Opportunities and Challenges that arise with digital endpoints: Floodlight as a case example

BBS Annual Meeting, 06 November 2023

Stanislas Hubeaux, Principal Statistical Scientist, Floodlight Project Lead Statistician

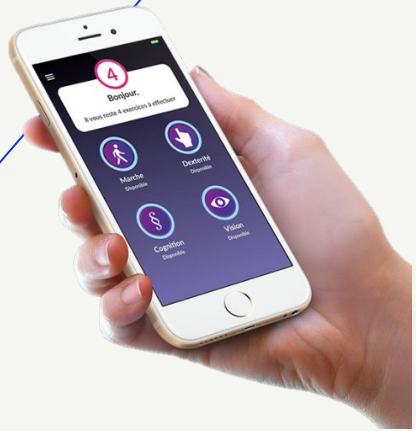


Table of contents

Introduction to Floodlight

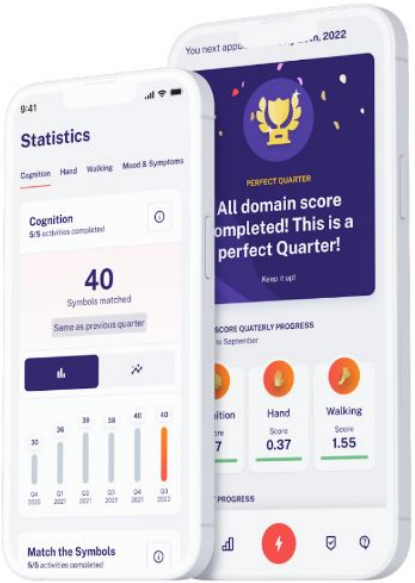
What is Floodlight?

Roche's initiative to develop DHT tools in Neuroscience

Floodlight (or FLOODLIGHT) is not one single device, product or technology.

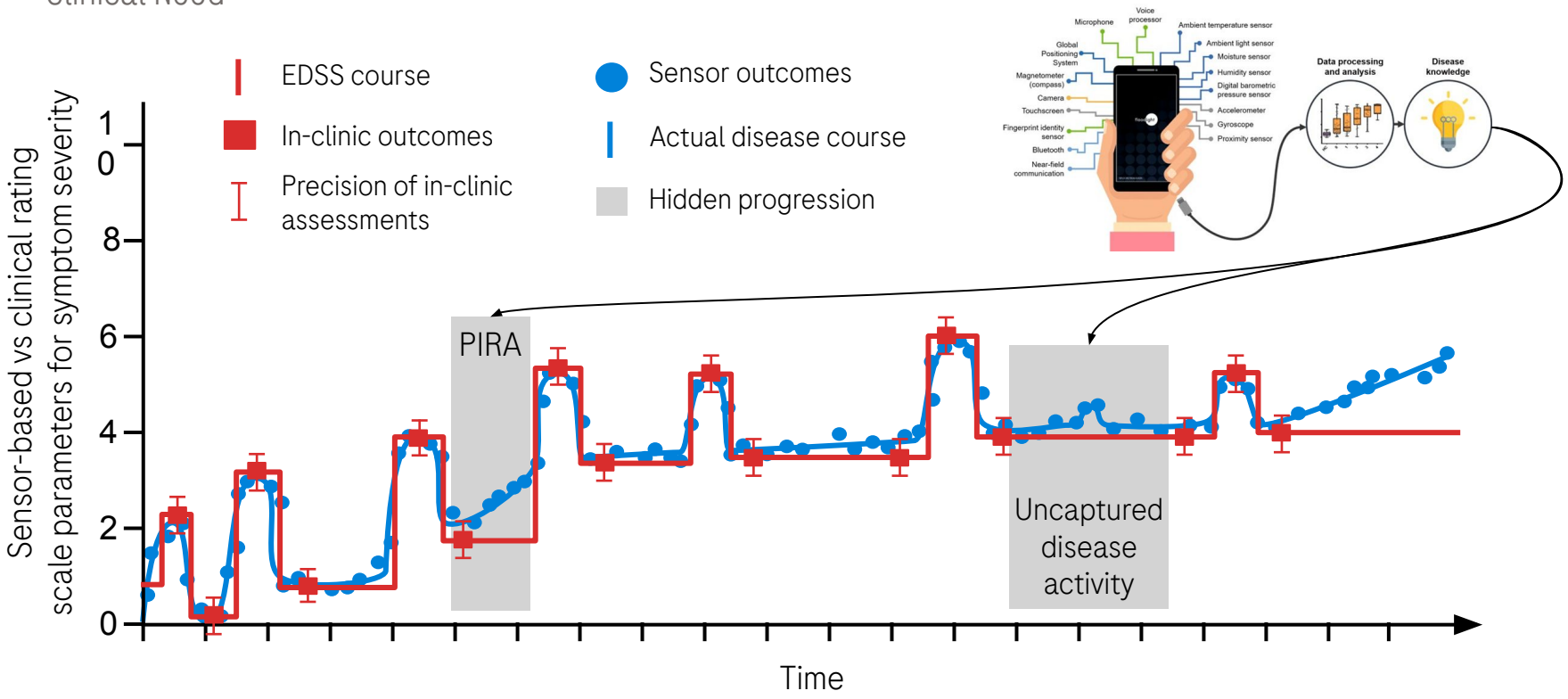
Many software and hardware technologies at various stages of research and development have emerged from the Floodlight pipeline.

