
# Biomarkers European Special Interest Group

**PSI Conference 2024**
**June 17, 2024**

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# Disclaimer

The opinions expressed in this presentation are those of the authors, and do not necessarily reflect the official policy of any of the employers listed.

# Biomarkers
# European Special Interest Group

Nicole Krämer[1] & Guillaume Desachy[2]
on behalf of the Biomarkers ESIG

[1]Global Biostatistics & Data Sciences, Boehringer Ingelheim GmbH & Co. KG, Biberach, Germany
[2]Research and Development, Pierre Fabre, Toulouse, France

June 17, 2024

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# 3 Goals:
# Methods, interactions & connections

1. Establish advanced analytical methods to analyse biomarkers for clinical development

2. Increase interaction with other disciplines (medicine, biology, academic research)

3. Connect with other Special Interest Groups

# A diverse group coming from industry & academia

👥 50+ members

🤩 Strong interest in biomarkers, with a focus on clinical development

👬 Co-Leads:
Guillaume Desachy (Pierre Fabre)
Nicole Krämer (Boehringer Ingelheim)

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# SIG kicked off in early 2022 & 3 priority topics identified!

🤝 Kick-off meeting in April 2022 - since then, monthly meetings

🔎 Priority topics so far:

Biomarker-based designs

Machine Learning for Biomarkers

Identification of publicly available biomarker datasets

# What a journey since 2022! 😍

📻 The Effective Statistician podcast with Alexander Schacht
https://bit.ly/3rqtA4I

📰 Poster presentation at the 2022 PSI conference

📺 PSI Webinars in January 2023 and November 2023
https://bit.ly/3Xo9ohw ; https://bit.ly/3ulRpjc

🎤 Biomarker ESIG Session at PSI conference 2023 & 2024

# Machine Learning as an enabler of precision medicine

## The PSI Biomarkers SIG Machine Learning / AI workstream group

Karl Köchert[1] and Nils Ternès[2]

[1]Bayer AG / Clinical Data Science & Analytics, Bayer, Berlin, Germany
[2]Sanofi R&D, Sanofi, Montpellier, France

June 17, 2024

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# Agenda

- Introduction of the PSI Biomarker ESIG AI/ML sub-stream
- AI/ML, biomarkers and drug development – a value proposition
- Getting hands on with the team – shaping the usage of AI/ML in clinical development

# Introduction of the PSI Biomarker ESIG AI/ML sub-stream

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# AI/ML workstream

- Group was set up in May 2023 as workstream of the PSI Biomarkers ESIG
- Currently ~25 members from ~15 different companies and academic institutions, attendance in meetings usually around 10-15 people
- Main focus of the group: liaising with like-minded peers to learn about, discuss and get hands on with applying AI/ML in clinical development, specifically in a precision medicine & biomarker context
- Monthly 1h early morning “all-comers” meetings + additional meetings of the “hands-on AI/ML projects core team”

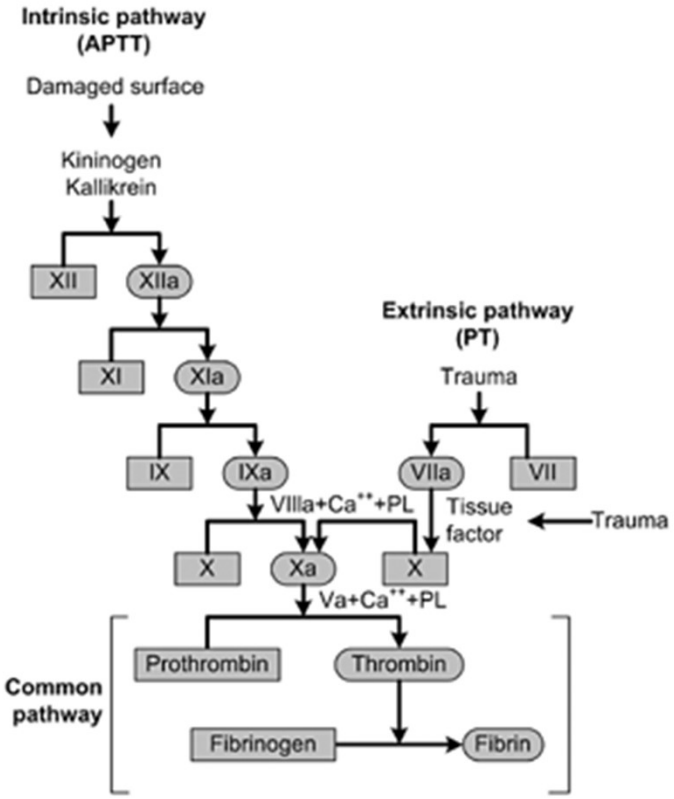
# Team members


*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# AI/ML, biomarkers and drug development – a value proposition

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# The complexity of biology




Some basic constituents of Homo Sapiens (far from being comprehensive):

- **19.300** protein coding **genes** (HGNC Oct 2023)
- A multitude of possible variations (~80m SNPs, splicing, post-translational modification) results in **~82.000 proteins** (UniProt Oct 2023)
- **1.900 miRNAs**
- **~300 cell types** in **78 organs**
- **…**

Over the last 2 decades technical progress enabled us to measure these building blocks of human biology at unprecedented resolution.

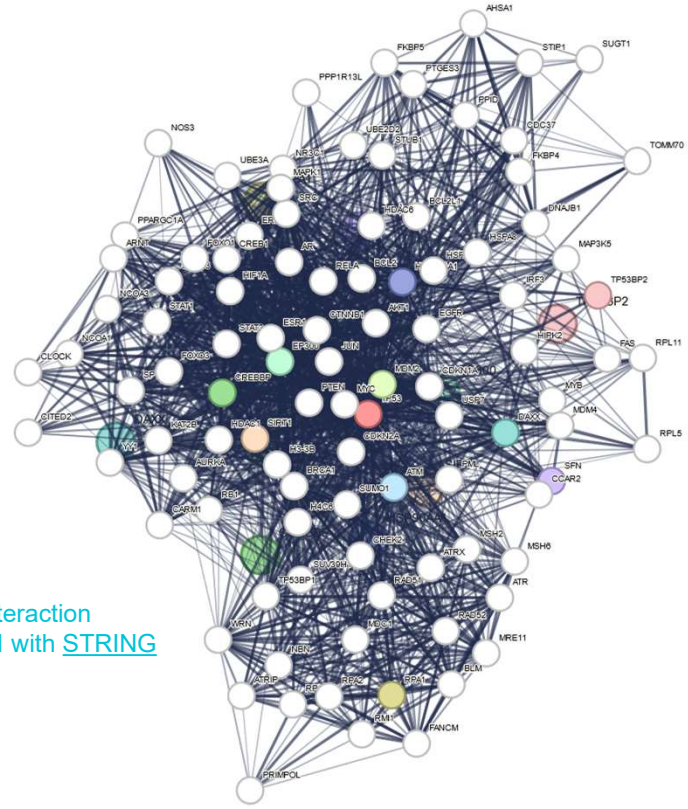
For some very important biological processes, we also have a good understanding how these building blocks interact and play together.

