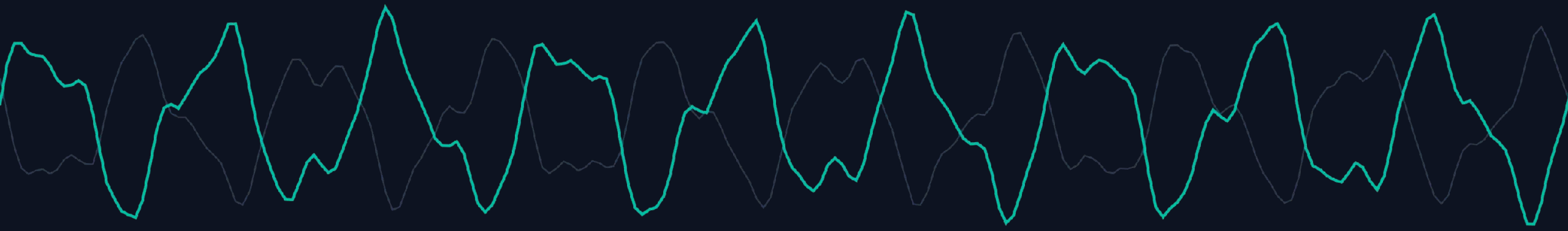
KOIOS CARE



Unlocking the Utility of Passive Digital Biomarkers of Patient Functioning

in Clinical Trials and Care Settings - a Parkinson's Disease case

Dimitris Iakovakis · Koios Care · BBS Digital Endpoints Workshop · 27 August 2026

A tremor score is not the ability to eat dinner

SYMPTOMS

What we rate in clinic.

MDS-UPDRS · supervised · quarterly

FUNCTIONING

What patients actually live.

eating · typing · walking · sleeping

Functioning = what you can do, in your environment, on an ordinary day.

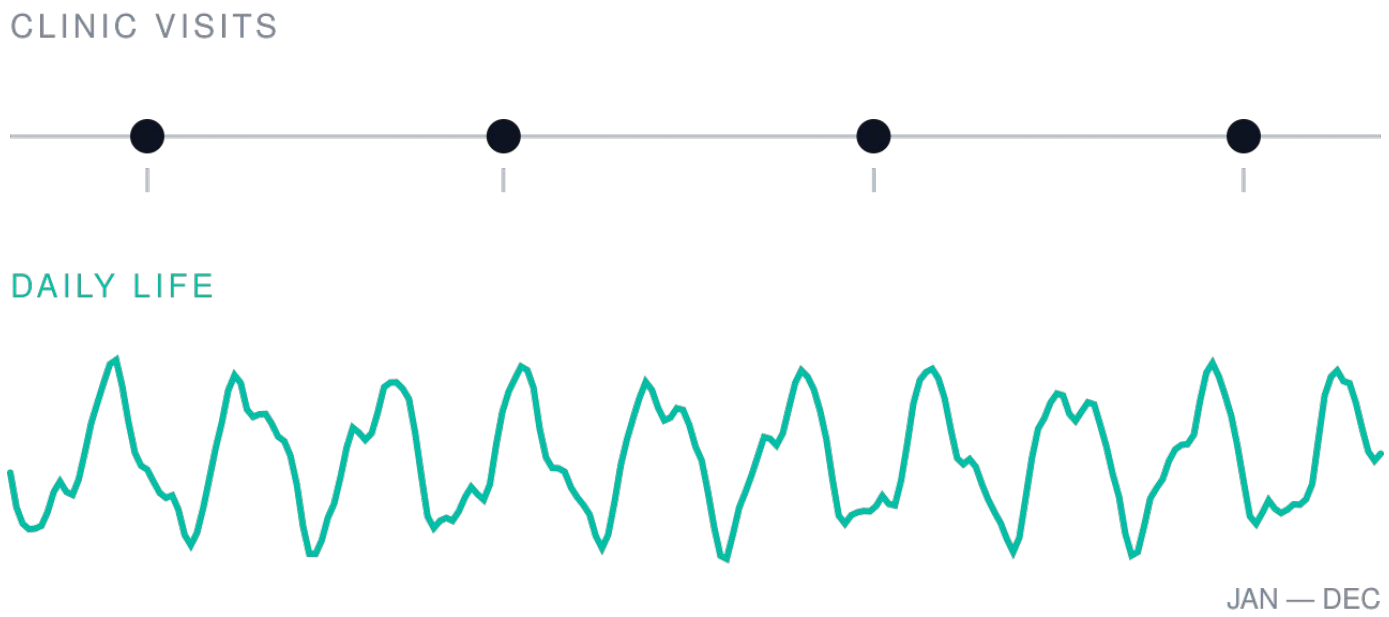
We assess 4 hours a year - the disease runs all 8,760

4

hours of assessment / year

8,760

hours of disease / year



Snapshots, best-day bias, recall - versus the actual signal.

Parkinson's is the proving ground for passive functioning measurement

MOTOR CORE

The sensor sees the disease directly

FLUCTUATING

Hours-scale ON / OFF, invisible to visits

TREATABLE

Every dose change is a natural experiment

INSTRUMENTED

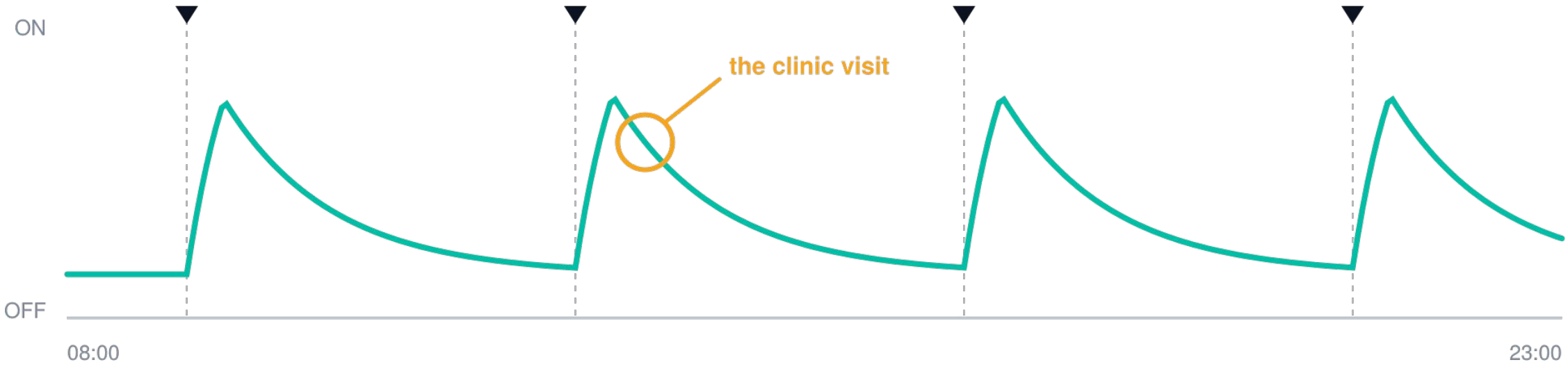
Validated scales to anchor against

BEHAVIORAL

It alters the things people do every day

PD is the beachhead, not the boundary - the methodology transfers to any condition that alters daily behavior.