The Floodlight programme's current approach to app development is a ***"bring your own device" software-based test battery for use in a home setting as a remote instrument for sensor-based measurement of symptoms of cognitive and physical disability, such as is experienced by PLwMS --***



What is Floodlight?

Clinical Need



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

Kappos et al. JAMA Neurol. 2020; https://doi.org/10.1001/jamaneurol.2020.1568

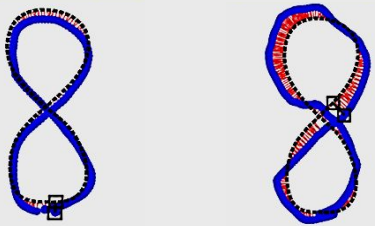
Advantages of Digital over Standard Clinical Assessments

Granularity of captured data allows deeper disease phenotyping

Accuracy

Pixel points sampled relative to interpolated reference coordinates

- Drawn shape
- Reference shape
- Hausdorff distance



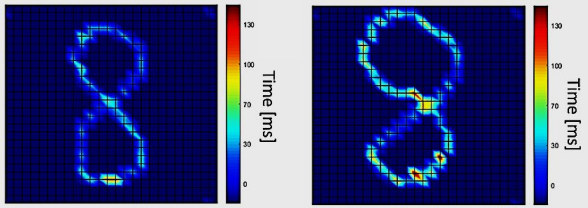
Healthy control
9HPT: 21.75 s

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

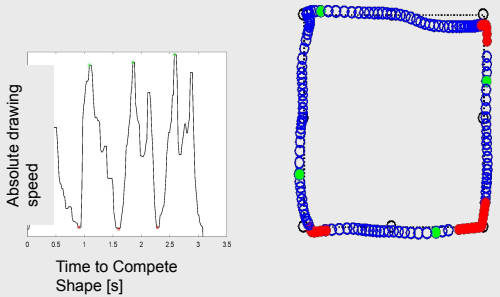
Subject 1
9HPT: 33.8 s

Areas of hesitation or non-movements are colored

Speed

Temporal drawing speed

- Dwell time (hesitation)
- Local maximum drawing speed



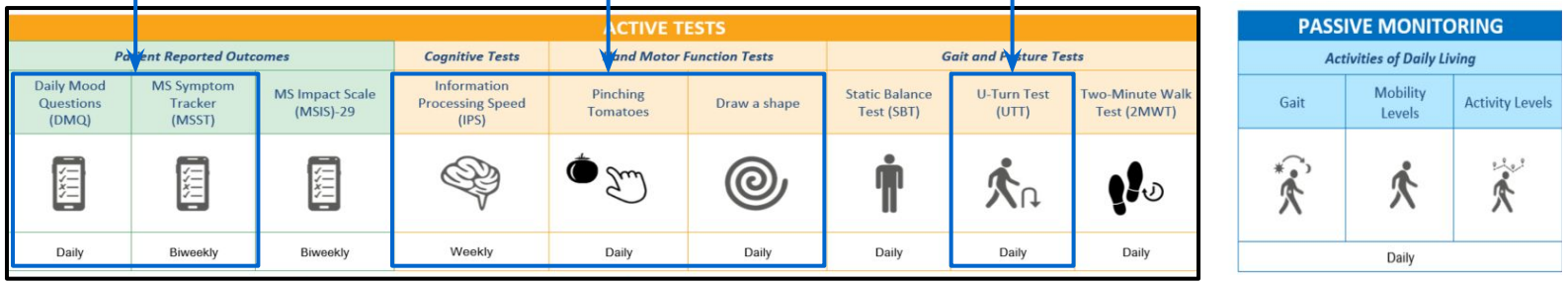
Healthy control
9HPT: 21.75 s

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

Overview of FLOODLIGHT™ assessments

Development of Floodlight