For most though, we don't.

The coagulation cascade. APTT: activated partial thromboplastin time; PT: prothrombin time

**Nephrology, 2009, DOI: (10.1111/j.1440-1797.2009.01128.x)**

# Diseases add another level of complexity – patients are extremely heterogeneous

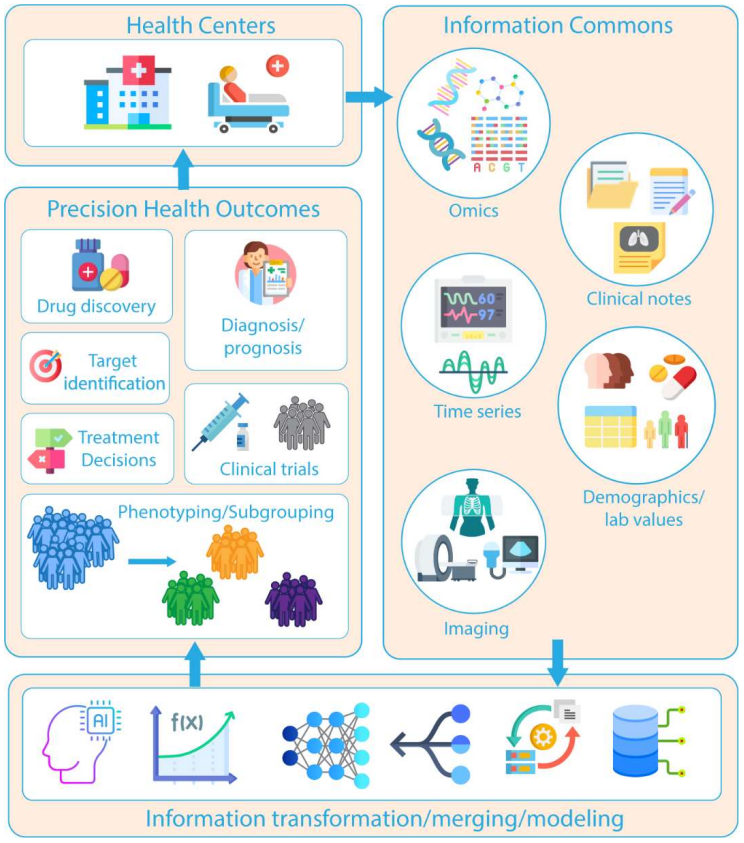



TP53 protein interaction network created with STRING v12.0

Given the utmost complexity of an individual patient's disease trajectory, in the setting of a multifactorial / multicausal disease, *precision medicine* is required to tailor a specific therapy.

As FDA describes it: "Precision medicine […] tailors disease prevention and treatment for *individual variability* (e.g., genetic and lifestyle differences among patients). The goal of precision medicine is to *match the right treatments at the right dosages for each individual patient at the right time*. The challenge for precision medicine is identifying the mechanistic basis for adverse events […] and differences in efficacy […]"

# Precision Medicine & quantifying disease complexity in the 2020s




Kline et al., Nature Digital Medicine 2022

In this vein, (pharma) research & clinical development has already made good progress in the last couple of years in terms of measuring things: -omics data, digital biomarkers, imaging etc. are being assessed increasingly.

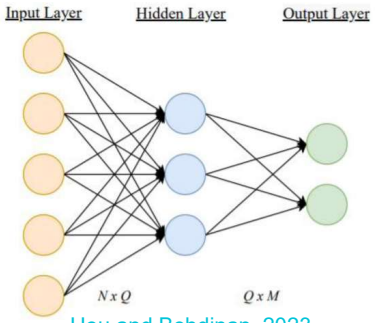
This has pushed the number of variables available in a typical interventional clinical trial from ~150 to much higher levels.

*In other words – the testable hypothesis space has grown almost exponentially compared to ~20 years ago.*

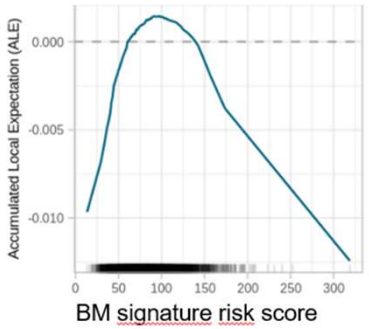
# So how does AI/ML fit into the vision of precision medicine?

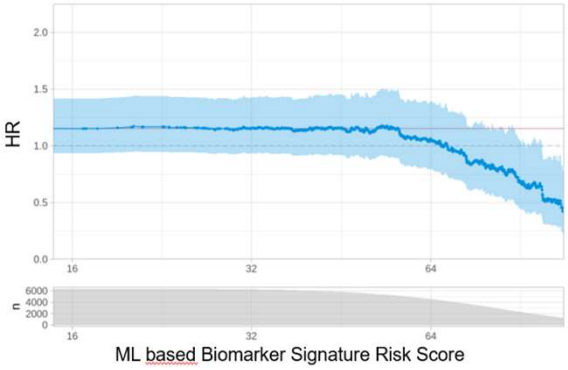
- Classical approaches are *"forcing" a hypothesis on the data* & are limited to detection of strong, low-complexity signals using only a small amount of the variables (measured data) at one time, e.g., they are usually restricted to treatment, age, BMI, disease severity at baseline etc.
- By *deriving the best hypothesis given the data*, Machine Learning (ML) is currently the best available methodology to create holistic mathematical models of complex (biological) systems using all available data and variables.
- As such, AI/ML methodology is key in enabling true *data driven decision making* and a *prerequisite for precision medicine* since it complements classical approaches to enable as thorough as technically possible *insights generation*.

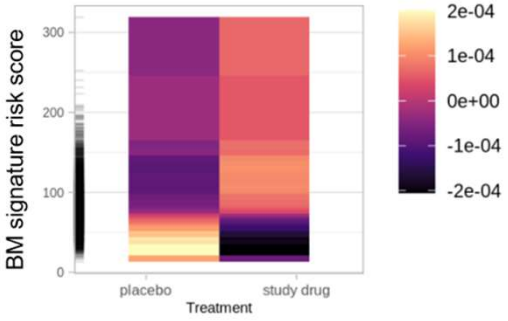
# Where are we using AI/ML already?




