Development of a digital endpoint in Multiple Sclerosis – challenges and opportunities

Fabian Model
Director Biostatistics & Franchise Lead Neuroimmunology

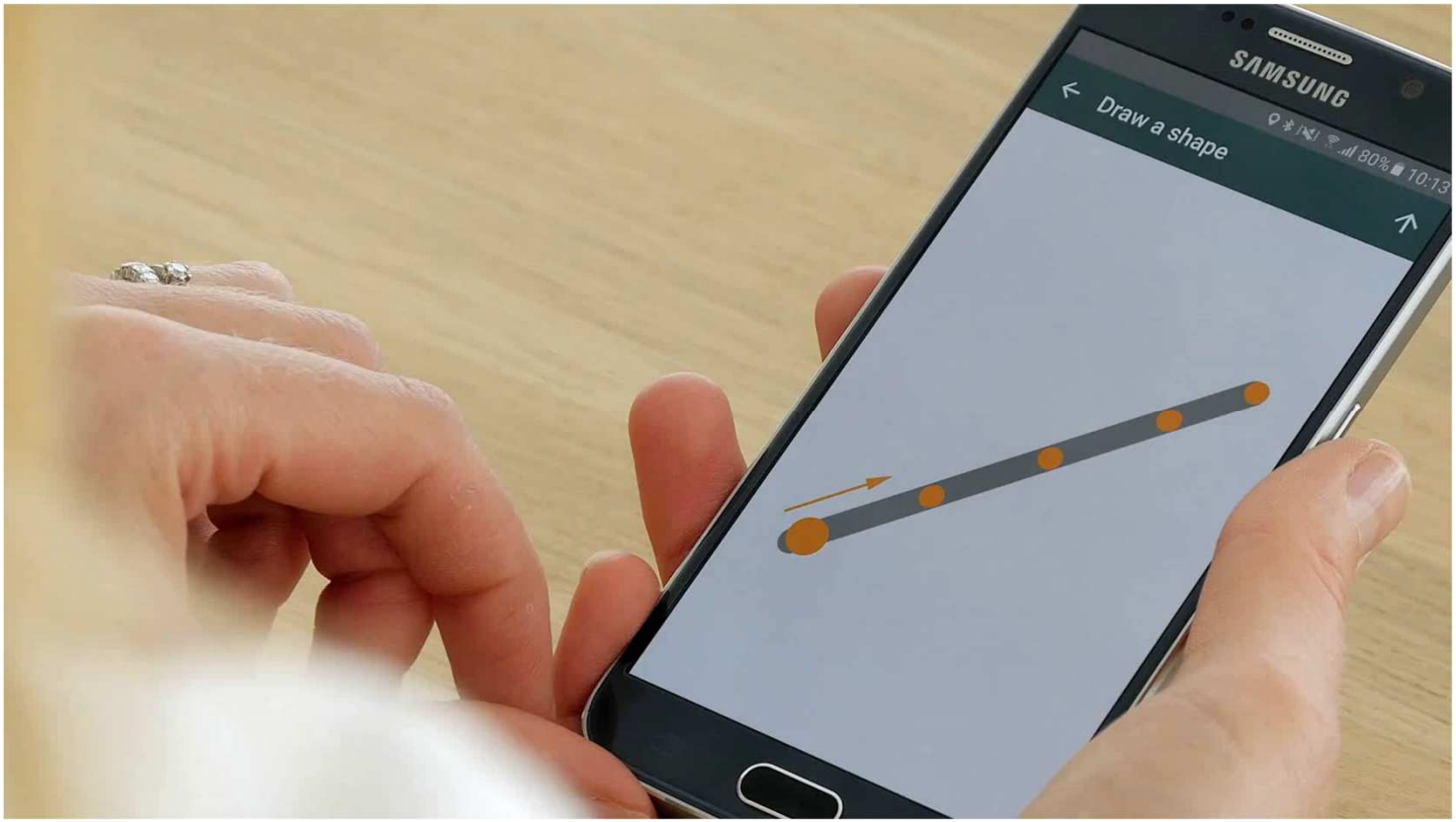
4th EFSPI regulatory statistics workshop, Basel, 23rd September 2019