- A list of Active Tests and Passive Monitoring are currently being developed within Floodlight
- As part of the development of Floodlight, in-clinic assessments are collected via clinical trials (6 studies, almost 2000 patients (target ~**3650 patients**), with some now using Floodlight for as long as 4 years)
- FLMS is now available in **10 countries** (AUS, AUT, CH, DE, ES, FIN, IT, PT, UK, USA) with almost **1100 patients** actively using Floodlight in clinical practice
  - Currently only a selected number of Active Tests are available in the commercial version of Floodlight

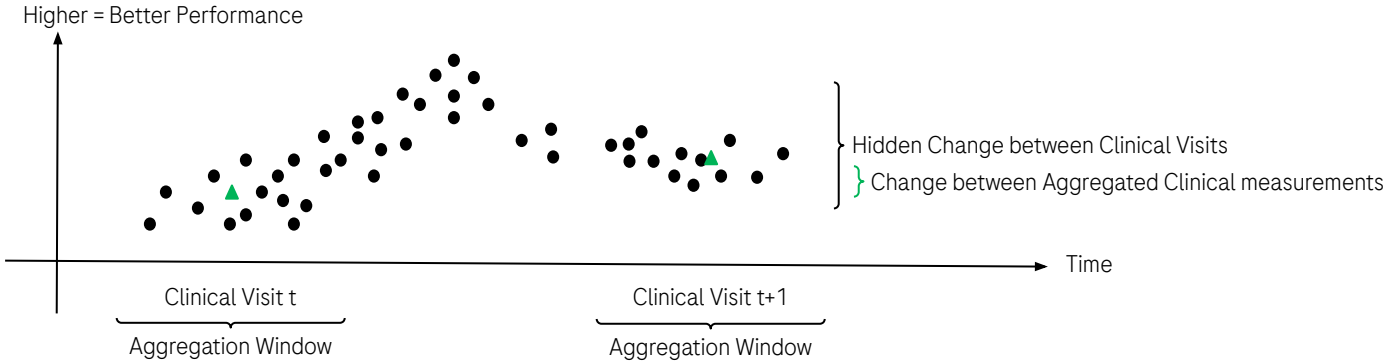


Longitudinal Use of High Density Data

Longitudinal Use of High Density Data

Motivation

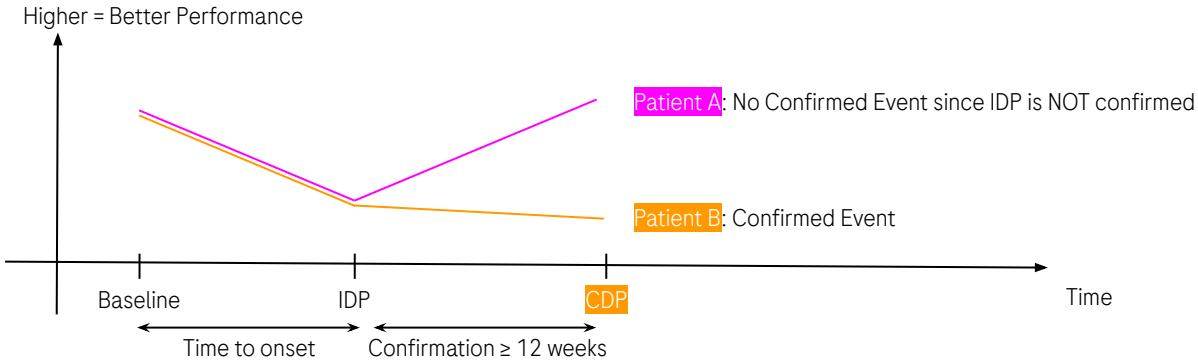
- **Longitudinal Association with a gold standard anchor** (e.g. a clinical measurement) is a key activity of the development of a DHT
- Since a DHT enables collection of continuous data (no “hole” between clinical visits), simple aggregation of data around time points (e.g. clinical visits) would be suboptimal (see example below)
  - There is a clear need to define a way of using the complete amount of collected data when defining change based on the DHT (e.g. detection of disease progression)