Hou and Behdinan, 2023



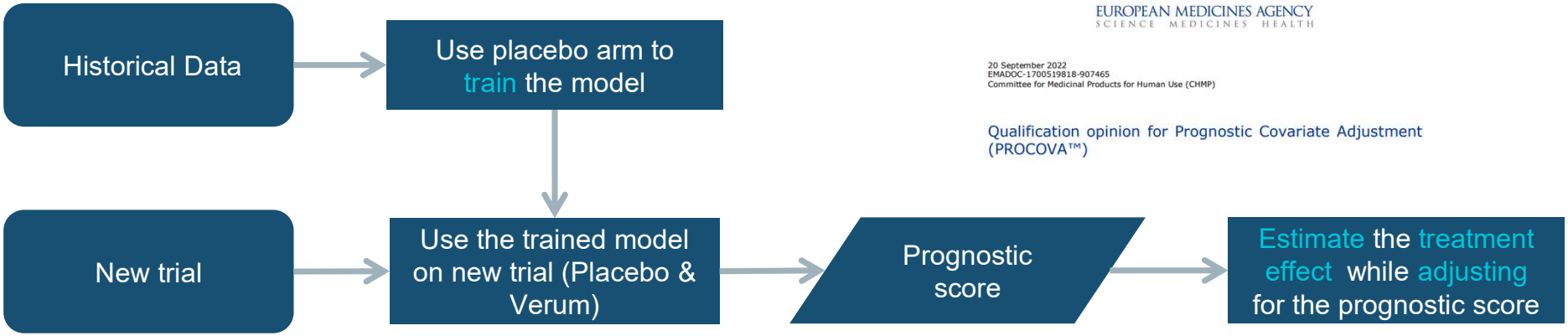




- Dimensionality reduction / information integration of complex digital biomarker data sources such as images and digital health technologies.
- Assessment and quantification of treatment effect heterogeneity instilled by complex biomarker-treatment interactions.
- Data driven hypothesis generation / identification of complex prognostic biomarker signatures for patient stratification.
- ML based prognostic covariate adjustment for increase of effective sample size / treatment effect estimate precision in interventional trials.

# ML based prognostic covariate adjustment

- A variety of approaches exist regarding covariate adjustment and can leverage the potential of ML to capture complex, non-linear relationships between (biomarker) covariates and outcome.
- One is to use historical data to train a prognostic ML model and then to use it to predict complex prognostic scores in a new trial. The score can be used to increase the effective sample size of the trial with respect to treatment effect estimation.



# ML workstream: What have we done so far?

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# Discussion papers on the use of AI/ML in drug development

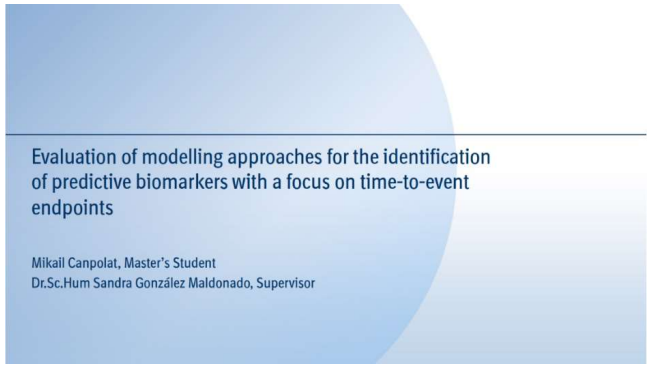
- Release of discussion and reflection papers by FDA and EMA on the use of AI/ML in drug development, with request for feedback
  - FDA: “initial communication […] to promote mutual learning and discussion”
  - EMA: “to reflect on the scientific principles that are relevant for regulatory evaluation”

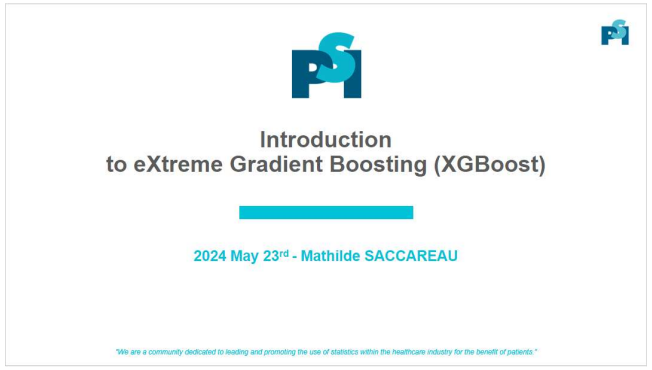
- Consolidated review from the ML workstream and feedback sharing to HAs
  - Comprehensive range of subjects relating to fit-for-purpose application of AI in the drug development life cycle
  - Example of comments:
    - is it acceptable to utilize a black-box prognostic model for PROCOVA?
    - Model performance testing: is prospectively generated data (future calendar time) too strict and might be infeasible in practice?

# Regular series of talk

## Share AI/ML experiences, learnings and expertise through regular meetings




Upcoming talks:

- ❑ Determination of cut-offs using AI
- ❑ Explainable AI
- ❑ Introduction to Neural Networks
- ❑ Deep learning for image analysis
- ❑ Probability of trial success using AI

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# Predictive biomarker identification in a realistic biological setting

- Exploration of treatment effect heterogeneity (TEH) has attracted a lot of attention in clinical trial setting in the last decade
- Increasing number of innovative (AI/ML) approaches is being proposed (Lipkovich et al., 2023) in this context
- Current publications related to TEH and/or predictive BMs identification often consider a simplified problem, while biomarker/biological setting is much more complex ($p >> n$, high-order interactions, non-linear relationships, etc.)
- **Lack of existing framework to simulate realistic (biological) data and evaluation of variable selection properties of methods in identifying / quantifying feature-feature interactions**

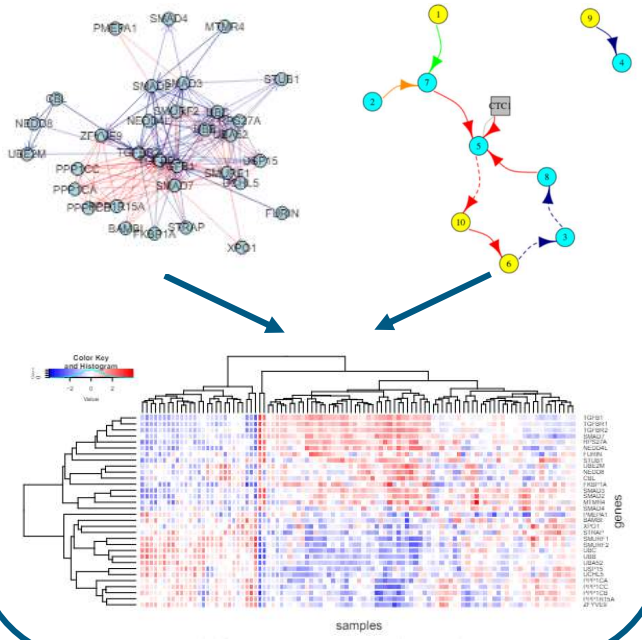
# Hands-on project – detecting TEH in complex biological systems

**Goal**: Identify and/or develop a framework for complex predictive BMs detections by assessing variable selection properties of innovative methods in identifying/quantifying feature-feature interactions