The disease fluctuates faster than care can look



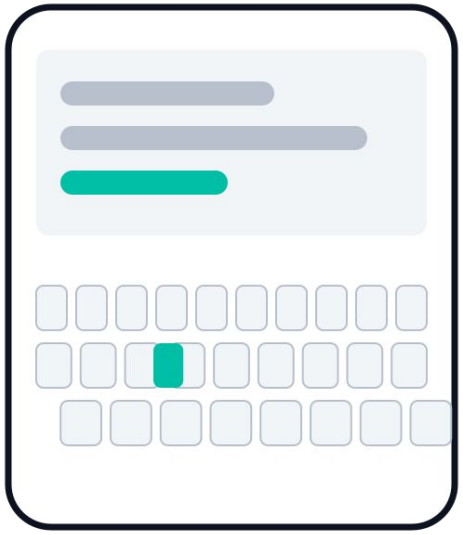
▲ levodopa dose ○ what the visit sees

One visit samples one point of a moving distribution - usually the best one.

The assessment is embedded in life - not added to it.

Every daily behavior is a motor test the patient self-administers.

Typing is a bradykinesia test hiding inside a message



KEYSTROKE TIMING · MILLISECONDS

- **Millisecond timing**

  Fine-motor kinetics at keystroke resolution

- **Hundreds of trials a day**

  Natural repetition no clinic protocol achieves

- **Zero instruction**

  No learning curve, no performance behavior

Eating is an upper-limb assessment hiding inside lunch

- **Goal-directed reach**

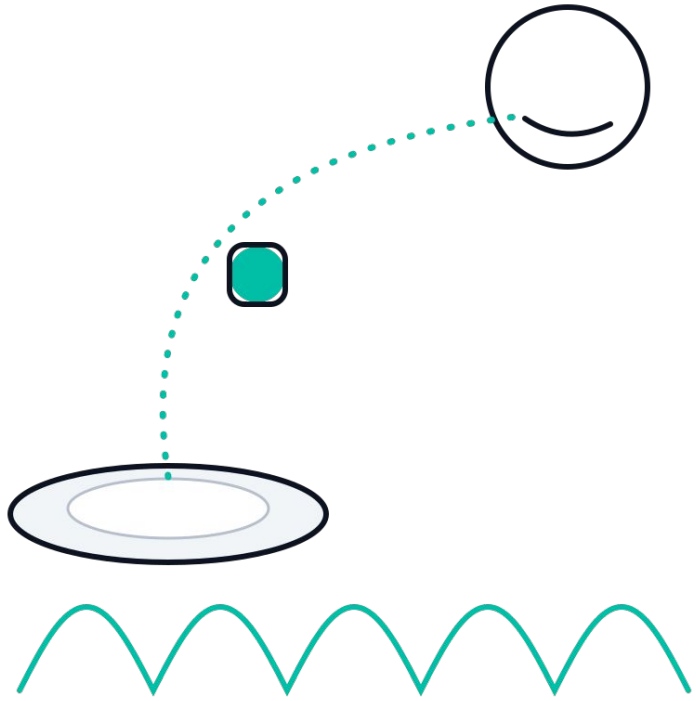
  Plate-to-mouth kinematics, cycle by cycle