Standard Clinical Metrics in Multiple Sclerosis

Definition Time to Confirmed Events

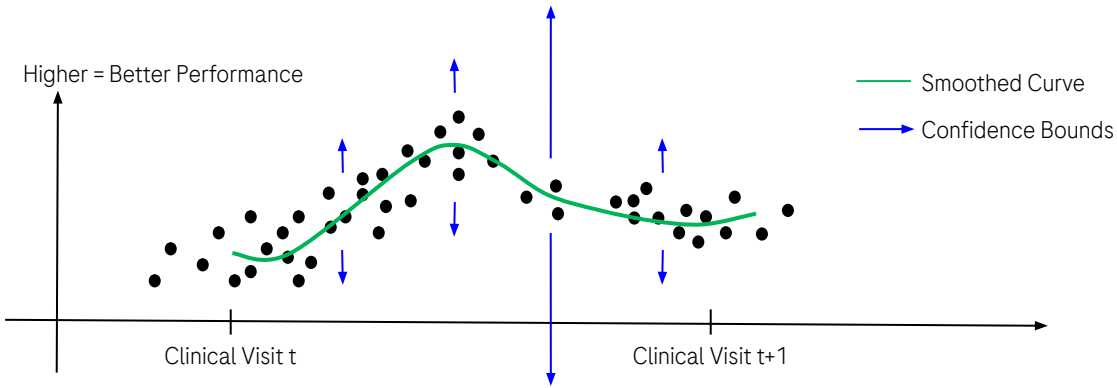
- Let's first evaluate how change (and more specifically disease progression) is currently defined for clinical trials in MS
  - Time to onset of the first 12-week CDP is the time from baseline to the onset of the first disability progression that is confirmed at the next clinical visit (for example ≥ 12 weeks) after the initial disease progression (IDP)
- **Concept**: Create a similar Time to Confirmed Events but based on high density data collected via a DHT



Longitudinal Use of High Density Data

Confidence Interval around Smoothed Curve

- **Goal: Transform the continuous DHT data into a time-to-event endpoint**
- Requiring all collected data points to be below a certain Threshold isn't realistic given the variability of data collected in an at-home setting
  - **Proposal**: Construct Confidence Bounds around smoothed DHT data which is representative of the local number of data points. Then require the Confidence Bounds to remain below a certain Threshold for a specific time period.



Longitudinal Use of High Density Data

Implementation of methodology

- Methodology developed by Prof. Peter Bühlmann (Department of Mathematics, ETH Zurich)*
- It was implemented in R by Zheng Chen Man during an internship at Roche under the supervision of Fabian Model, Frank Dondelinger, and Stanislas Hubeaux
- Event Detection Using Confidence Bounds (edecob) **R Package available on Rcran**
  a. Construct moving median
  b. Use autoregressive model on residuals of the moving median
  c. Bootstrap to find pointwise confidence intervals
  d. Use those pointwise confidence intervals to construct simultaneous Confidence Bounds for the estimated moving median

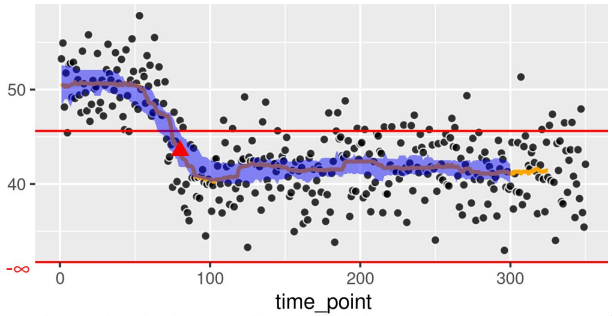
*Peter Buhlmann. Sieve bootstrap for smoothing in nonstationary time series. Ann. Statist., 26(1):48–83, 02 1998.

