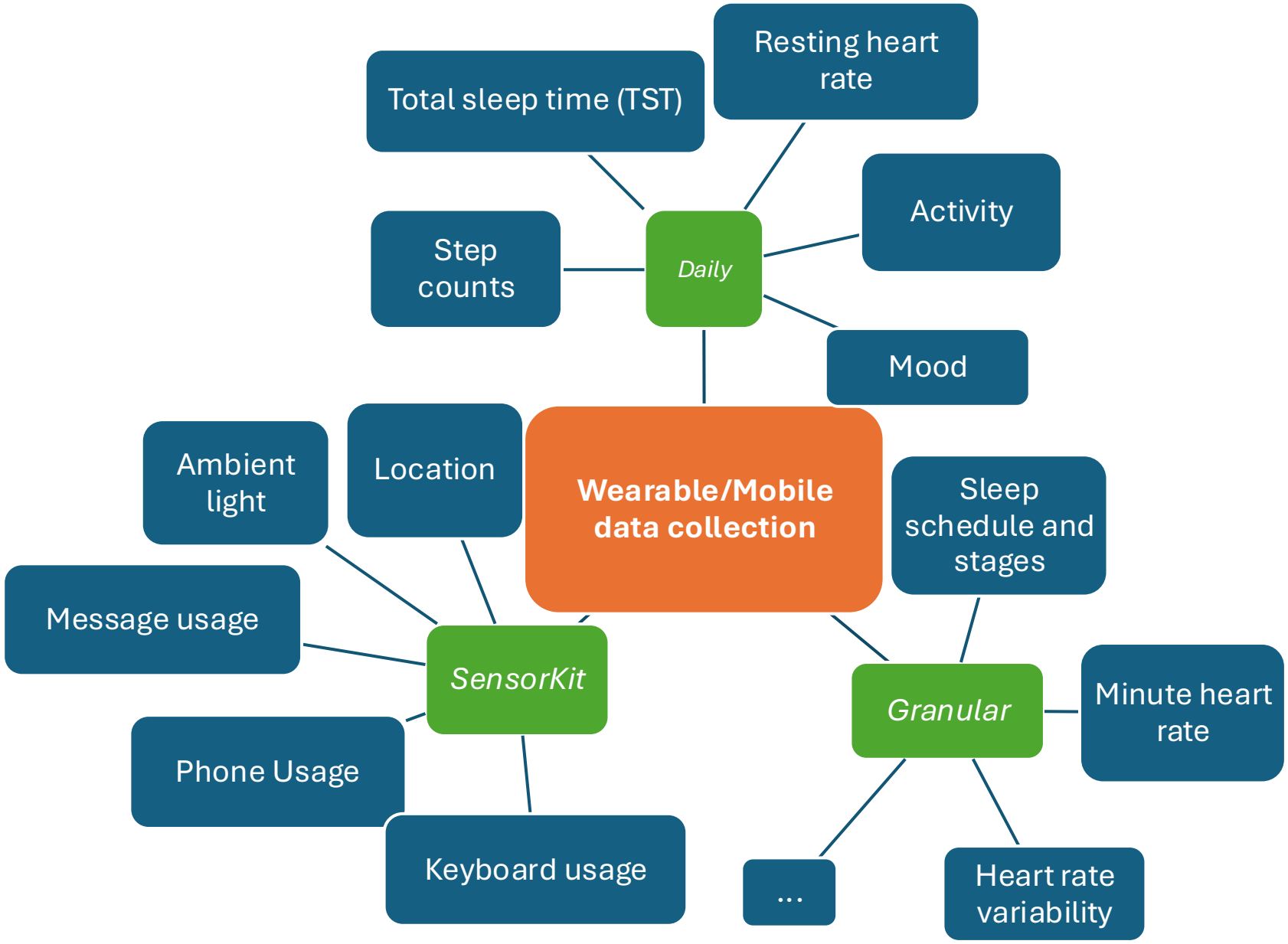
COMPASS – Digital Phenotyping

University of Michigan

COMPASS Study Patient Flow



COMPASS Mobile Data Collection

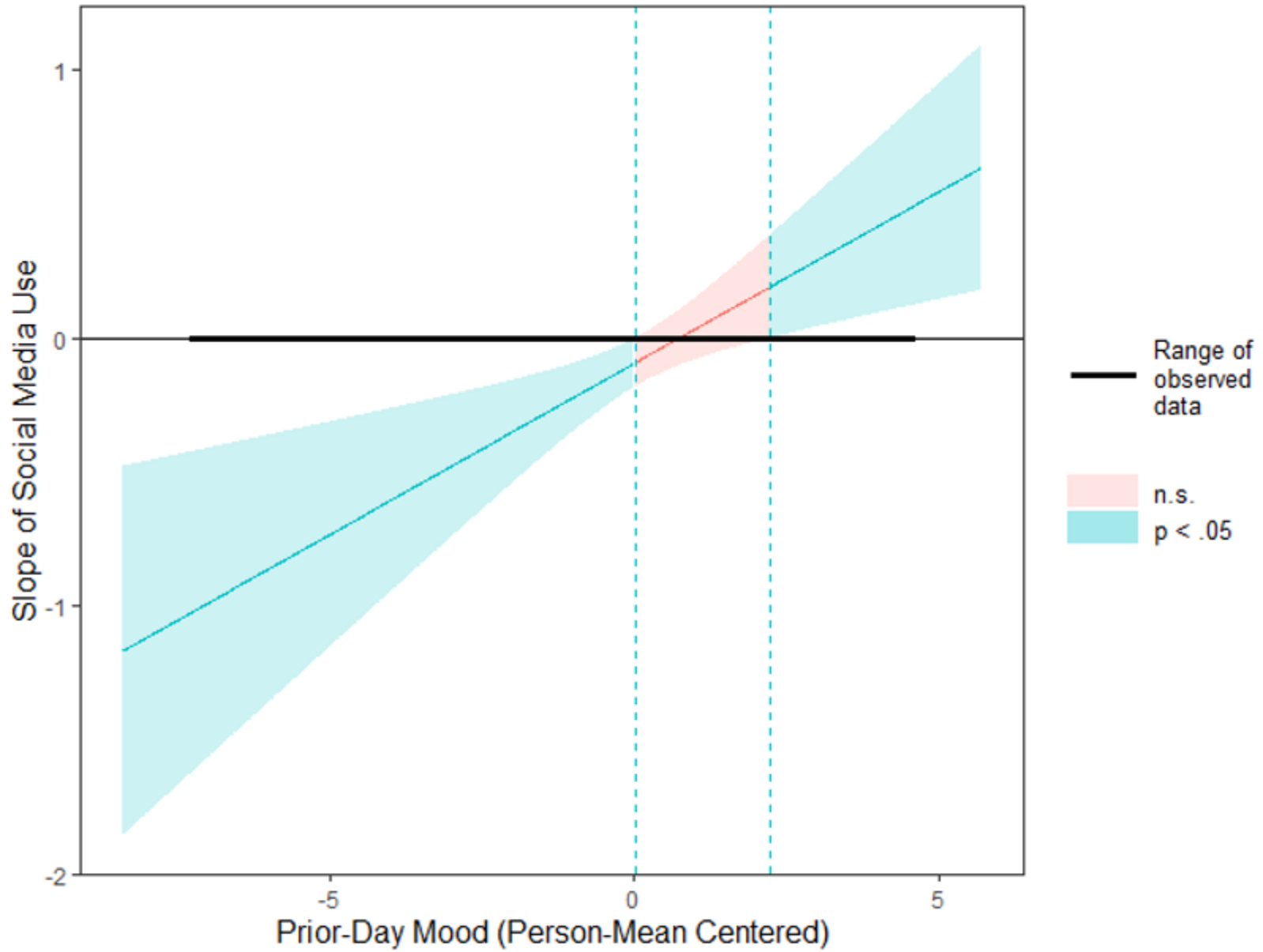


Sensorkit data - iPhones

- participant consents to study, with specific permissions for data sharing for each data type
- participant completes onboarding in MDH app, directs to “Research Sensor & Usage data” in settings
- MyDataHelps app/CareEvolution (CE) obtains sensorkit data
- CE transfers raw data to UM
- IT processing and subsetting to reduce data size

Data Completion (estimated)

- Fitbit data
  - 82% with any data
  - 58% with at least 3 days of usable activity data
- Sensorkit
  - 85% sharing some data
  - 71% sharing most sensitive (e.g., location)



eFigure 1. Johnson-Neyman plot of the interaction between social media use and centered prior-day mood as related to daily mood

HALO: Developing Data-Driven Clinical Signatures for People Who Experience Hallucinations