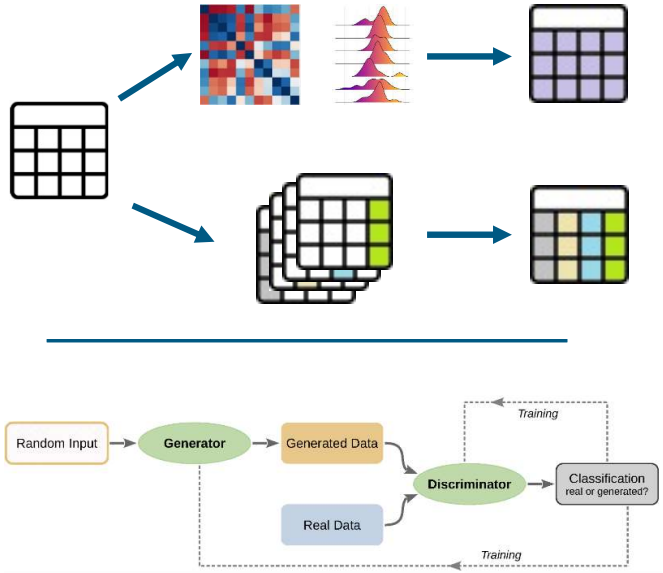
- ❑ Extensive literature review to get a comprehensive view of state-of-the-art TEH detection in complex settings to avoid redundant work
- ❑ Set-up a collaborative GitHub environment for program/code sharing
- ❑ Screen existing frameworks for simulating realistic biological data, evaluate pros/cons
- ❑ Create a short list of initial methods to be benchmarked
- ❑ Initiate a simulation protocol to test and evaluate methods
- ❑ Specify and implement a software framework in a plug and play fashion that ultimately can be used to support planning and analyses of biomarker studies in context of interventional trials

# Simulate data from a complex but also realistic biological setting

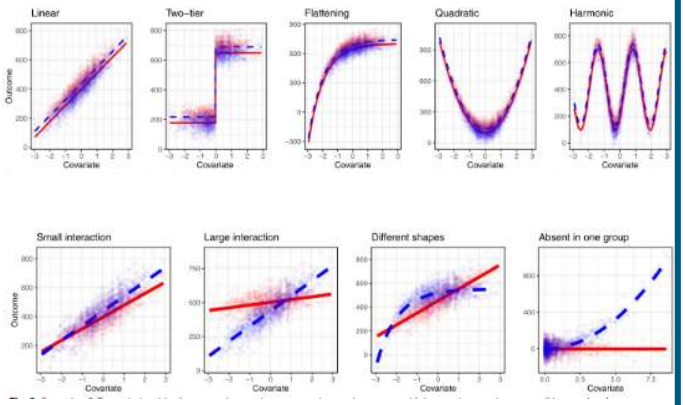
## Simulation based on existing biological networks[1,2]




## Simulation based on existing biological data[3,4]



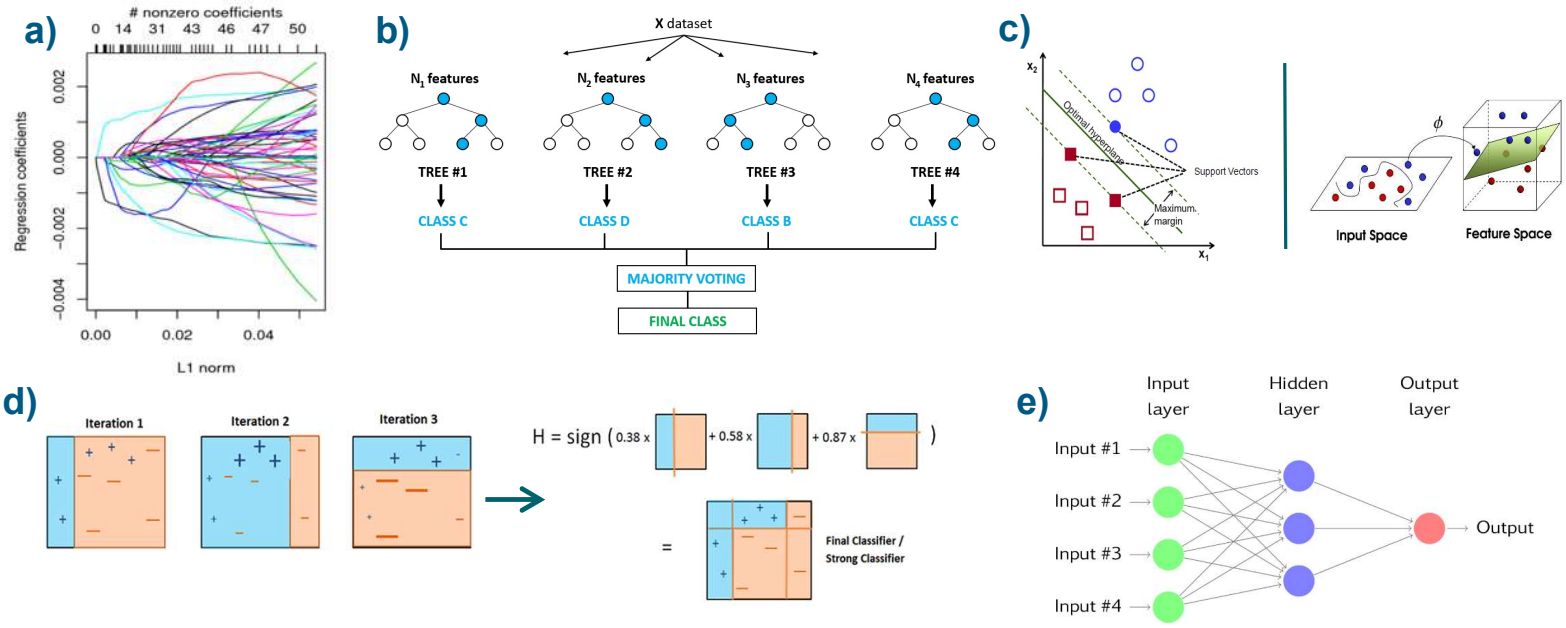

## Simulation based on artificial data generation[5,6]




1. graphsim R package (JOSS, 2020); 2. sismonr R package (Bioinformatics, 2020); 3. Schulz A et al. (BMC Medical Research Methodology, 2017); 4. RGAN R package (2022); 5. Tackney MS et al. (Trials, 2023)

# ML/AI: very attractive for handling outcome prediction tasks

- Penalized regression[a]
- Random forest[b]
- Support vector machine[c]
- Boosting[d]
- Neural network[e]
- …




**Direct application of ML methods may not fit with TEH and predictive BM detection**

# Methods for predictive biomarker detection

Growing literature of ML methods for TEH and predictive biomarker detection!