Floodlight™ – Roche´s Digital Platform in MS

Smartphone based data collection: Suite of Active Performance Tests, Passive Monitoring & ePROs

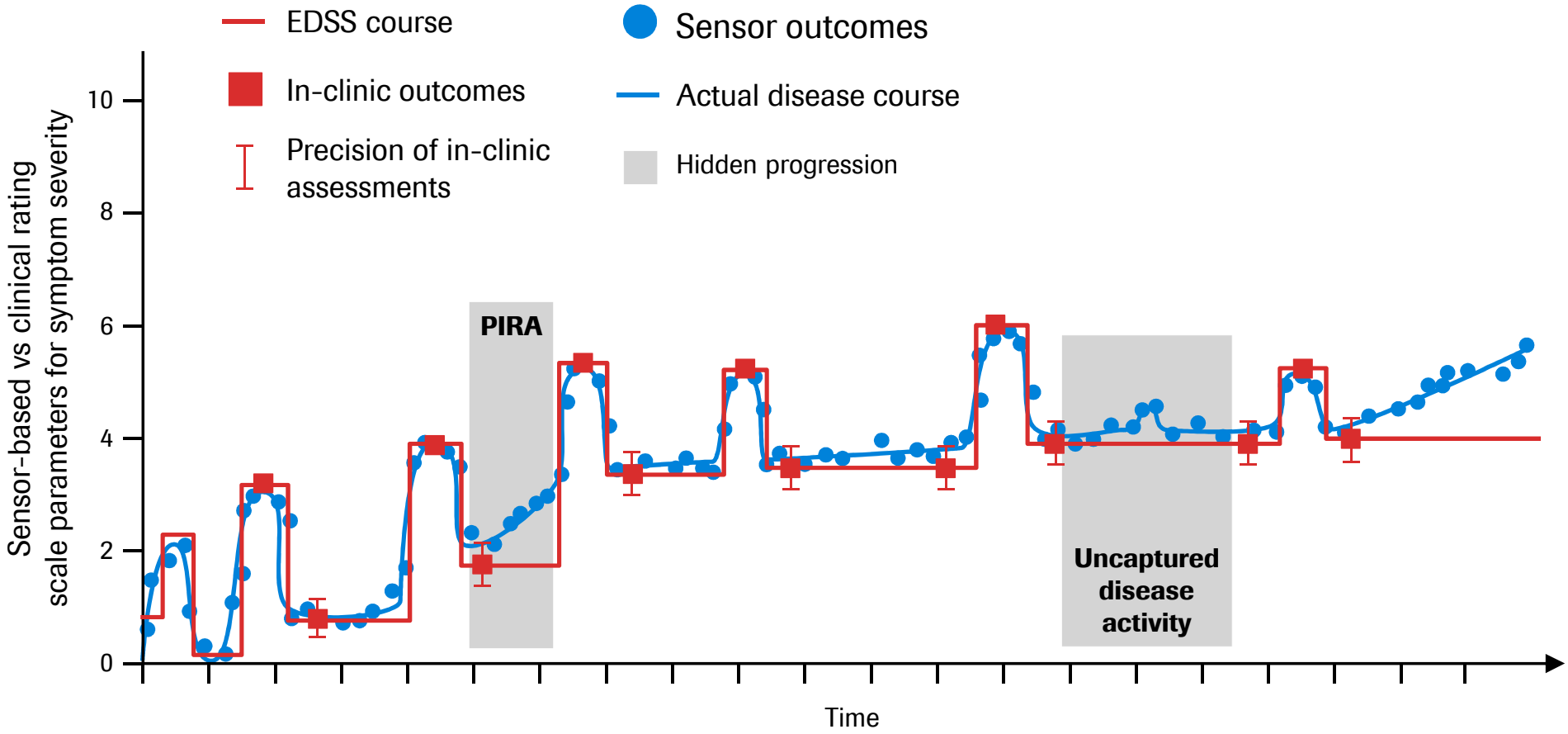
	Active tests									Passive monitoring	
Test type	Experience sampling			Cognition	Hand & arm		Gait & posture			Gait & posture	
Test name	Daily Mood Question (DMQ)	Symptom Tracker (ST)	Multiple Sclerosis Impact Scale (MSIS-29)	Information Processing Speed (IPS) Test	Pinching Test	Draw a Shape Test	Static Balance Test (SBT)	5-U-Turn Test (5UTT)	2-Minute Walk Test (2MWT)	Gait behavior	Mobility pattern
Frequency	Daily	Fortnightly & ad hoc	Fortnightly	Weekly	Daily	Daily	Daily	Daily	Daily	Continuous	Continuous