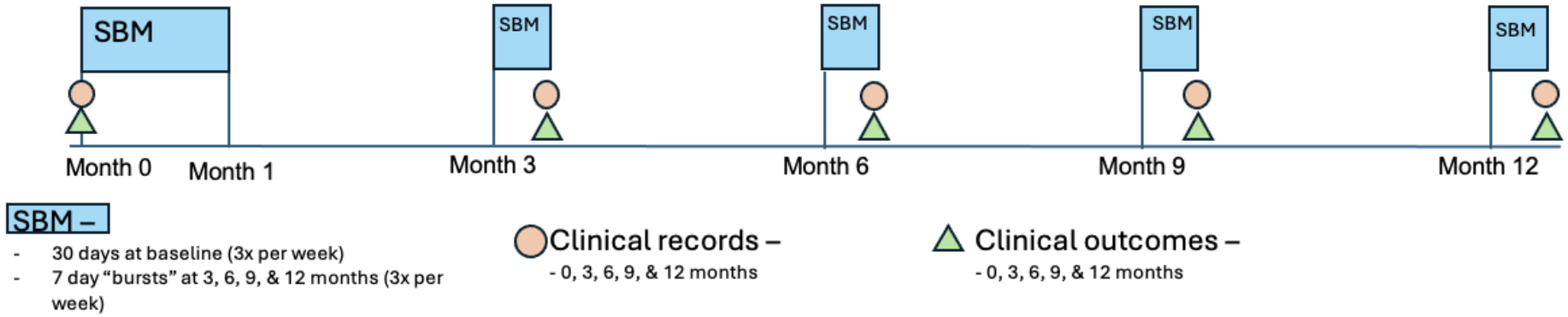
HALLUCINATION ASSESSMENT THROUGH LONGITUDINAL OBSERVATION

MPIs: Dror Ben-Zeev, PhD; Trevor Cohen, MD, PhD

HALO was funded by NIMH U01 MH135901-01

Modalities Collected + Sampling/Duration

Population: Adults experiencing hallucinations (auditory or visual); transdiagnostic, not diagnosis-gated
Target N: 2,000

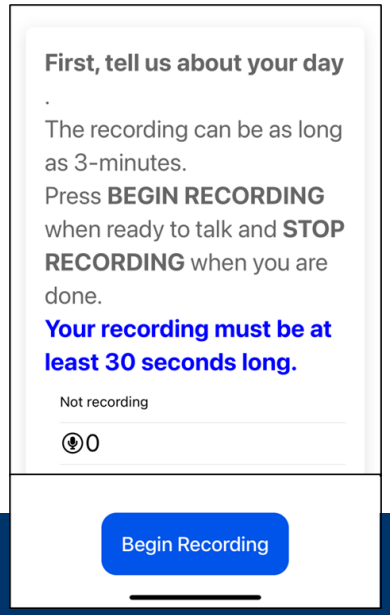


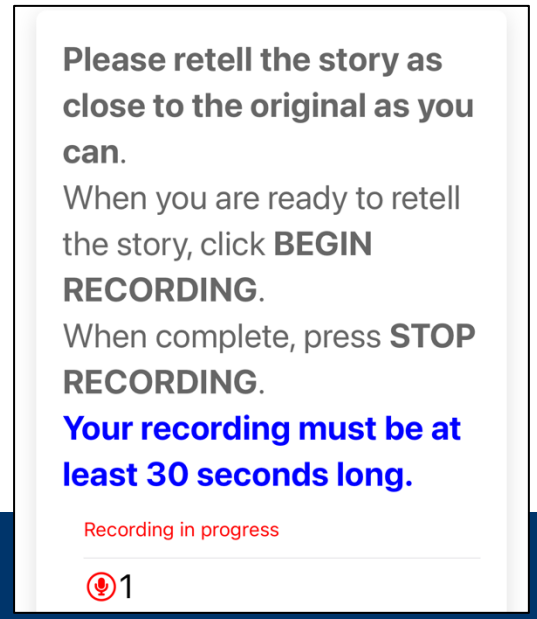
Smartphone Behavioral Measurement:

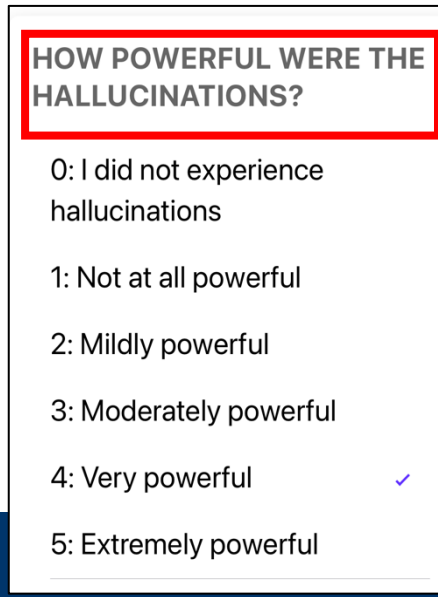
EMA (x3): Hallucination presence, negativity, distress, power

Verbal (story) Recall: automated scoring

Audio Diaries (x3): General day description, social interactions, hallucinations







Feature Pipeline: What we're extracting

Three tasks → 200+ features across established and exploratory NLP and speech families
Currently: 13,095 recordings from 342 participants

Feature Family	What it Captures	# Features	Status
EMA (hallucinations)	Presence, negativity, distress, power	~4	Integrated
Semantic coherence	Speech organization, topic consistency	~36	Integrated
Speech graphs	Word connectivity, repetition, cycling	8	Integrated
Perplexity	Language unexpectedness via LLM	5	Integrated
Cognitive distortions	Maladaptive thinking styles	5	Integrated
Emotion classification	Expressed emotion categories	28	Integrated
LIWC	Psycholinguistic word categories	117	Integrated
Self-harm / acuity	Risk-related language signals	13	Integrated
Verbal recall	Memory + speech coherence	partial	In development
Delusion / hallucination content	Content categorization	TBD	Exploratory
Pause duration	Acoustic fluency	TBD	Exploratory