Event Detection Using Confidence Bounds (edecob)

Analytical Validation

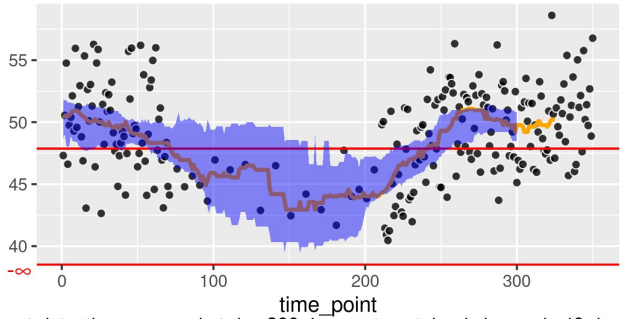
- To evaluate if the developed methodology works as expected, we generate multiple simulations
- Below we present 3 of them. We can observe that the Confidence Bounds (i.e. purple area) is well dependant (i.e. larger when fewer data points) on the local number of data points (i.e. black dots) used to construct the moving median (i.e. the orange line)

Simulation 1: Ideal Event



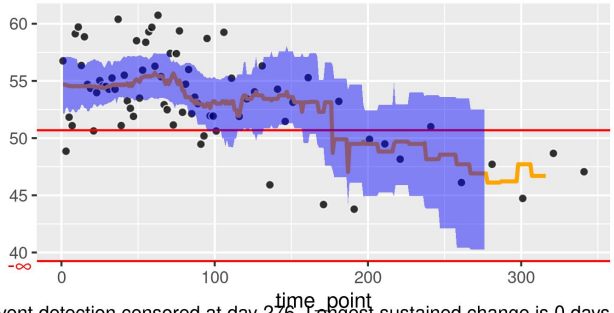
Event detected at day 80, sustained for 221 days.
Minimal duration of change for event detection: 84 days
Level of confidence band: 95%

Simulation 2: Temporary Decrease in Frequency



Event detection censored at day 300. Longest sustained change is 40 days
Minimal duration of change for event detection: 84 days
Level of confidence band: 95%

Simulation 3: Continuous Decrease in Frequency



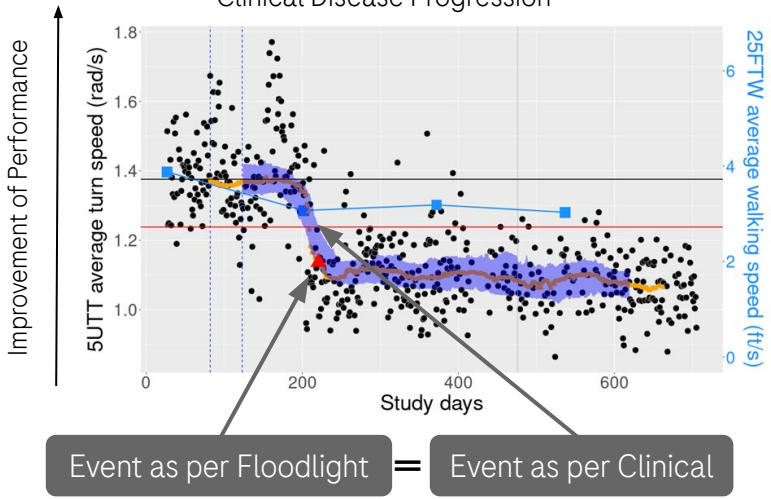
Event detection censored at day 276. Longest sustained change is 0 days.
Minimal duration of change for event detection: 84 days
Level of confidence band: 95%

Event Detection Using Confidence Bounds (edecob)

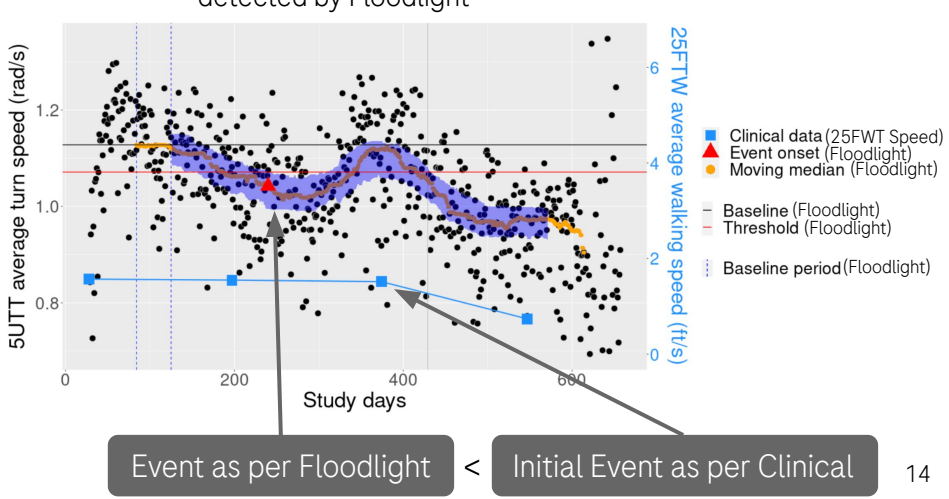
Clinical Validation

- It is now required to evaluate if the events (i.e. disease progression) detected using the proposed methodology is concordant with the events detected using the standard clinical metric
- This is still ongoing for Floodlight, but we observe very promising results when visually inspecting the individual patient graphics