Active Test Example: *Draw a Shape Test*



Advantages of Digital over Standard Clinical Assessments

1) Remote, continuous monitoring of patients in their daily life



EDSS, Expanded Disability Status Scale; PIRA, progression independent of relapse activity

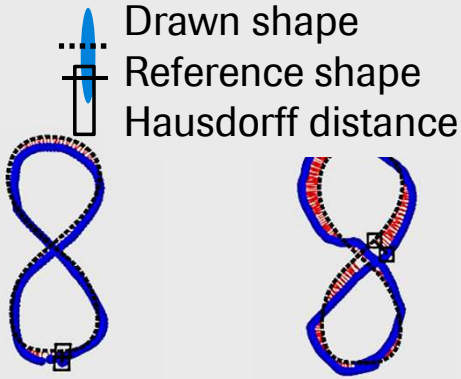
MSE UCSF et al. Ann Neurol. 2019;85:653–66

Advantages of Digital over Standard Clinical Assessments

2) Granularity of captured data allows deeper disease phenotyping

Accuracy

Pixel points sampled relative to interpolated reference coordinates



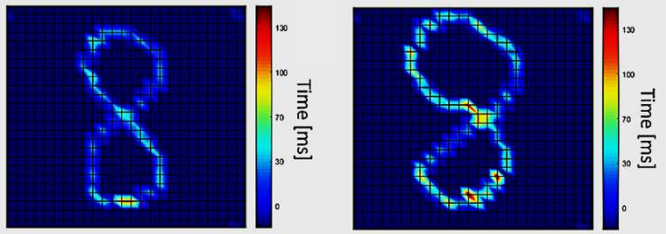
Healthy control
9HPT: 21.75 s

Subject 1
9HPT: 33.8 s

Query points and maximal Hausdorff distances are shown in red and black squares, respectively: surface of red area is measured