- **A meal is a functional task**

  Repeated three times a day, every day

- **No extra hardware**

  Captured by the watch already on the wrist



WRIST KINEMATICS · CYCLE BY CYCLE

Passive wins where it matters: burden, bias, repetition

ACTIVE TESTS

Scheduled effort · learning & performance effects

1

supervised trial per quarter

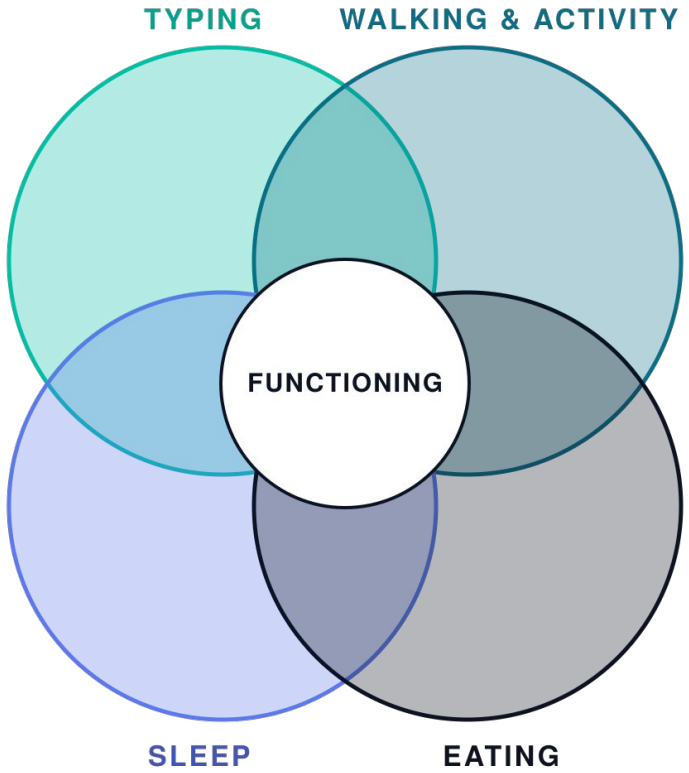
PASSIVE SENSING

No burden · no learning curve · life as usual

10,000+

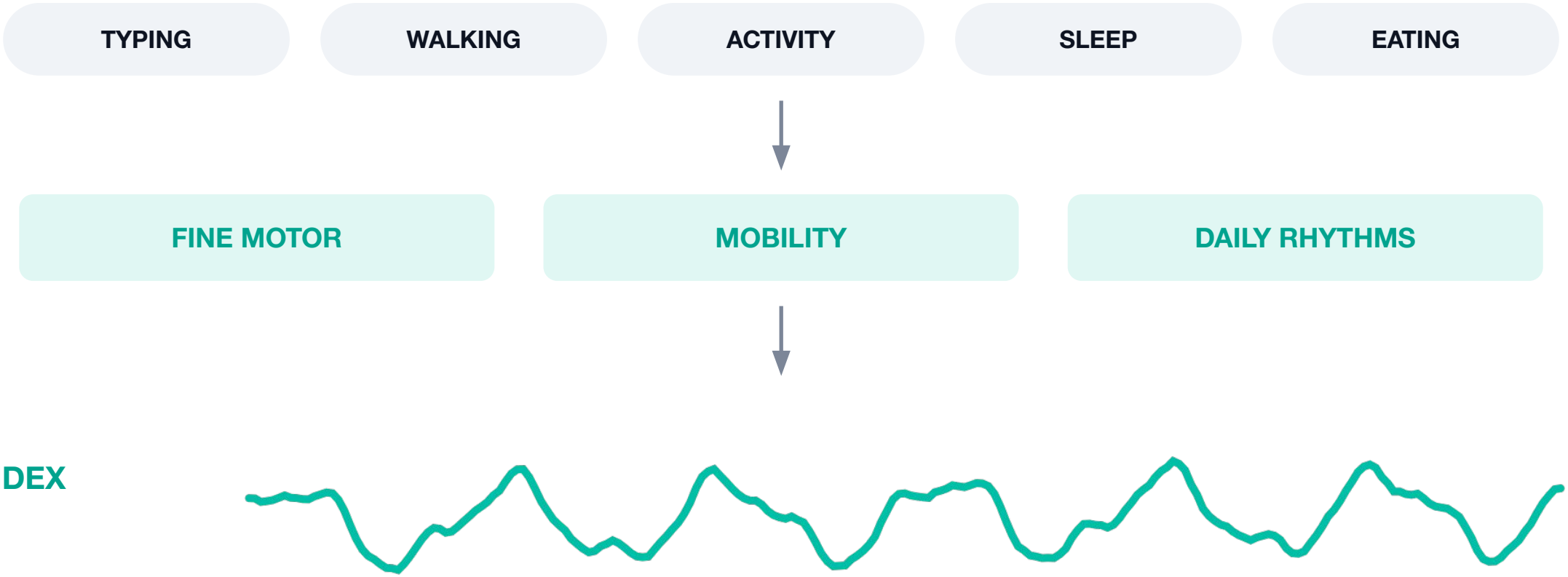
natural trials per quarter

No single behavior is functioning



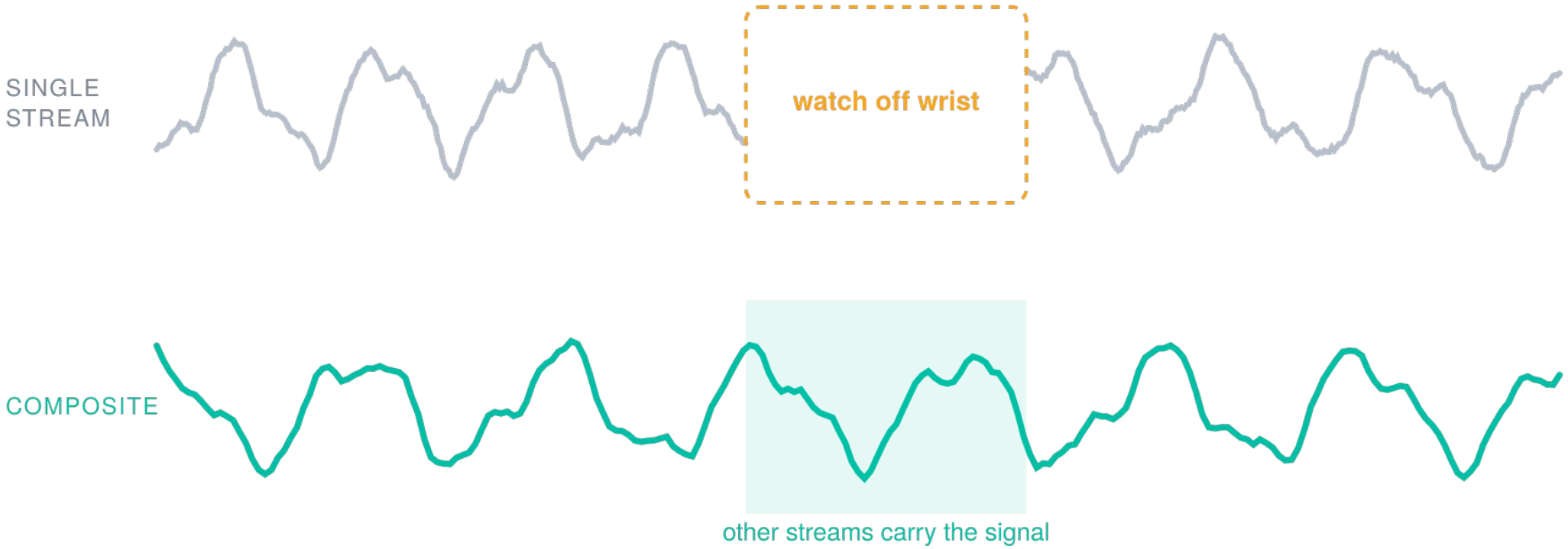
Each stream sees one facet - patients decline in one while stable in another.

The Q-index: many behaviors, one trajectory



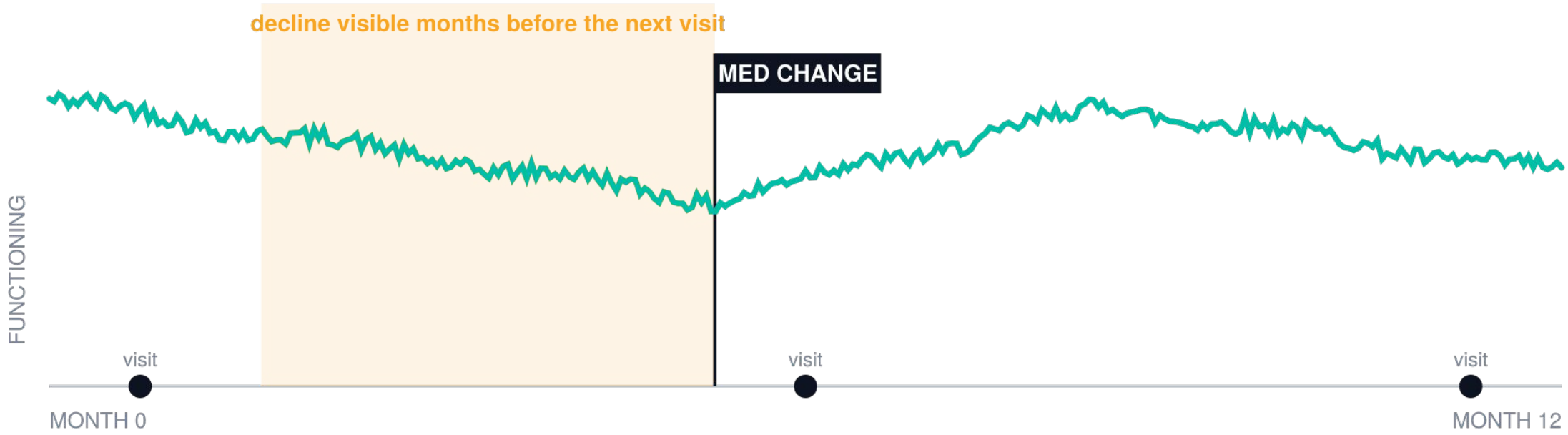
One number - always unfoldable back into the behaviors driving it.