Patient A: Synchronized Floodlight & Clinical Disease Progression



Patient B: Earlier Disease Progression detected by Floodlight

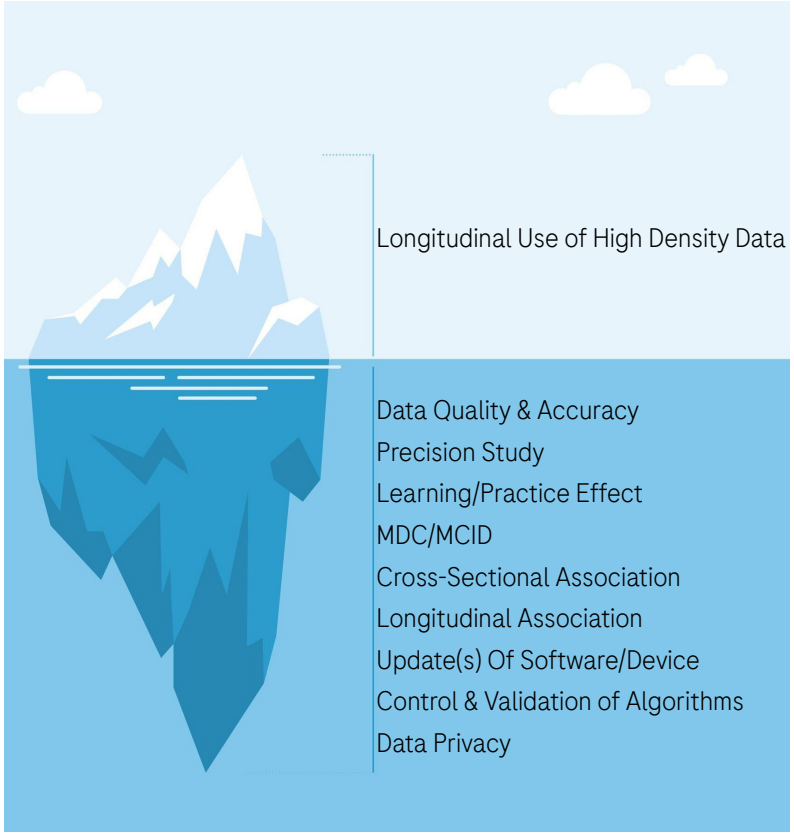


Longitudinal Use of High Density Data

Remarks

- **Next Steps**: The Clinical Evaluation (i.e. correlation with and up to potentially surrogacy of a clinical metric) needs to be continued and finalized
- **Sky it the limit** when defining the notion of change on high density data (non-exhaustive list)
  - Change in variability could also be indicative of a disease progression (evaluation for Floodlight are already ongoing)
  - Missing data points might be as well very informative (e.g. a patient stops doing a Floodlight Test due to loss of Hand Motor Function, patient doesn't feel the need of doing Floodlight since feeling good)
- High density data, especially collected via a DHT in an at-home setting, also comes with **challenges** (non-exhaustive list)
  - Requiring high density data (e.g. daily) can be very cumbersome for the patients. This might be suitable for a short study (e.g. a few months) but not for standard MS studies which can last for years.
  - Deployment of an updated version of the DHT (e.g. to correct a bug, to improve the DHT) while maintaining comparability with data already collected

Conclusions



- Today we only scratched the surface of the number of activities which are key for for Data Sciences when developing a DHT
- DHTs enable to collect high density/frequency data. **This has the potential to uncover hidden progression!** It comes with a bunch of important challenges that need to be tackled.
- **Final Take Home Message**: before starting the development of a DHT, are these 3 main components of the context of use defined?
  - Disease and target patient population
  - Study type (e.g. non-pivotal, pivotal, and trial design including endpoint positioning);
  - Measurement/Administration principles (e.g. timing and frequency of assessment))