Direct ML methods for estimating conditional average treatment effects (CATE)

- Penalized regression including prognostic and predictive effects
- Causal forest and its bayesian version
- Causal boosting and its bayesian version

Meta-learners: to decompose CATE into several regression sub-problems

- Conditional mean regression methods: S-, T- learners
- Pseudo-outcome methods: X-, DR-, R- learners

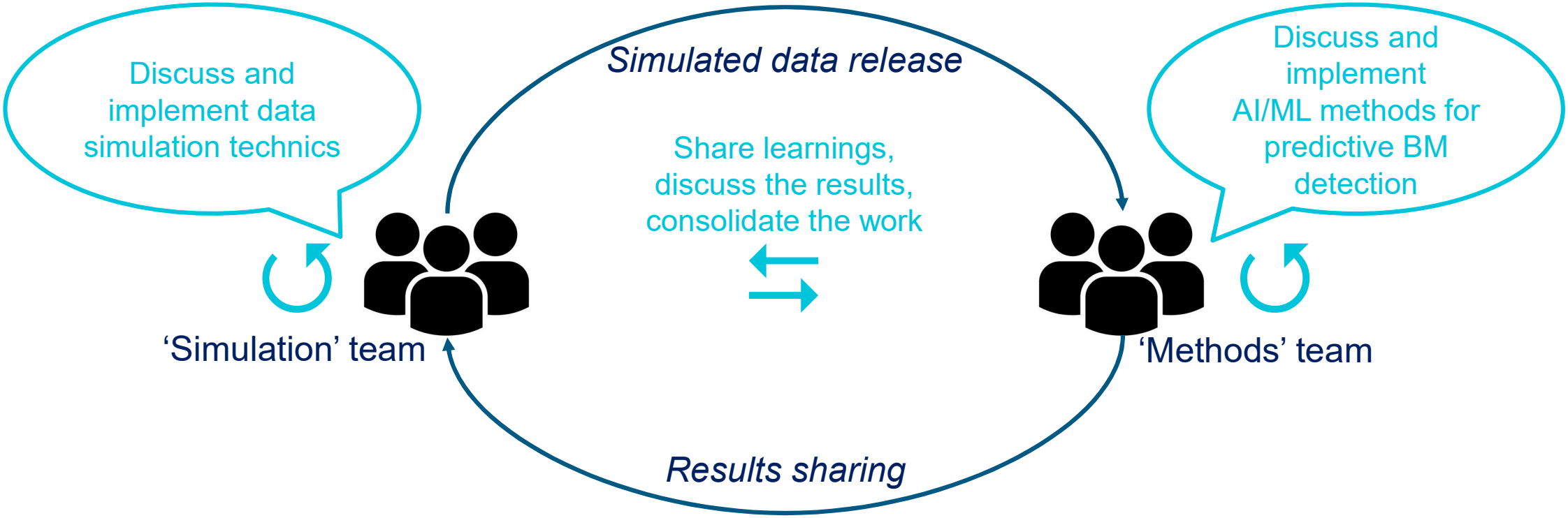
Ensemble methods (stacking): to combine results from different ML methods

- SuperLearner

**Need to explore and evaluate the behaviors of these methods in a realistic (and complex) biological setting**

# A ‘Kaggle’ initiative

To continuously create synergies and learnings, a gentle competition is being set-up:




*“We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients.”*

# Conclusion and next steps

- ‘AI/ML’ and ‘biomarkers/precision medicine’ are two hot topics!
- Biomarker SIG - ML workstream: a great mixture of people to be creative and get inspiring perspectives, feedbacks, ideas etc.

***Interested to join?*** Please contact us:
Karl Köchert (karl.koechert@bayer.com)
Nils Ternès (nils.ternes@sanofi.com)


*“We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients.”*

# Digital biomarkers: the essential guide for statisticians

Marzia Antonella Scelsi, PhD
Roche Products Ltd, Welwyn Garden City, UK

June 17, 2024

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*


# Digital biomarkers: the essential guide for statisticians

PSI Conference 2024 - Biomarkers ESIG Session

Marzia Antonella Scelsi, PhD
Senior Statistician, Roche Products Ltd, Welwyn Garden City, UK

17th June 2024 | Public

# Disclosures

Marzia A. Scelsi is a full-time employee of Roche Products Ltd.

The views and opinions expressed in this presentation are solely those of the author and do not necessarily reflect the official policy or position of F. Hoffmann-La Roche AG.

# What will you learn from this talk?

- Digital technologies are getting more and more ubiquitous in the healthcare sector and in drug development
- To turn them into useful tools (e.g., biomarkers), their quantitative properties have to be investigated rigorously
- You will learn about **four classes of statistical methods to study these properties**:
  - Test-retest reliability
  - MDC and MCID
  - Choice of aggregation window
  - Longitudinal modelling

# FDA definition of digital biomarker

*"A characteristic or set of characteristics, collected from digital health technologies (DHT), that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions."*


- Vasudevan, Srikanth, et al. "Digital biomarkers: convergence of digital health technologies and biomarkers." NPJ digital medicine 5.1 (2022): 36.
- US Food and Drug Administration. "Patient-Focused Drug Development: Collecting Comprehensive and Representative Input." Final guidance document www.fda.gov/media/139088/download (2020)

# The promise of digital biomarkers in drug development and healthcare in general

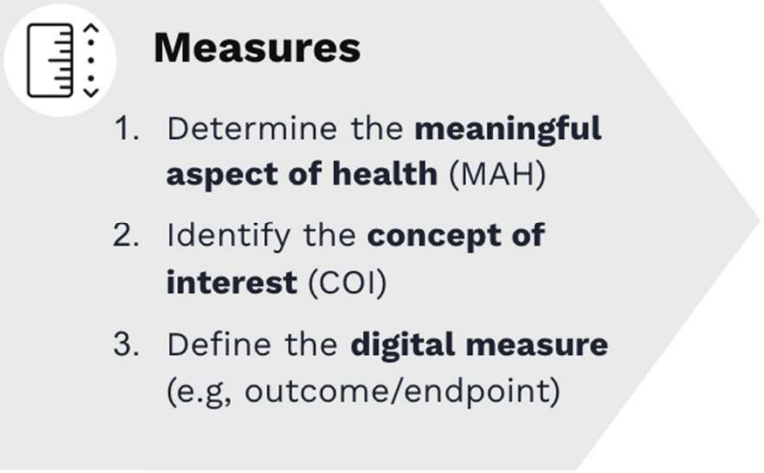
In the **healthcare sector** in general:

- Measure health outcomes with high frequency, in a sensitive and non-invasive way
- Enable detection of early signs of disease
- Capture the daily fluctuating nature of certain diseases
- Transmit patient data to healthcare providers in real-time, increasing quality of care