Celerity

Pixel density heat map grids



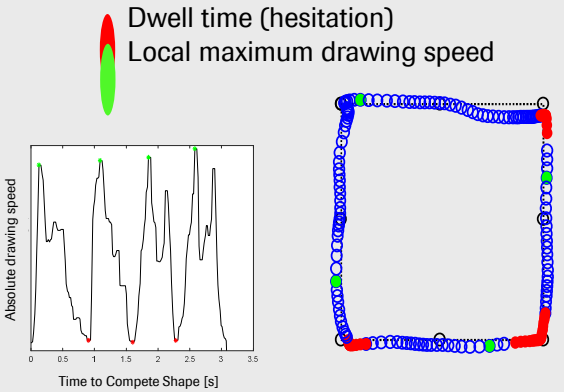
Healthy control
9HPT: 21.75 s

Subject 1
9HPT: 33.8 s

Areas of hesitation or non-movements are colored

Speed

Temporal drawing speed



Healthy control
9HPT: 21.75 s

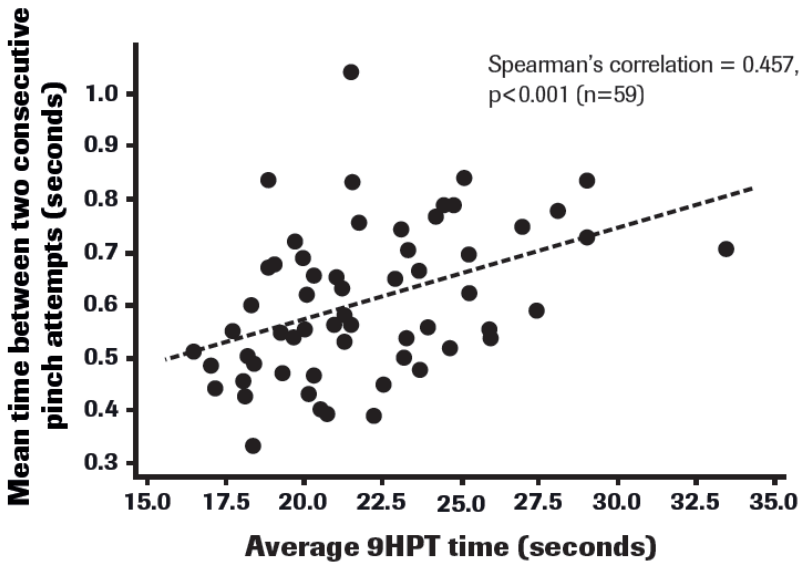
Local maximum speed highlighted in green; Dwell (hesitation) time is marked in red at corner locations

NB: Images included are only intended as illustrative examples from Roche digital outcomes studies
9HPT, 9-Hole Peg Test. Roche data on file

Digital Outcomes Validity – Current Evidence in MS

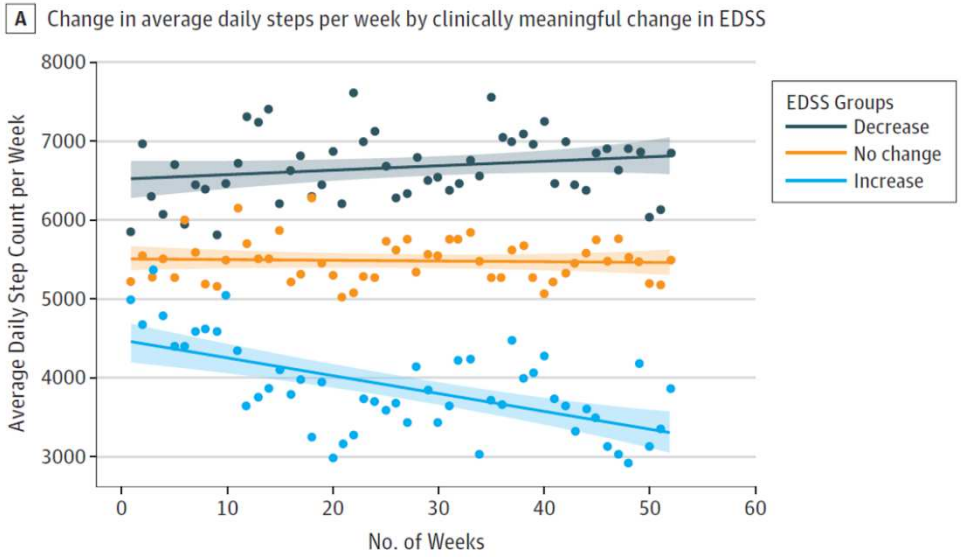
First Proof of Concept Studies in Multiples Sclerosis

Floodlight active & passive tests show cross-sectional correlation with conventional in-clinic outcome measures



Montalban X et al. ECTRIMS, 10–12 October 2018, Poster 382. Berlin, Germany

Continuous monitoring of Step Count: Longitudinal correlation with disability progression



Block VJ et al, Association of Continuous Assessment of Step Count by Remote Monitoring With Disability Progression Among Adults With Multiple Sclerosis, JAMA, 2019

Challenges with current assessments & Digital Ambition in MS

Clinical trial endpoints

- MS is characterized by phenotypic heterogeneity
- Current outcome measures have limitations in precision and sensitivity to change
- Outcome measures that capture improvement are not available
- Enable Decentralized Clinical Trials

Assessing MS in clinical practice

- Limited use of quantitative measures
- No feasible solutions for frequent monitoring of disease activity or progression
- Full administration of current tools are costly
- Better tools to predict disease course are needed
- Enable ubiquitous RWD collection