Iceberg image: https://www.freepik.com/free-vector/iceberg-infographic-concept_8514874.htm#query=iceberg%20drawing&position=40&from_view=keyword&track=ais

Doing now what patients need next

Back-Up

edecob Methodology

Statistical model for a single time point t

We use the following model:

$$Y(t) = S(t) + \eta(t),$$

$$\eta(t) = \sum_{j=1}^{p} \phi_j \eta(t-j) + \varepsilon(t) \quad \text{AR(p) model}$$

where

Y(t) are the data points,

S(t) the patient's hidden real performance,

$\varphi_1, \ldots, \varphi_p$ the parameters of the autoregressive model

$\varepsilon(t)$ i.i.d. errors with expectation zero.

We assume the $\eta(t)$ to be stationary.

edecob Methodology

Bootstrapping

3.4.1 Bootstrapping

We follow the article by Bühlmann [22]. We assume the data points consist of the sum of the patient's performance and some residuals which we model using an AR model as described in Section 3.3.2.The goal is to obtain confidence bounds for the unobserved signal $S(t)$. We then do the following procedure to obtain the bootstrapped $\hat{S}(t_i)_b^*$, $b = 1, \ldots, B$, where B is the number of bootstrap repetitions:

1. Compute $\hat{S}(t)$
2. Compute the residuals $\hat{\eta}(t_i) = X_{t_i} - \hat{S}(t_i)$.
3. Fit an AR(p) model to $\hat{\eta}(t_i)$ and obtain estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and residuals $\hat{\varepsilon}(t_i) = \hat{\eta}(t_i) - \bar{\eta} - \sum_{j=1}^{p} \hat{\phi}_j \hat{\eta}(t_i - t_{i-j})$.
4. Resample

$$\varepsilon(t_i)^* \text{ from } \hat{\varepsilon}(t_{p+1}), \ldots, \hat{\varepsilon}(t_n);$$
$$\eta(t_i)^* = \bar{\eta} + \sum_{j=1}^{p} \hat{\phi}_j \eta(t_{i-j})^* + \varepsilon^*(t_{i-j});$$
$$X_{t_i}^* = \hat{S}(t_i) + \eta(t_i)^*$$

5. Compute $\hat{S}(\cdot)^* = g(X_{t_1}^*, \ldots, X_{t_n}^*)$.
6. Repeat steps 4–5 B times to obtain $\hat{S}(t_i)_b^*$, $b = 1, \ldots, B$

edecob Methodology

Calculation of the confidence bounds

3.4.2 Construction of the simultaneous confidence bounds

After using bootstrap to compute the $\hat{S}(t_i)_b^*$ $b = 1, \ldots, B$, we will now construct a simultaneous confidence band of level α for $S(t_i)$. In the following we assume that $\eta(t)$ is a stationary noise process with mean zero. Note that we do not check this assumption when applying this method.

1. We compute the quantiles

$$\hat{q}_{\frac{\alpha_P}{2}}(t_i), \hat{q}_{1-\frac{\alpha_P}{2}}(t_i) \qquad i = 1, \ldots, N,$$

where $q_{\alpha_p}(t_i) = \inf\{u;\ \mathbb{P}^*[\hat{S}(t_i)_b^* - \hat{S}(t_i) \leq u] \geq \alpha_P\}$ is a pointwise bootstrap quantile and N is the number of measurements.

2. We vary the pointwise error α_P until

$$\mathbb{P}^*[\hat{q}_{\frac{\alpha_P}{2}}(t_i) \leq \hat{S}(t_i)_b^* - \hat{S}(t_i) \leq \hat{q}_{1-\frac{\alpha_P}{2}}(t_i) \quad \forall i = 1, \ldots, N] \approx 1 - \alpha.$$

In other words, until the ratio of bootstrap curves around $\hat{S}(t_i)$ within $[\hat{q}_{\frac{\alpha_P}{2}}(t_i), \hat{q}_{1-\frac{\alpha_P}{2}}(t_i)]$ is approximately $1 - \alpha$.

3. We let

$$I(t_i) = [\hat{S}(t_i) + \hat{q}_{\frac{\alpha_P}{2}}(t_i), \hat{S}(t_i) + \hat{q}_{1-\frac{\alpha_P}{2}}(t_i)] \quad \forall i = 1, \ldots, N.$$

Then $\{I(t_i)\}_{i=1,\ldots,N}$ is a consistent simultaneous confidence band of level $1 - \alpha$.