A composite degrades gracefully - a single sensor goes silent



Real life has gaps. Coverage is a design property.

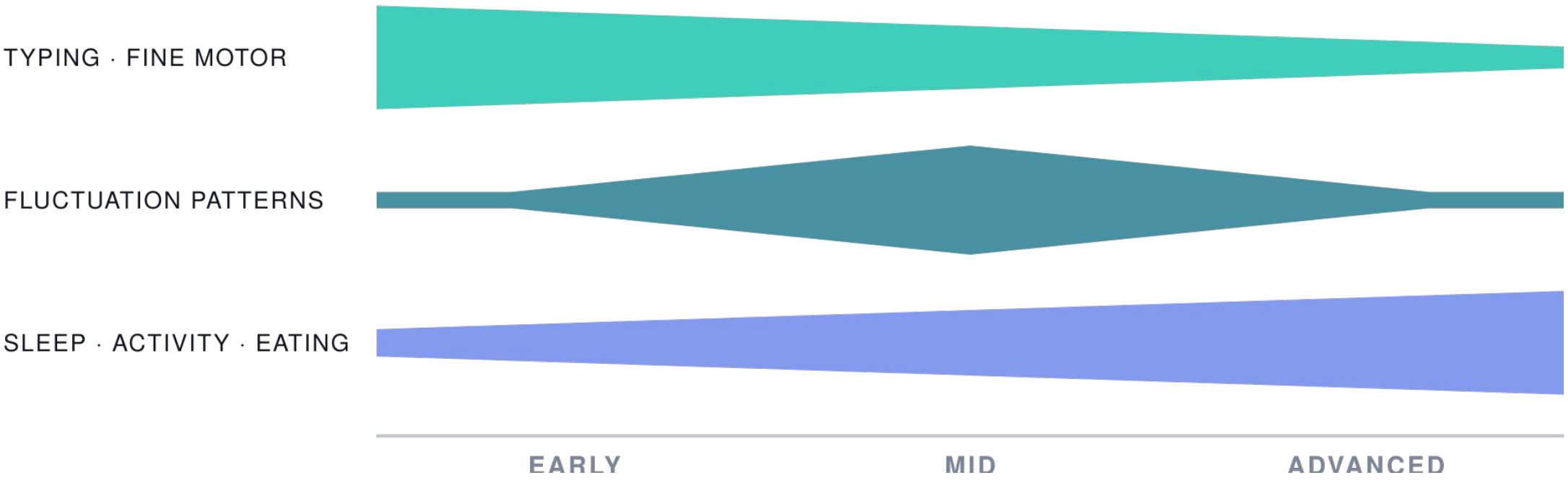
The comparison that matters: you vs. you



Change detected between visits - not at them.

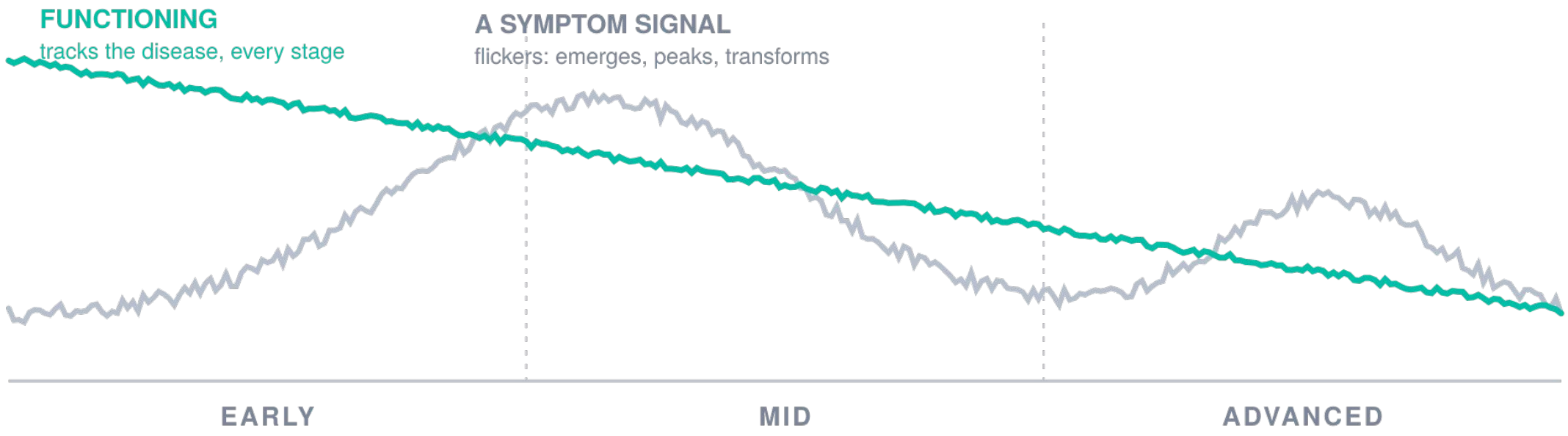
Illustrative trajectory

The disease moves through stages - so does the signal



Every single measure has a best stage - a composite shifts weight to where the signal lives.

Symptoms flicker across stages - functioning tracks the disease



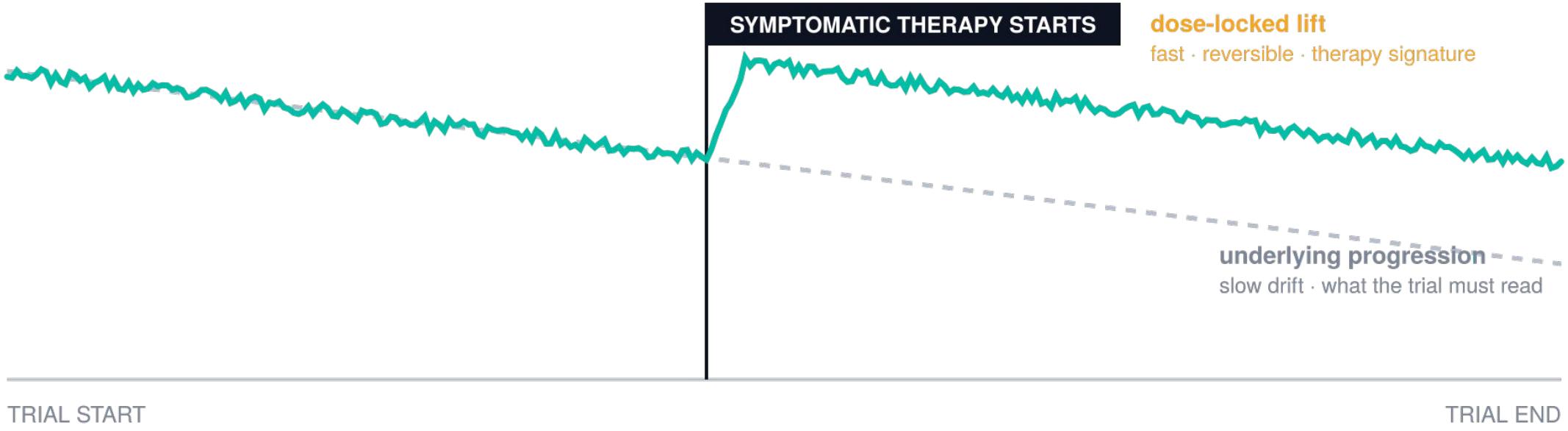
PHASE I · EXPLORATORY

PHASE II · SUPPORTIVE

PHASE III · CONFIRMATORY

Integrate functioning from the first-in-patient study - endpoints mature with the program.