Validation note: Each feature family is evaluated for validation needs before downstream modeling: define → technical quality → reliability → construct validity → bias assessment → modeling utility

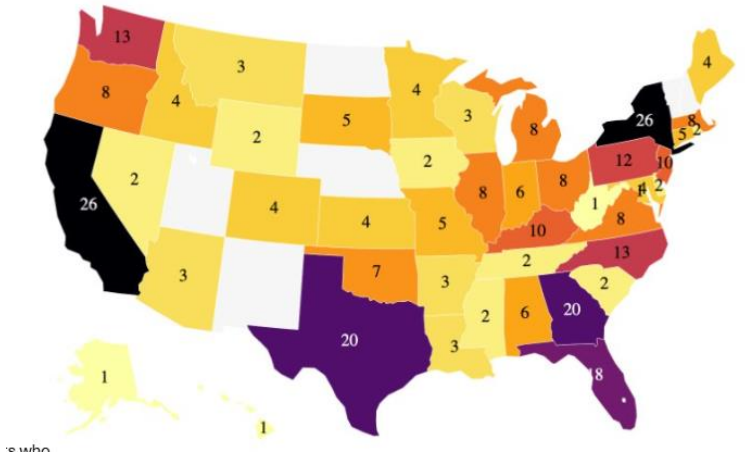
Compliance, What's working & what we're working through

Participant completion:

82.5% M1
73.4% M3
58.3% M6
52.9% M12

13,095 recordings from 342 participants
~145 hours of audio processed
~967,000 words transcribed, roughly even distribution across tasks

71% recordings come from hallucination positive check-in



Pipeline quality:

97.9% feature completion on submitted recordings

47% of recordings fall in the good-to-excellent AMOS range (3–5)

Primary missingness: **ASR transcription failures** in small subset of participants w/poor audio quality

What we're working through:

Longitudinal attrition & data quality: 21 participants through 1 year, monitoring as cohort moves through final phase

Audio quality variability: AMOS mean 2.9/5 with meaningful lower tail; may constrain feature reliability

ASR Bias: differential transcription performance across dialects and speech patterns; bias evaluation and mitigation ongoing

Feature validation: defining timing & "good enough" threshold before modeling; some features less reliable in specific contexts (e.g., short transcripts)

TRACC-MH: Trajectories of Risk and Cognitive Change in Mental Health

MPIs: Laura Germine, PhD; Kerry Ressler, MD, PhD; Agustin Yip, MD, PhD

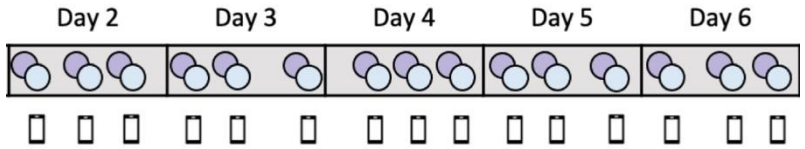
TRACC-MH was funded by NIMH UF1 MH125991

Modalities Collected + Sampling/Duration

Oura ring: activity, sleep, resting HR, HRV

MetricWire w/ SensorKit: EMA survey, speech samples (open ended prompt + weekly semantic fluency test), SensorKit (device usage, texts/calls frequency, keyboard metrics)

TestMyBrain Cognitive Measures: Digit Symbol Matching, Grid Memory, Choice RT, Gradual Onset CPT, Symbol Search, Self-Referential Encoding and Recognition Memory



Day 1: wearable

Day 2-6: 3/day

Months 1 – 3: 3/day for 7 days (x four bursts)

12 month hospitalization

INPATIENT (MCLEAN)

POST DISCHARGE

N=1500

N=250

COMMUNITY (MONTEFIORE HEALTH SYSTEM)

N=250

TRACC-MH was funded by NIMH UF1 MH125991

(Based on prior work w/similar cognitive EMA protocols)

Compliance?