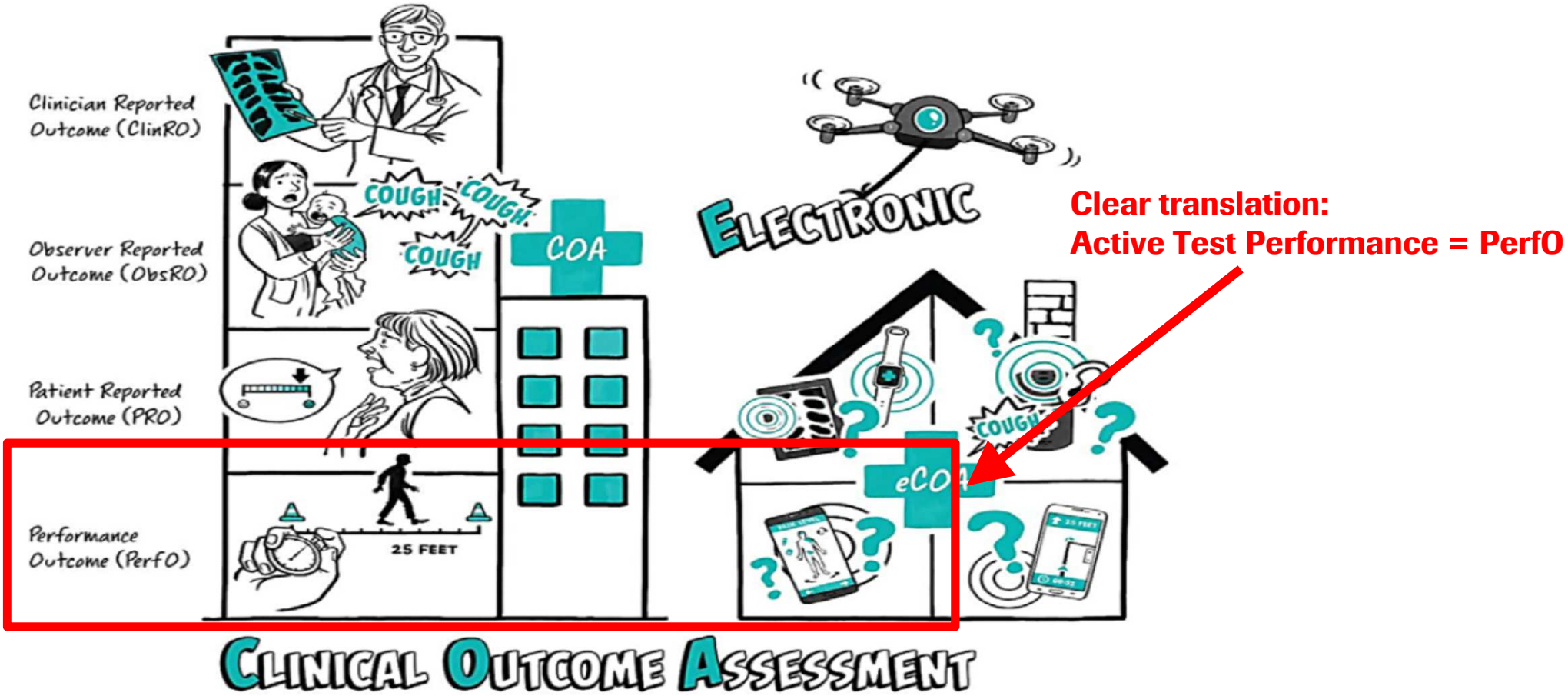
Our ambition: qualify digital measures as regulatory-grade label-enabling endpoints and make them available as measurement tools in clinical practice