Early therapy moves the floor — passive sensing separates drug from disease



Two processes, two timescales — separable because we never stop watching. That is what raises the probability of regulatory and technology success.

Disease Modifying Trials on Parkinson's Disease:

Current Endpoints Are Statistically Blind to **Early PD Progression and rarely quantify patient functioning**

Insensitivity at Early Stage

Clinician-administered UPDRS Part II detects only a small effect over 5 months of natural history progression.

This drives the Probability of Success that matters for **Disease Modifying Therapies**

A recent analysis of Parkinson's Progression Markers Initiative (PPMI) data by Evers L. et al. reported that while the UPDRS within-subject reliability is notably weaker when tracking subject-specific changes over time.

d = 0.17

Clinical UPDRS at 5 months

Visit-Dependent, Snapshot Assessments

Standard PD trials use 2–4 clinical visits per year. Each visit captures a single-day score, missing the fluctuation, compensation, and slow drift between appointments.

Missing data from non-attending subjects is common, and informative dropout is a known confounder.

2–4×/yr

Typical clinical visit frequency

Respondent Calibration Bias in Self-Report, Literature is noisy

Within-subject variability (CV) of self-reported UPDRS Part II and PDQ-39 is markedly elevated during the first ~50 days of **repeated administration before stabilizing**. Using this noisy window as baseline inflates Cohen's d by ~83%.

Literature hasn't produced fair evidence so far on what's a reliable, clinically meaningful and sensitive to change outcome measure for early PD trials

~50 day

SR calibration artifact window

The Q-index study design

130 patients / 4 Clinics

ISPPC azdelta CHU de Liège AZ Oostende

External Datasets for exploration

PPPMI Studywatch (50 PD @ H&Y =1) | iPROGNOSIS

Validation study:

The Q-index: Validating Passive-sensing-derived biomarkers of patient functioning

Longitudinal ~ 9 months

3 Visits

H&Y Stage:

H&Y 1: 10 (11.1%)

H&Y 1.5: 11 (12.2%)

H&Y 2: 32 (35.6%)

H&Y> 2: (40%)

Weekly: UPDRS II

Monthly: PDQ-39

Continuous Passive BYOD Wearables + Smartphone-derived biomarkers

(accelerometer/gyroscope/HR/GPS/Touchscreen typing)

Scientific Analysis Principles:

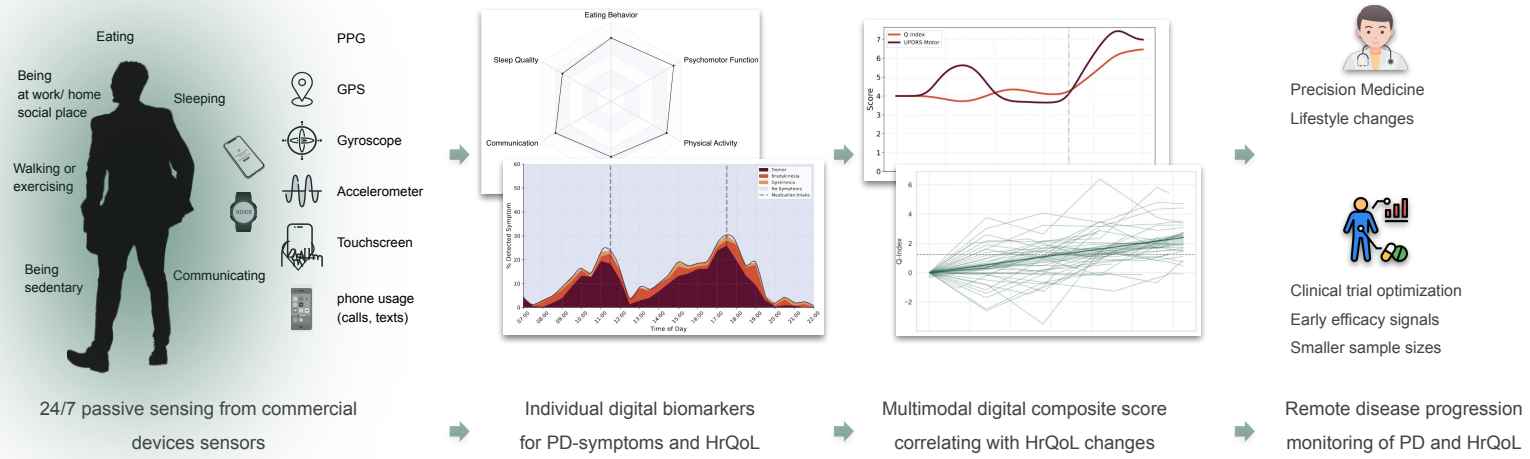
A) Clinically-meaningful for Patient-functioning (i.e. UPDRS II items)

B) Excellent Reliability (ICC > 0.75)

C) Sensitive to change (through cohen's D lense)

D) Clinical Relevance (with existing items)

Regulatory pathway: CE Class IIa (device) + dPerfO qualification as exploratory endpoint under FDA COA framework