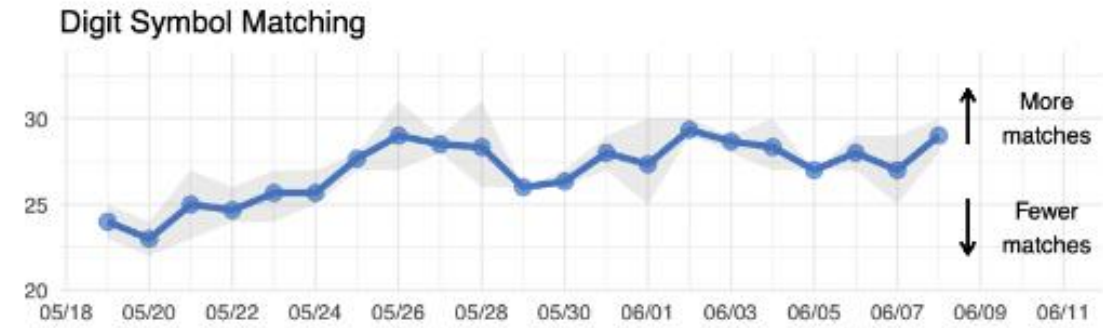
75-85%

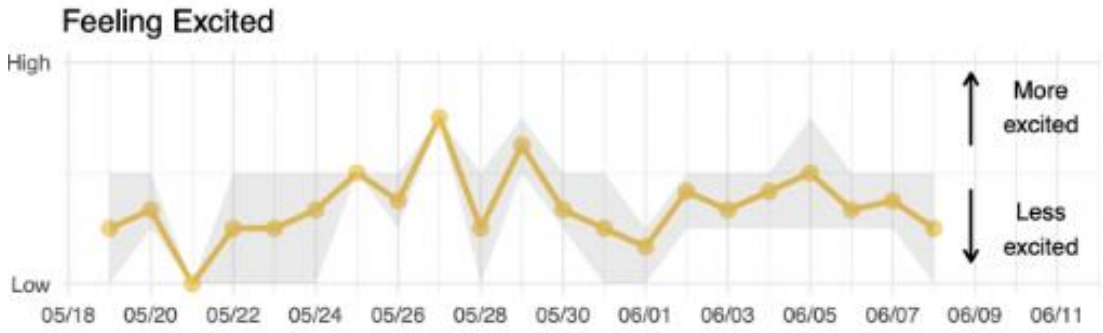
One thing that worked

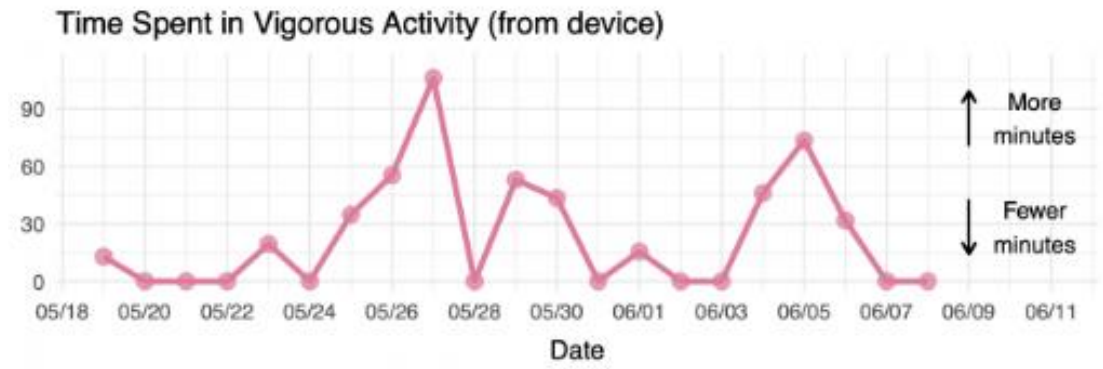
Return of individual-level results

One thing that didn't

No F2F contact (zoom is ok).