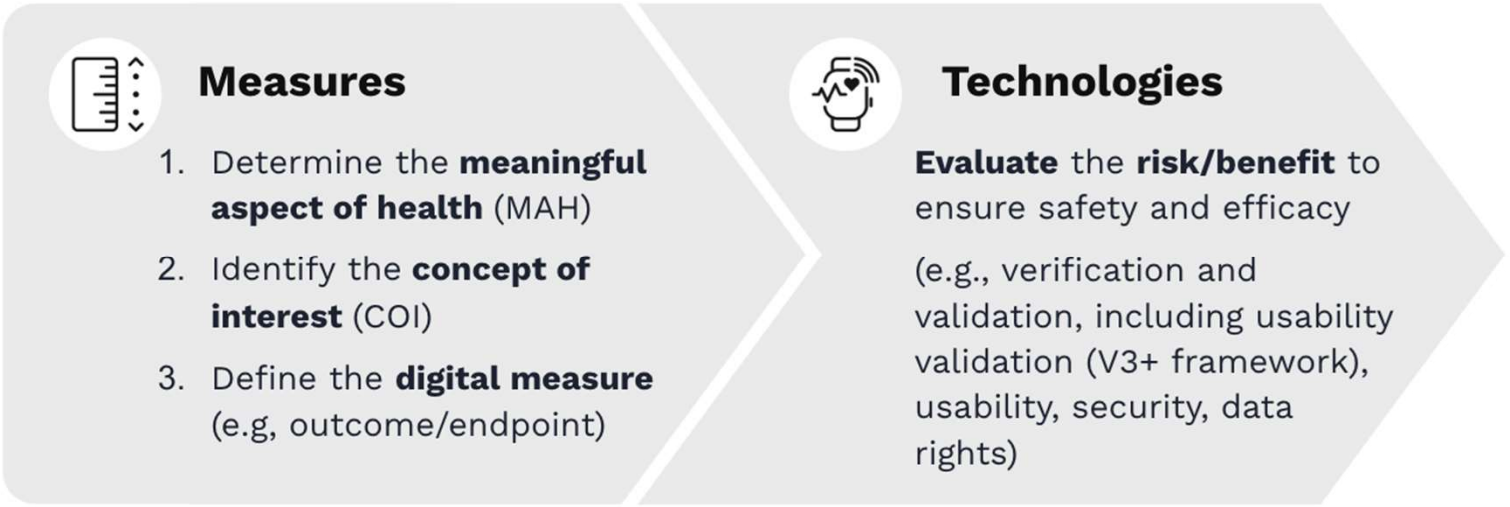
In **drug development** in particular:

- Quantitative, objective outcome measures that are more sensitive to change than established COAs
- Potential to characterize domains of disease usually understudied but relevant to patients (e.g., fatigue in MS)
- Potential to run shorter and/or smaller clinical trials
- Potential to use for internal strategic decision making (e.g., increase PTS, derisk ph3 programs)

# The DiME Playbook guides validation and adoption of DHTs in drug development



# The DiME Playbook guides validation and adoption of DHTs in drug development



# The DiME Playbook guides validation and adoption of DHTs in drug development

## Measures

1. Determine the **meaningful aspect of health** (MAH)
2. Identify the **concept of interest** (COI)
3. Define the **digital measure** (e.g, outcome/endpoint)

## Technologies

**Evaluate** the **risk/benefit** to ensure safety and efficacy

(e.g., verification and validation, including usability validation (V3+ framework), usability, security, data rights)

## Operations

Plan for the **jobs to be done** during deployment (e.g., purchasing, distribution, monitoring, data analysis)

PRO TIP

### Order Matters

You can avoid common, order-related pain points if you follow sequentially the steps we outline in *The Playbook*!

# The DiME Playbook guides validation and adoption of DHTs in drug development

## Measures

1. Determine the **meaningful aspect of health** (MAH)
2. Identify the **concept of interest** (COI)
3. Define the **digital measure** (e.g, outcome/endpoint)

## Technologies

**Evaluate** the **risk/benefit** to ensure safety and efficacy

(e.g., verification and validation, including usability validation (V3+ framework), usability, security, data rights)

## Operations

Plan for the **jobs to be done** during deployment (e.g., purchasing, distribution, monitoring, data analysis)

PRO TIP

Order Matters

You can avoid common, order-related pain points if you follow sequentially the steps we outline in *The Playbook*!