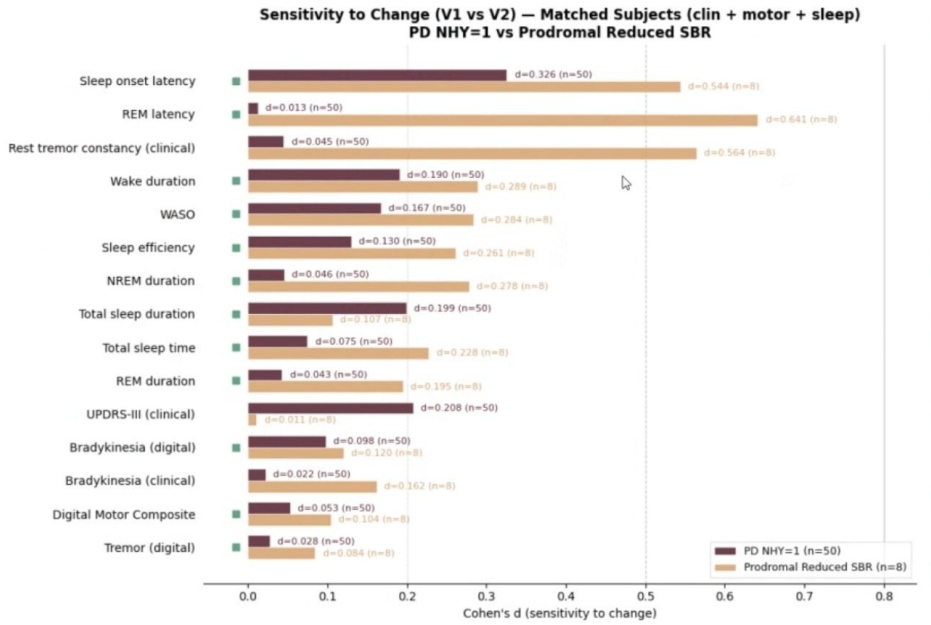


PPMI Dataset: Results from Koios Care Algorithms in PPMI Dataset

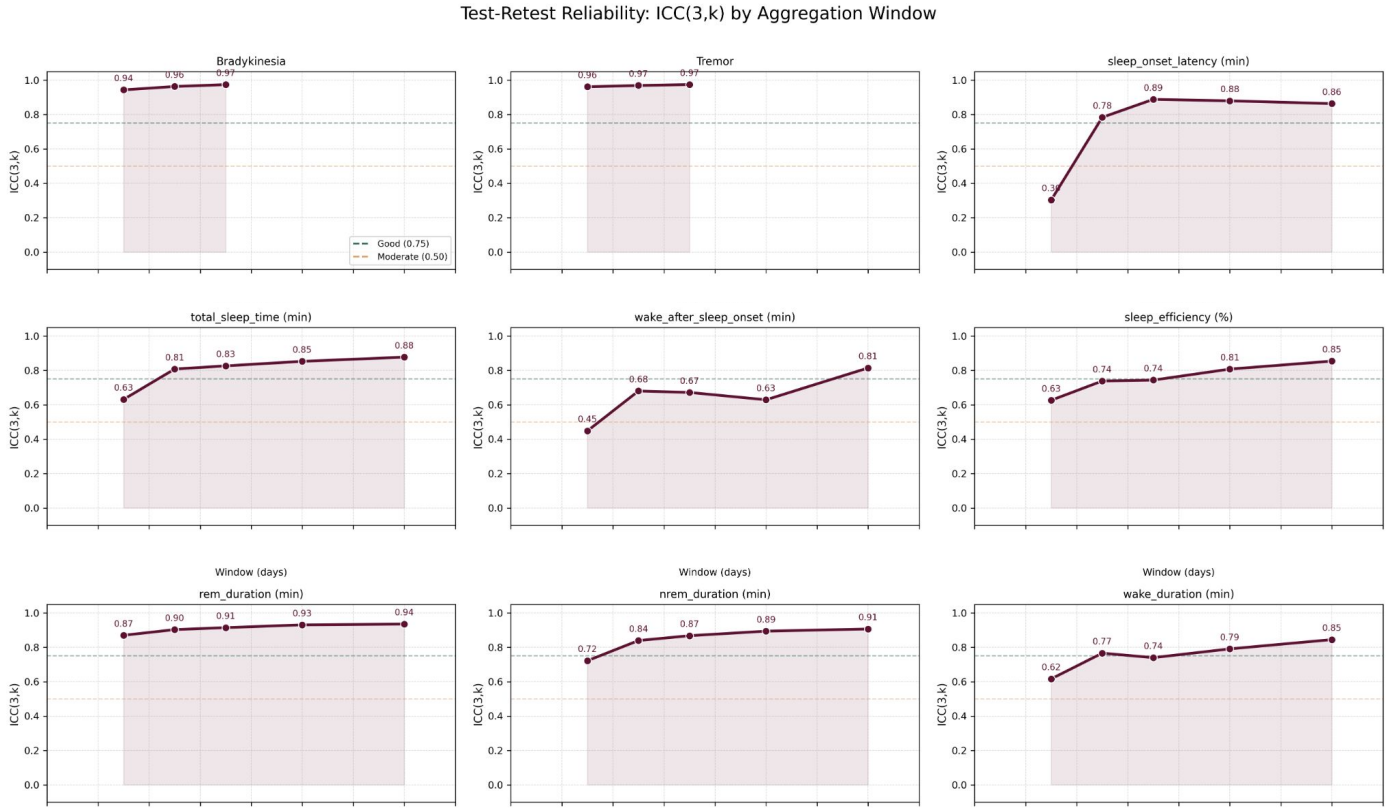
Reliability is computed using 30-days windows after baseline

Sensitivity to change is computed through Cohen's d is evolving when we compare sliding 3-week windows v.s. baseline

Investigating data from PPMI Verily Study Watch substudy to assess sensitivity to change over 6 months across sleep, daytime activity and behavioral domains in 50 individuals with early PD and 8 at-risk individuals with DaTscan-derived reduced SBR.



Using exploratory longitudinal data from PPMI with Verily Study Watch substudy in a total of 50 individuals with early PD and 8 at-risk participants with positive SBR findings, we examined sensitivity to change within 6-months across sleep, daytime and motor aspects.



Composite of Passive Monitoring Amplifies the Detectable Signal

How Q-Index Works

01

Passive multimodal sensing

Continuous accelerometry-derived measures of upper-limb hand-movement, tremor, and daily activity are captured using wearable and smartphone sensors, without active patient input

02

Muli-idomain features monitoring is the only way to go

A machine-learning model estimates changes in UPDRS Part II between consecutive wearable monitoring windows (W1 → W2), using paired clinical assessments.

03

Cumulative longitudinal trajectory

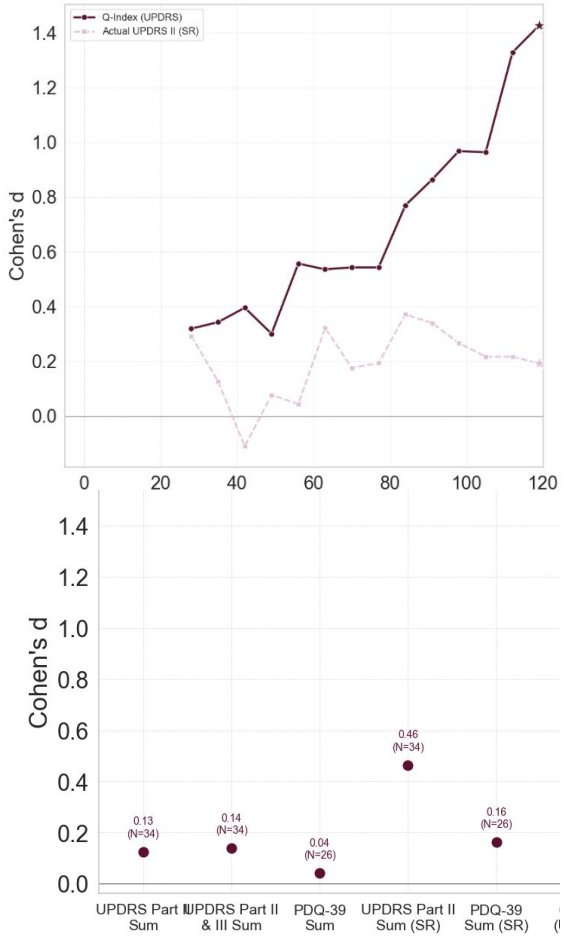
Predicted changes from consecutive monitoring windows are integrated into a cumulative trajectory, enabling characterization of of change across the full monitoring period between clinical visits.