Endpoint Qualification Procedure
FDA (CDER) & EMA

Software as a Medical Device
FDA (CDRH) & EU Notified Bodies

Clinical Outcome Assessments

Digital tools fit into the framework but borders get blurred



Clear translation:
Active Test Performance = PerfO

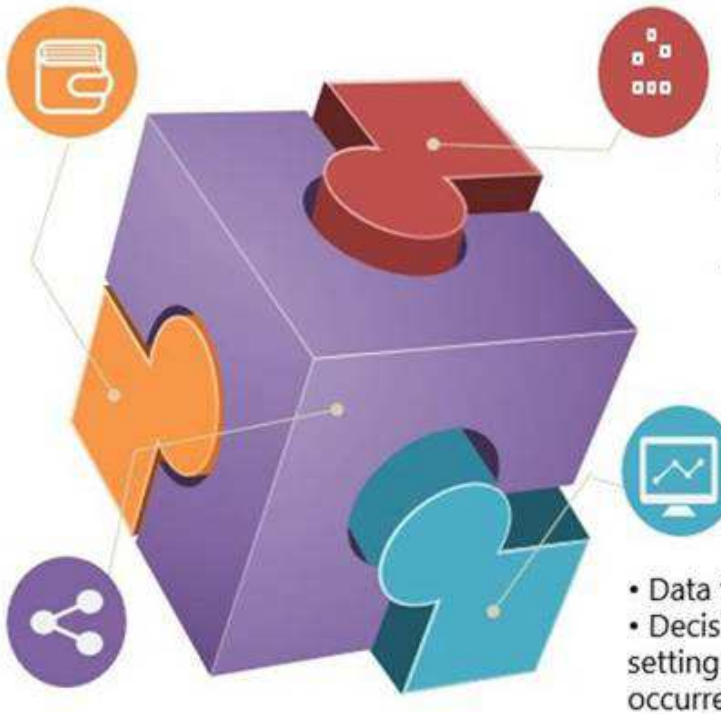
Coravos A et al. Digital Medicine: A Primer on Measurement, Digit Biomark, 2019

Evidence required for COA qualification

Content validity	Construct validity	Reliability	Sensitivity to change
▪ Develop concept of interest and context of use	▪ Correlation with other related measures	▪ Test-retest variability	▪ Mean-to-SD ratio of decline
▪ Generate evidence that the instrument measures the concept of interest	▪ Discrimination of known groups	▪ Day-day, Device-device variability	▪ Worsening during clinical/sub-clinical activity
▪ Patient understanding		▪ Biological variability	▪ Longitudinal correlation with clinical assessment
▪ Patient burden			▪ Longitudinal change predicts (long-term) disability

Key Design, Implementation & Analysis Considerations

FDA Regulatory Perspective: Digital Health Technology Tools



Instrumentation and Instrument Validation

- Device model and manufacturer
- Documentation of instrument validation

Data Collection

- Data collection environment
- Duration of data collection period
- Days of the week for monitoring

Variable Selection and Endpoint Definition

- Concept to be assessed
- Clear definitions of selected variables
- Well-defined, reliable, and clinically meaningful endpoint(s)

Data Processing, Scoring, and Analysis

- Data file preparation and transfer
- Decisions regarding time interval setting (daily diary, episodic event occurrence)
- Scoring criteria
- Missing data rules
- Clinically meaningful within-patient change

Challenge: Reliable collection of digital data outside the clinic

How can we adapt good data and record management to the digital remote patient monitoring setting?