***Statisticians are key stakeholders and can provide crucial input to teams at all stages of the adoption process!***

# Statistical Methods for Digital Biomarkers

*(a non-exhaustive guide)*

# How to Quantify Reliability of a Digital Biomarker

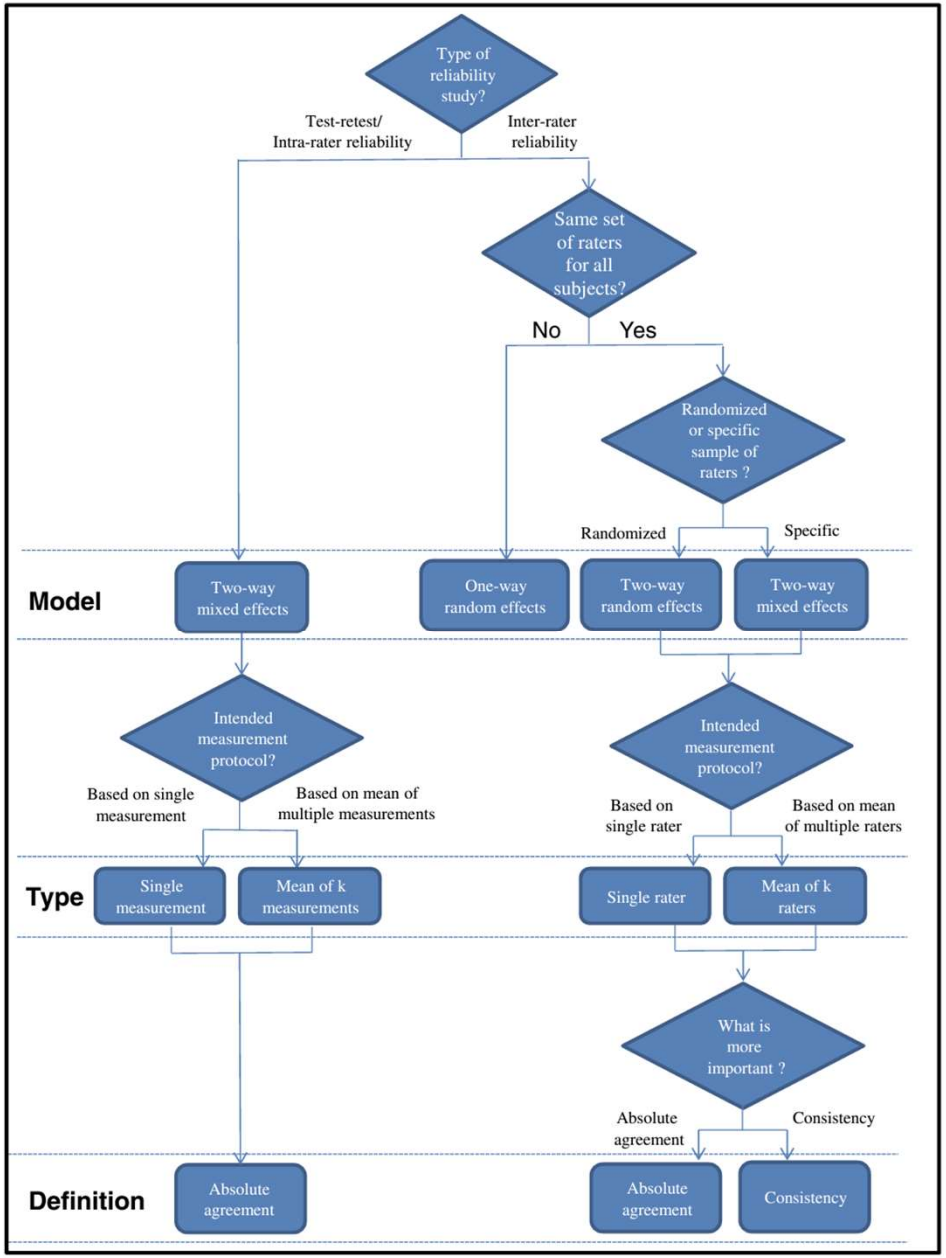
# Intraclass Correlation Coefficients

## How to choose the most appropriate?

- **What**: quantify how reliable/repeatable is your digital biomarker measurement

- **How**: conduct a test-retest reliability study
  - Measure your digital biomarker twice, in the same experimental conditions, in a short interval of time
  - Calculate ICC

- Sounds easy, but… there are **10 different versions of ICC available**[1] differing in terms of model, type, and definition! Which one to choose?




(1) Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine, 15(2), 155-163.

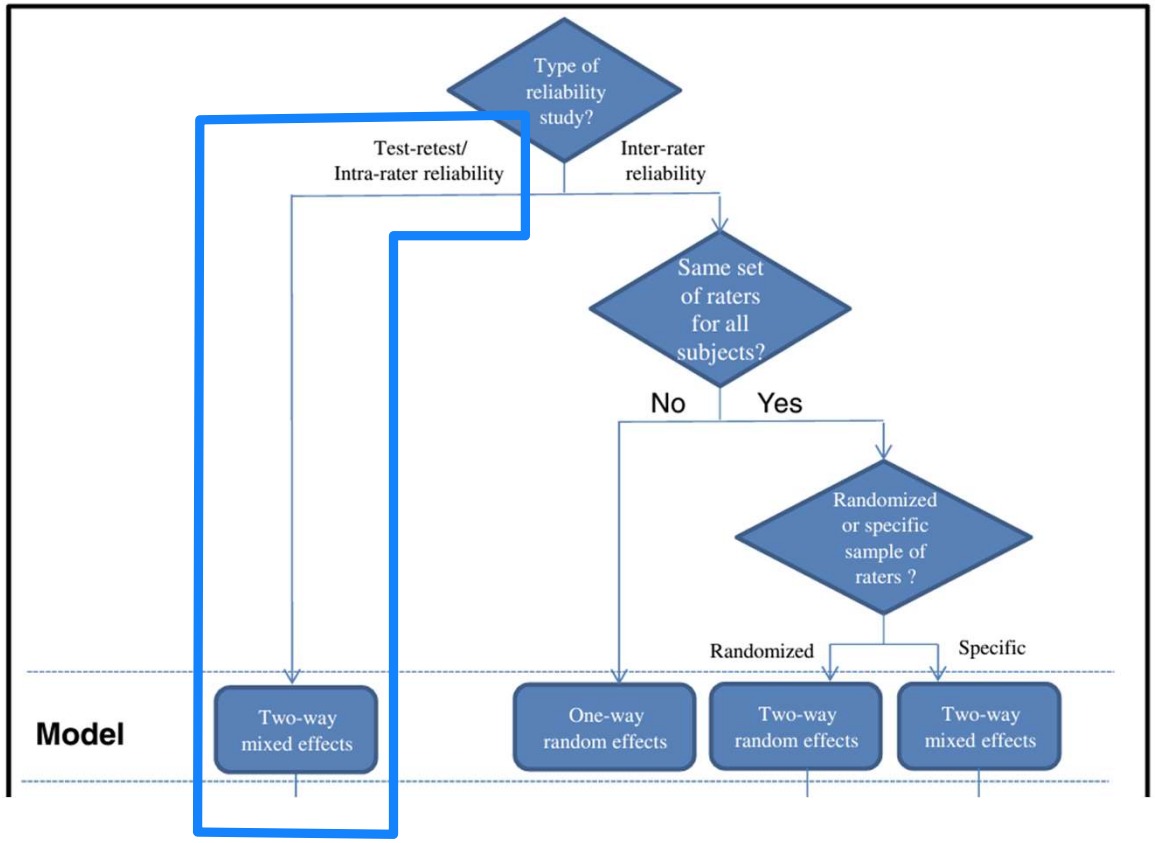
# Intraclass Correlation Coefficients

## Model, type, definition

### MODEL

*"FDA recommends that, in most cases, intraclass correlation coefficients be calculated using **absolute agreement**, **two-way mixed-effects model with the time as a fixed effect** (McGraw and Wong 1996; Shrout and Fleiss 1979), as suggested by Shrout and Fleiss (1979) and Qin et al. (2019)."*

*"**2-way mixed-effects model** should be used in test-retest reliability study because repeated measurements cannot be regarded as randomized samples."* (Koo and Li 2016)