04

Reduced susceptibility to response-shift bias

Passive sensing avoids reliance on retrospective self-report. In our analyses, self-reported outcomes showed greater variability during the initial ~30-day calibration period, with raw effect sizes inflated by up to 83%.

Sensitivity to Change: Q-index v.s. UPDRS II



Evidence Base

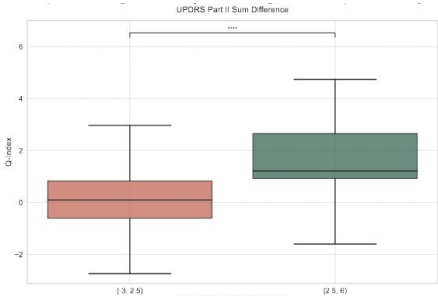
Analytical Validation

› Retained features that demonstrated ICC > 0.75

› At F1, Q-Index-related measures showed Cohen's d = 1.04–1.62 (N = 33–34), compared with d = 0.13–0.17 for clinical UPDRS measures.

› Sleep- and motor-related digital features captured longitudinal differences across the prodromal-to-early PD spectrum in PPMI.

› MCID classification: sensitivity ≥ 75%, specificity ≥ 70%



Confound & bias characterisation

› Self-reported UPDRS II showed evidence of response-shift effects, with the raw effect size decreasing from d = 0.44 to d = 0.24 after correction.

› An approximately 30-day calibration period was characterized in N = 94 participants.

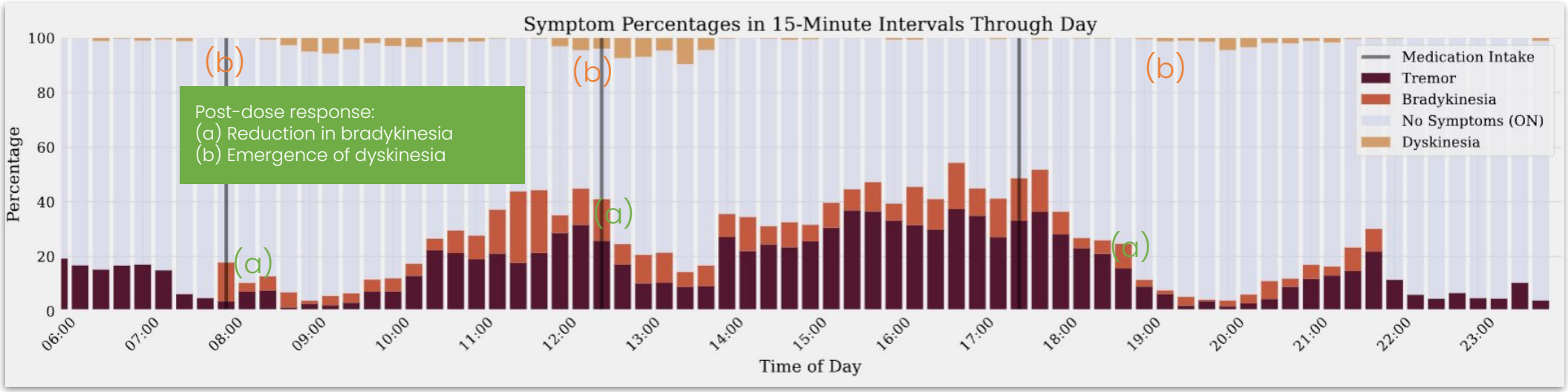
Clinical correspondence

› The cumulative Q-Index trajectory was associated with clinician-rated UPDRS change at temporally matched visits (median assessment proximity: 1–8 days).

› Consistent signal across 4 study sites (AZD, CHA, LIE, STND cohorts)

› As a passive measure, the Q-Index does not rely on symptom recall and is therefore less susceptible to self-report calibration bias.

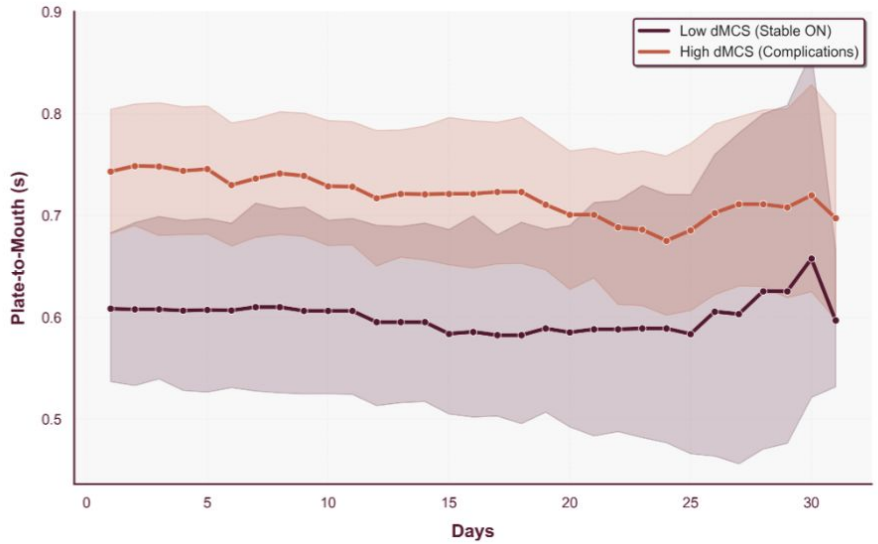
Can the measures be invariant of DMT signal detection despite levodopa variation?



Intraday relationship between **medication intake timing** and **validated digital measures** allow for personalized denoising of indicators.
The measures are sensitive enough to characterize efficacy in an intraday fashion.

This preemptively showcase that the system can capture changes related to minute level dopaminergic changes; changes that a successful **DMT trial will halt/reverse in the long run.**

For example; extracting the Plate-to-mouth (P2M) signal from a single subject can be denoised based on the timing of the medication.