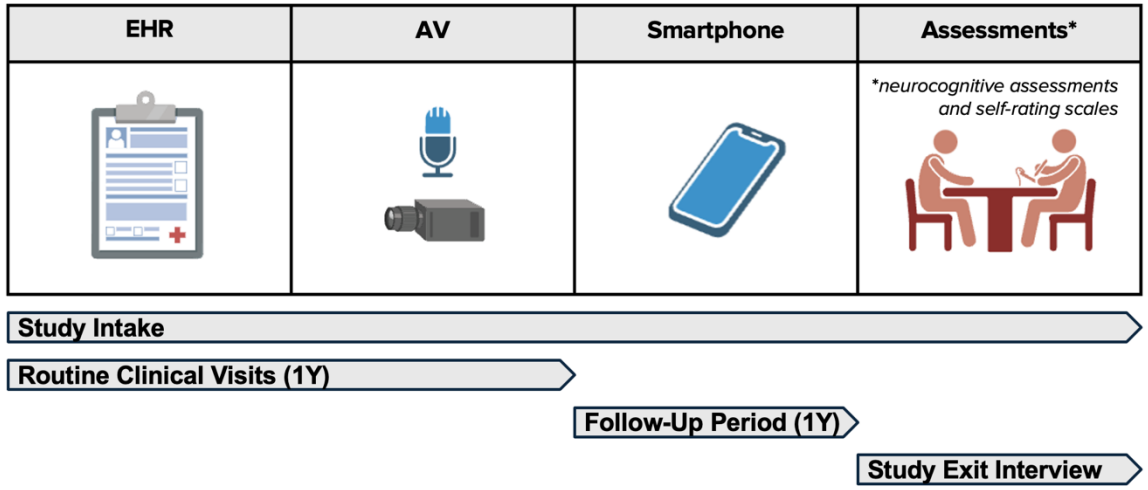


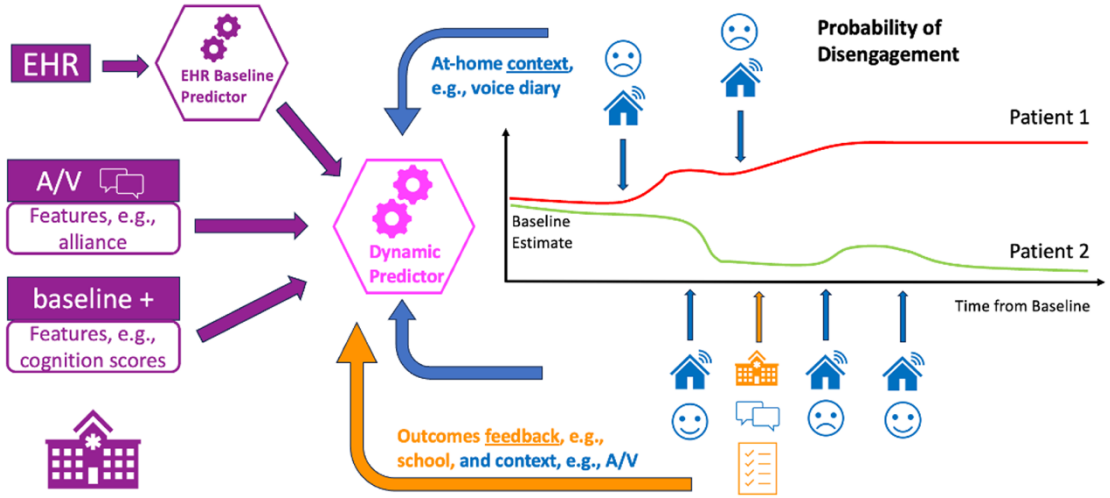
TRACC-MH was funded by NIMH UF1 MH125991

PREDICTOR

MPIs: Rene Kahn, MD; Cheryl Corcoran, MD, PhD; Guillermo Cecchi, PhD

-First-time psych patients aged 15-30 (n=2100)





Ecological momentary assessment & digital phenotyping

Passive sensing assesses GPS, screen state, accelerometer

Text journal

Audio journal

Video journal (coming soon!)

Compliance (as of April 2026)