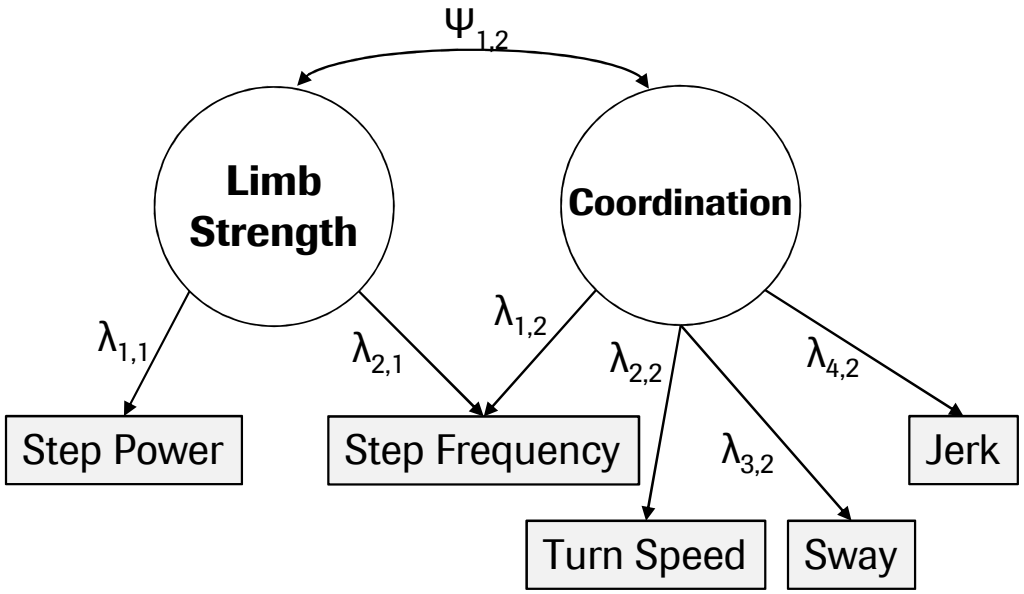
ALCOA Principle	Advantages for Digital	Open Questions
Attributable	• Reliable documentation of measurement device	• How do we prove data comes from the patient? e.g. eSignatures or Biometric fingerprints?
Legible (traceable and permanent)	• No human interaction in data handling • Possibility to reach 100% legibilaty & traceability	
Contemporaneous	• Modern cell phones have ability to reliably sync their clocks • Possibility to guarantee 100% reliable time stamps	
Original	• Possibility to store full source sensor data	• Potential conflict with data privacy requirements (e.g. full GPS location or environmental audio)
Accurate		• How do we prove patient accurately performed assessment? e.g. patient eSignature, statistical quality control, outlier detection?

Challenge: Deep digital phenotyping vs. COA with face validity

Example: Potential eCOA for Gait Domain

Latent Disability Constructs

Indicators: Sensor Data Derived Features



Structural Equation Model (SEM)

$$\boldsymbol{\Sigma = \Lambda \Psi \Lambda' + \Theta}$$

- How to establish link between sensor data features and disease concepts meaningful to patients?
  - Qualitative patient studies
  - Movement disorder & disease experts
- Endpoint performance vs. Face validity
  - Power of digital lies in deep and rich phenotyping of patients
  - Likely requiring multivariate sensor feature scores

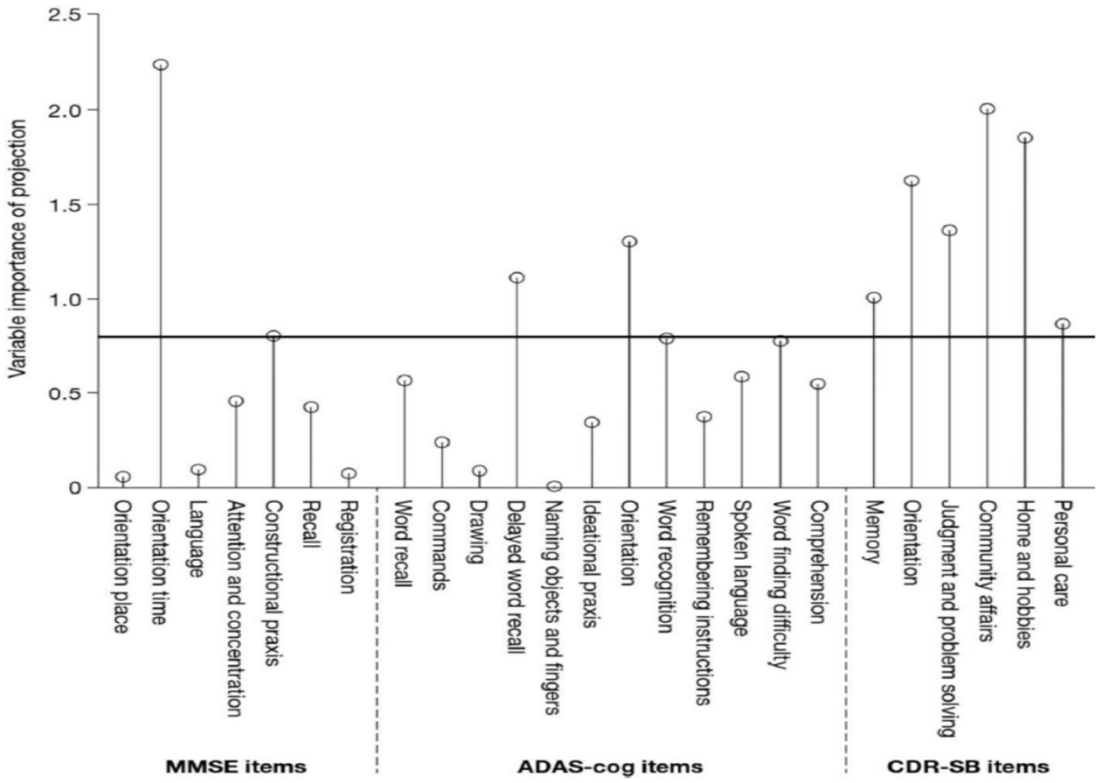
Challenge: Deep digital phenotyping vs. COA with face validity