Kyritsis, K.,et al. Assessment of real life eating difficulties in Parkinson's disease patients by measuring plate to mouth movement elongation with inertial sensors. Nature Sci Rep 11, 1632 (2021)

Composite sensitivity is engineered, not inherited

DILUTION

Coverage and responsiveness pull in opposite directions — an insensitive stream can blunt a sensitive one

WEIGHTING

The right weights follow the signal — stage, patient, and data availability. Fixed weights go stale

ANCHORING

Sensitivity must be anchored to change patients feel — not to whatever moves most

Not a reason to avoid composites — the reason their construction is a discipline.

Not everything that matters is measurable

FROM SENSORS

- Mobility
- Dexterity
- Sleep
- Eating & activity

NOT FROM SENSORS

- Stigma
- Emotional wellbeing
- Social life

Passive measures complement the patient's voice — they don't replace it.

Bring your own device: adherence by design

- **No new device**

  The watch is already on the wrist

- **No new habit**

  The phone is already in the pocket

- **No daily reminder of disease**

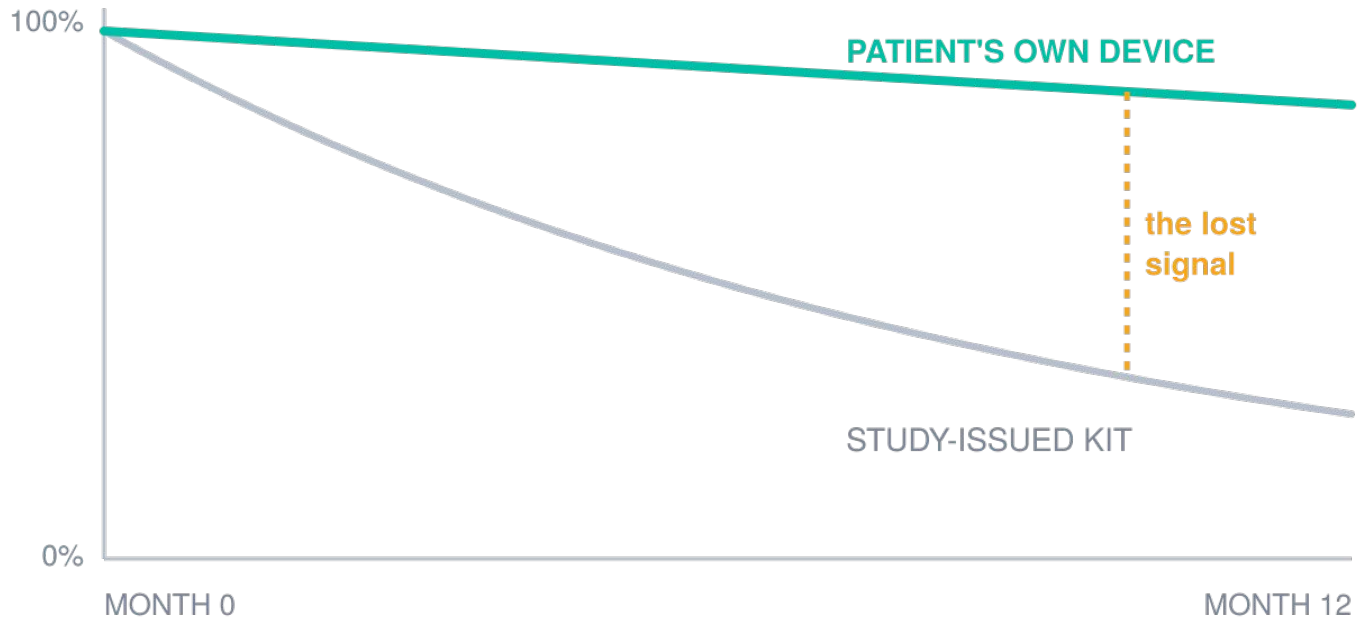
  Measurement disappears into routine

the patient's own devices

study-issued kit: one more charger, one more reminder

The best adherence tool is a habit the patient already has.

Dropout isn't noise — it's the signal you lose



Attrition is never random

The patients who drop out are the ones declining

Retention is bias prevention

Not logistics — longitudinal validity depends on it

Illustrative curves

One measurement stream, two lives

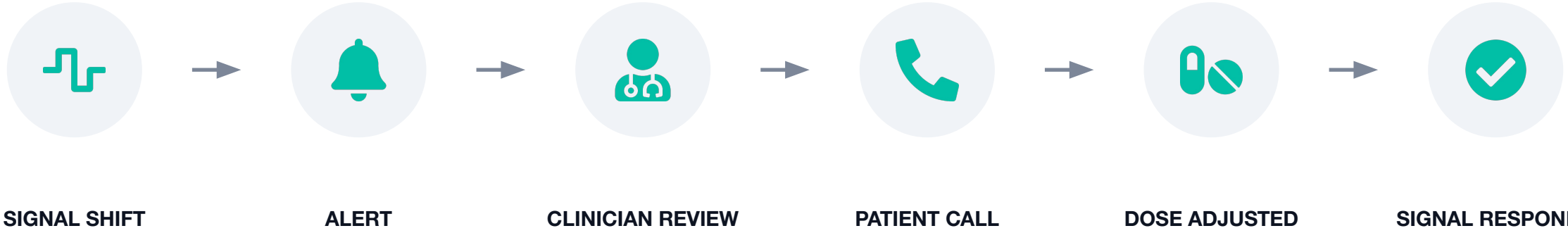
IN TRIALS

- Continuous functioning endpoint
- Reflects what patients value
- Change between visits, not at them

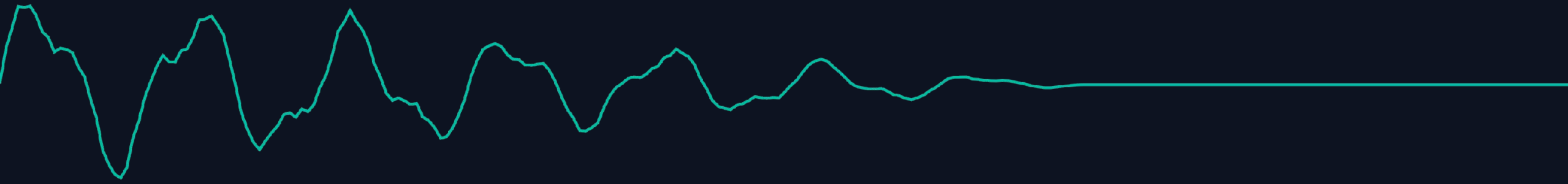
IN CARE

- Trends & alerts between visits
- The neurologist sees real life, not recall
- Treatment judged against daily functioning

From the wrist to the clinic: what an alert becomes



the loop closes — the next signal tells you whether it worked



The best assessment is the one the patient never has to perform — because they're already performing it, every day, just by living.

Thank you

koios.care