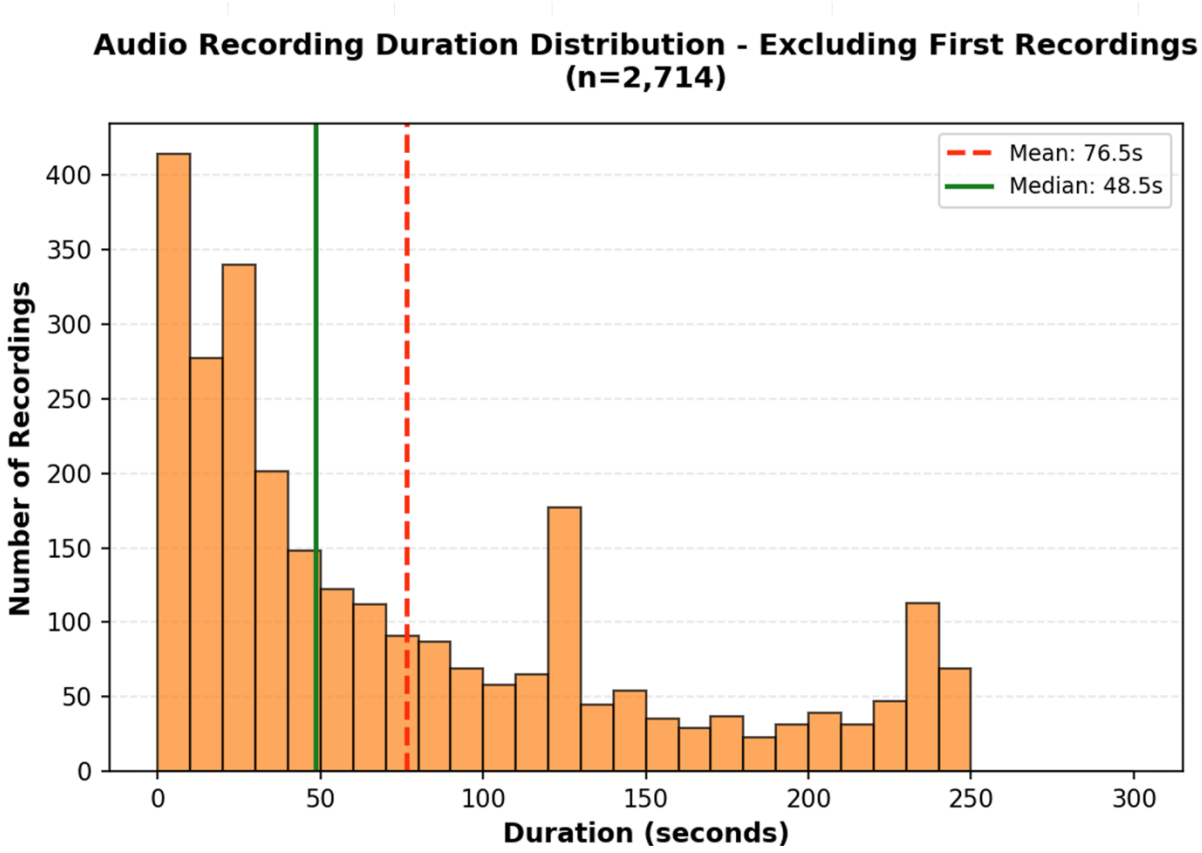
Summary:
- Active Participants: 393

Incoming Audio Journals:
- All Time: 3545
- Last 14 Days: 209
- Yesterday: 13

Incoming Text Journals:
- All Time: 984
- Last 14 Days: 100
- Yesterday: 5

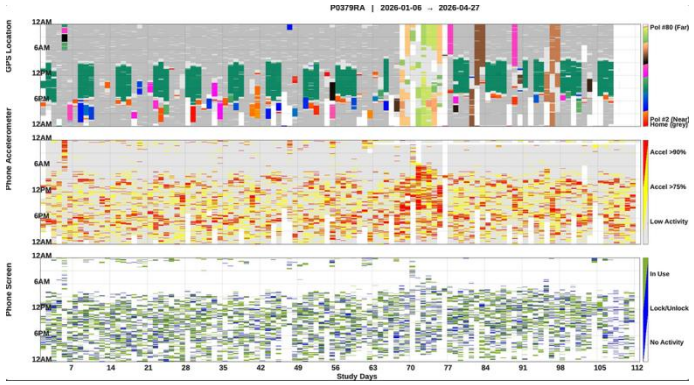
Sensor Data Gaps (>14 days):
- Total: 143
- GPS: 114
- No GPS: 27
- Unknown GPS Status: 2