Weighted Composite Score Example: ADCOMS (Alzheimer's Disease)

Weighted score of establsihed PerfO, ClinRO & ObsRO optimized to detect linear longitudinal change:

$$(t - t_0) = \sum_{i=1}^{12} d_i \Delta A_i(t) + \sum_{i=1}^{7} e_i \Delta B_i(t) + \sum_{i=1}^{6} f_i \Delta C_i(t) + \ldots$$

Implemented in clinical trials, e.g.
primary outcome in phase II BAN-2401



Wang et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials, J Neurol Neurosurg Psychiatry 2016

Challenge: Deep digital phenotyping vs. COA with face validity

Different views on value of weighted composite COAs

FDA scientists recently published a critical review on weighted composite scores in AD:
(Jin K, Cameron B, Dunn B. On weighted composite scores for early Alzheimer's trials. Pharmaceutical Statistics. 2019)

Active group: $Y_i^w = \sum_{l=1}^{k} w_l Y_{il}$ Placebo group: $X_j^w = \sum_{l=1}^{k} w_l X_{jl}$

Under normality assumptions test statistics for weighted and unweighted scores are:

$$H_0 : \mu_Y = \mu_X \qquad \frac{\sqrt{n}\left(\overline{Y}^w - \overline{X}^w\right)}{\sqrt{2\boldsymbol{w}^t\Sigma\boldsymbol{w}}} \sim N(0,1) \qquad \frac{\sqrt{n}\left(\overline{Y} - \overline{X}\right)}{\sqrt{2\mathbf{1}^t\Sigma\mathbf{1}}} \sim N(0,1)$$

Corresponding Power:

Weighted: $1 - \Phi\left(Z_{1-\alpha} - \frac{\sqrt{n}(\boldsymbol{w}^t(\boldsymbol{\mu_Y} - \boldsymbol{\mu_X}))}{\sqrt{2\boldsymbol{w}^t\Sigma\boldsymbol{w}}}\right)$ Unweighted: $1 - \Phi\left(Z_{1-\alpha} - \frac{\sqrt{n}(\mathbf{1}^t(\boldsymbol{\mu_Y} - \boldsymbol{\mu_X}))}{\sqrt{2\mathbf{1}^t\Sigma\mathbf{1}}}\right)$

Power maximized for: $\boldsymbol{w} = \Sigma^{-1}(\boldsymbol{\mu_Y} - \boldsymbol{\mu_X})$

Author argument: without knowing treatment effect no optimal choice of weights possible => use unweighted score

Counter argument: assume trt effect as common %reduction of pcb decline => weighting optimizes signal/noise & power

Optimal combination of high dimensional sensor data will be key for success of digital outcomes!

Summary

- Upiquitous digital technology offers tremendous potential for clinical research
  - Deep phenotyping of patient's symptoms & function
  - Remote patient monitoring and decentralized clinical trials
  - Efficient monitoring and management of disease
  - Efficient collection of meaningful RWD
- Many open questions remain
  - Best practices and regulatory framework for development and qualification of eCOAs
  - Reliable & demonstrable data quality
  - Score & endpoint definitions
  - Handling of missing data
  - …

Doing now what patients need next