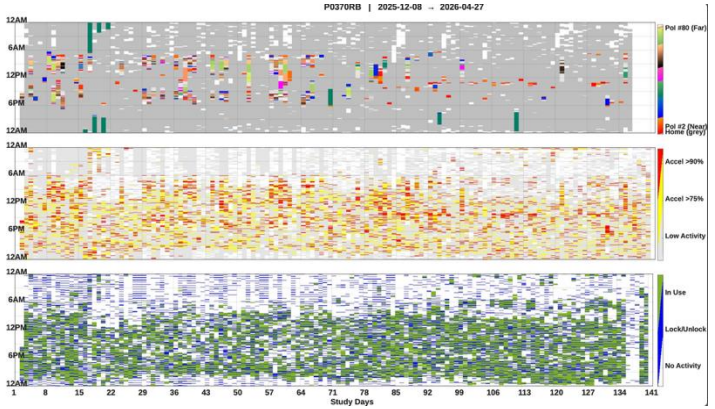
No MindLAMP Consent: 2

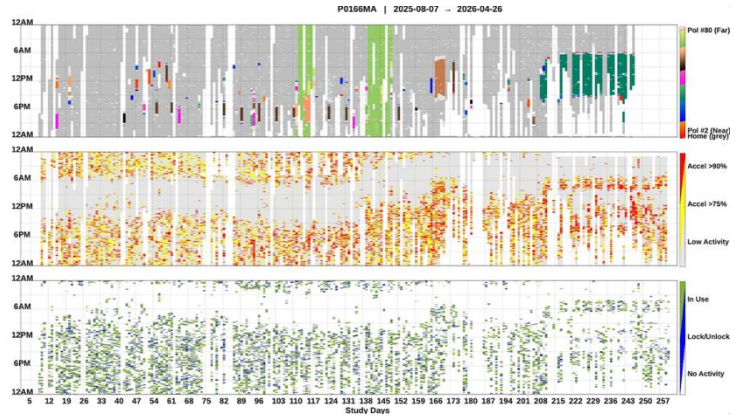


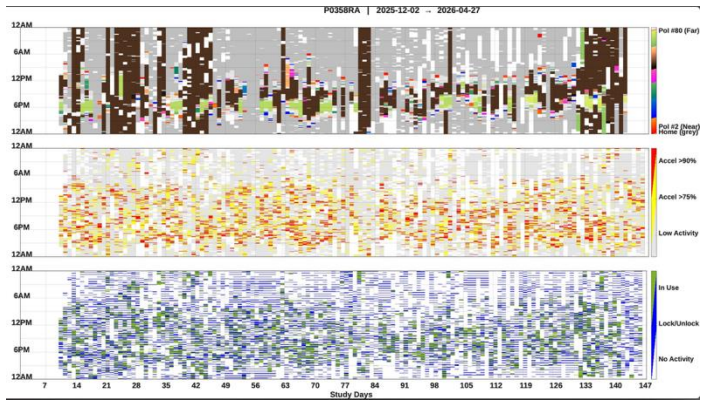
PREDICTOR

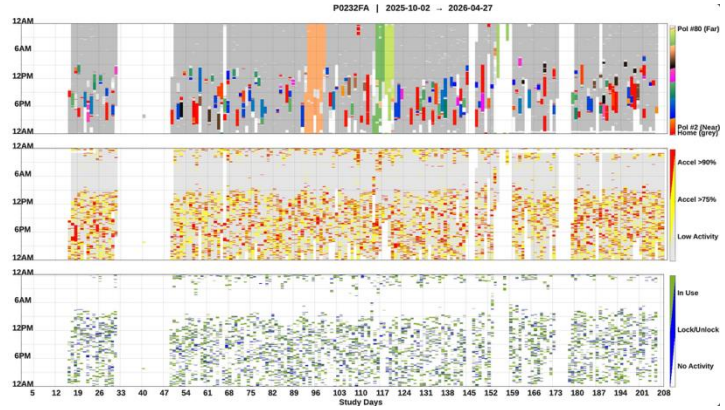
MPIs: Rene Kahn, MD; Cheryl Corcoran, MD, PhD; Guillermo Cecchi, PhD

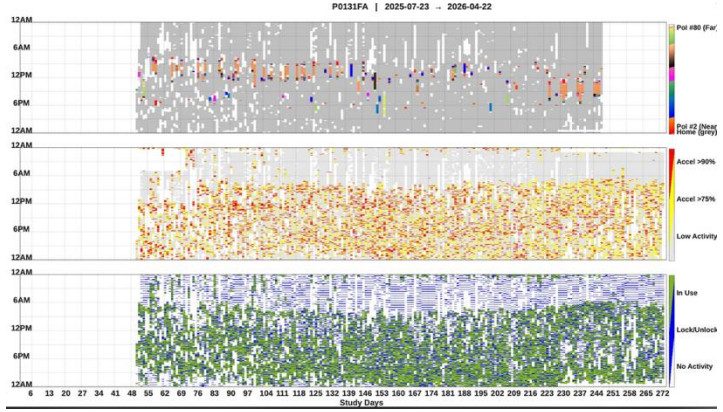












IMPACT Mental Health

Digital Phenotyping Project Matrix

Project	EMA	Passive Phone	Wearables	Audio	Cognitive Tasks
PREDiCTOR	✗	✓	✗	✓✓	✓ (non-mobile)
TRACC-MH	✓	✓	✓✓ (Oura)	✗	✓✓ (mobile)
HALO	✓	✓	No	✓	✓
ARTEMIS	✓	✓		✗	✓
COMPASS	✓	✓	✓	✗	✗
REACH	✓ (structured)	✗	✗	✗	✗
Duke-PMA	✓ (tablet)	✗	✗	✗	✓
PPSN	(possible)	✗	✓	✗	✗
ACE-D	✗	✗	✗	✗	✓✓

IMPACT Mental Health

Digital Phenotyping

Discussion:

Topic 1: Signal reliability and validity

What does signal reliability mean within the context of your work? How are you establishing reliability & validity?

For projects not yet collecting data: What signals are you most/least confident in?

IMPACT Mental Health

Digital Phenotyping

Discussion:

<u>Topic 2: Representation in data/sample</u>

Who is systematically not in your data/who are you noticing might be trending towards not represented?

For projects not yet collecting data: How are you thinking about reaching hard those that are hard to reach?

IMPACT Mental Health

Digital Phenotyping

Discussion:

Topic 3: Data quality, missingness, and current struggles

What is your biggest current data quality challenge? (e.g., app performance, audio quality, non-response, data harmonization)

For projects not yet collecting data: What data quality or missingness problem are you most concerned about? How are you designing to mitigate against it?

IMPACT Mental Health

Digital Phenotyping

Discussion:

<u>Topic 4: Participant experience, burden, and attrition</u>

What data are you finding is most worthwhile to be collecting (bang for the buck)?

What would you drop from your current procedures?

What would you add?

For projects not yet collecting data: how are you designing the participant experience to maximize retention across a diverse sample? What tradeoffs are you making between data richness and participant burden?

IMPACT Mental Health

Digital Phenotyping

Discussion:

<u>Topic 5: Cross-project harmonization</u>

With an eye toward minimum common digital phenotyping dataset, what could every project realistically contribute?

Mood EMA

Sleep timing

Activity data

Social interaction

Cognitive function

For projects not yet collecting data: what consultation, questions, or information would be helpful to inform where you are at now in setting up your data structures?