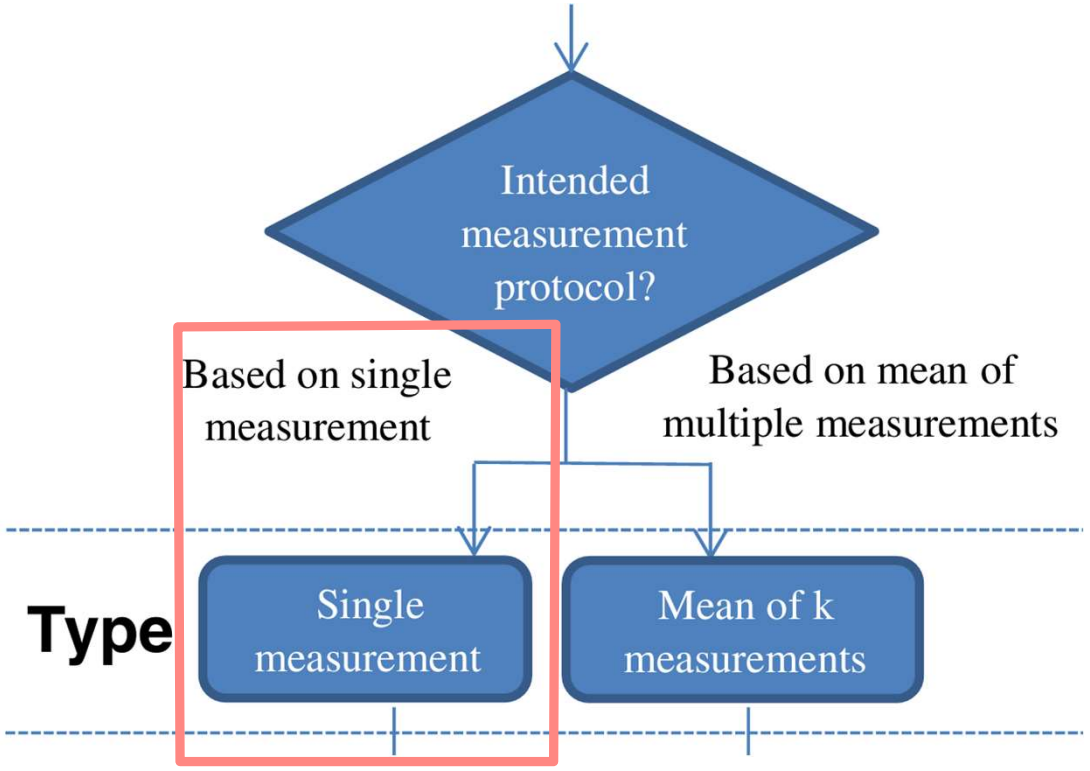


# Intraclass Correlation Coefficients

## Model, type, definition

**TYPE**

- **Single measurement** or **mean of k measurements**: depends on intended measurement protocol in the actual application
- **Usually single measurement is sufficient**, because digital features are already aggregated (mean/median) over a certain time interval



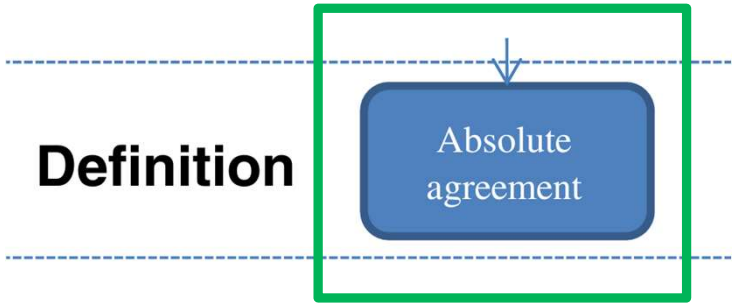
# Intraclass Correlation Coefficients

## Model, type, definition

**DEFINITION**

- **Absolute agreement**: y(t2) = y(t1)
- **Consistency**: y(t2) = y(t1) + c

*"**absolute agreement** definition should always be chosen for [...] test-retest [...] reliability studies because measurements would be meaningless if there is no agreement between repeated measurements."* (Koo and Li 2016)



# ICC implementation in R

## The ICC() function in the *psych* package - Materials courtesy of Giuseppe Palermo

✅ As opposed to other implementations, which use `'aov'` as the backbone, it offers the choice to use `'lmer'`

✅ It can handle missing data (unbalanced design)

✅ The `ICC()` function calculates all flavours of ICC for you →

simply choose **ICC2** which corresponds to **ICC(A, 1)** [agreement] in the notation of McGraw & Wong

| From ICC() documentation (psych package) | Formulas from McGraw & Wong |
|---|---|
| ICC1: Each target is rated by a different judge and the judges are selected at random.<br>`ICC(1,1) = \rho_{1,1} = \frac{\sigma^2_r}{\sigma^2_r + \sigma^2_w}`<br>(This is a one-way ANOVA fixed effects model and is found by (MSB- MSW)/(MSB+ (nr-1)*MSW)) | |
| ICC2: A random sample of k judges rate each target. The measure is one of absolute agreement in the ratings.<br>`ICC(2,1) = \rho_{2,1} = \frac{\sigma^2_r}{\sigma^2_r + \sigma^2_c +\sigma^2_{rc} + \sigma^2_e}`<br>Found as (MSB- MSE)/(MSB + (nr-1)*MSE + nr*(MSJ-MSE)/nc) | $\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$ ICC(A,1) |
| ICC3: A fixed set of k judges rate each target. There is no generalization to a larger population of judges.<br>`ICC(3,1) = \rho_{3,1} = \frac{\sigma^2_r}{\sigma^2_r + \sigma^2_c + \sigma^2_e}`<br>(MSB - MSE)/(MSB+ (nr-1)*MSE) | $\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$ ICC(C,1) |

# Minimum Detectable Change (MDC) and Minimal Clinically Important Difference (MCID)

# Minimum Detectable Change (MDC) and Minimal Clinically Important Difference (MCID)

For every health-related measurement (digital or otherwise) there are two key notions to evaluate results:

| Minimum Detectable Change | Minimal Clinically Important Difference |
|---|---|
| The smallest difference sufficiently large to be sure that it is not noise | *"The smallest difference [...] that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's health management"* (Guyatt et al., 2022) |

# Methods to calculate MDC

## Minimum Detectable Change

Three variants:

- Preferred **ANOVA** computation (1):

  MDC = 1.96 * √2 * SEM

  SEM = Standard Error of Measurement = √WMS, where WMS is the mean square error term from ANOVA

- **Standard** computation:

  1.96 * sd(Difference between 2 test occasions) = 1.96 * √2 * ME

  ME = Measurement Error = sd(Difference between 2 test occasions)/√2

- **ICC** computation:

  SEM = sd(measure) * √(1-ICC)

(1) Reliability, minimal detectable change and responsiveness to change: Indicators to select the best method to measure sedentary behaviour in older adults in different study designs - PMC

# Methods to calculate MCID

In text quoted from FDA guidance: MSD = Meaningful Score Difference; COA = Clinical Outcome Assessment
(2) PATIENT-FOCUSED DRUG DEVELOPMENT GUIDANCE PUBLIC WORKSHOP - Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making - Workshop Date: December 6, 2019

| Distribution-Based Methods | Anchor-Based Methods |
|---|---|
| ■ **Not recommended by FDA**:<br><br>*"Distribution-based methods (e.g., effect sizes, certain proportions of the standard deviation and/or standard error of measurement) do not directly consider the patient voice, and as such, are insufficient to serve as the sole basis for identifying an MSD. Distribution-based methods can provide helpful information about measurement variability." (2)* | ■ **Recommended by FDA**:<br><br>*"An anchor is some external variable, not derived from the COA whose scores require interpretation, for which meaningful differences are directly interpretable or already known. Meaningful differences on the anchor can then be mapped onto differences in terms of the COA scores." (2)*<br><br>To establish a threshold(s) or a range of thresholds:<br>■ Mean (longitudinal) difference in subgroup of patients with event of interest (e.g. progression)<br>■ Empirical cumulative distribution function (eCDF) and PDF curves<br>To evaluate the performance of the chosen threshold(s):<br>■ ROC curves |

# Example of MCID calculation

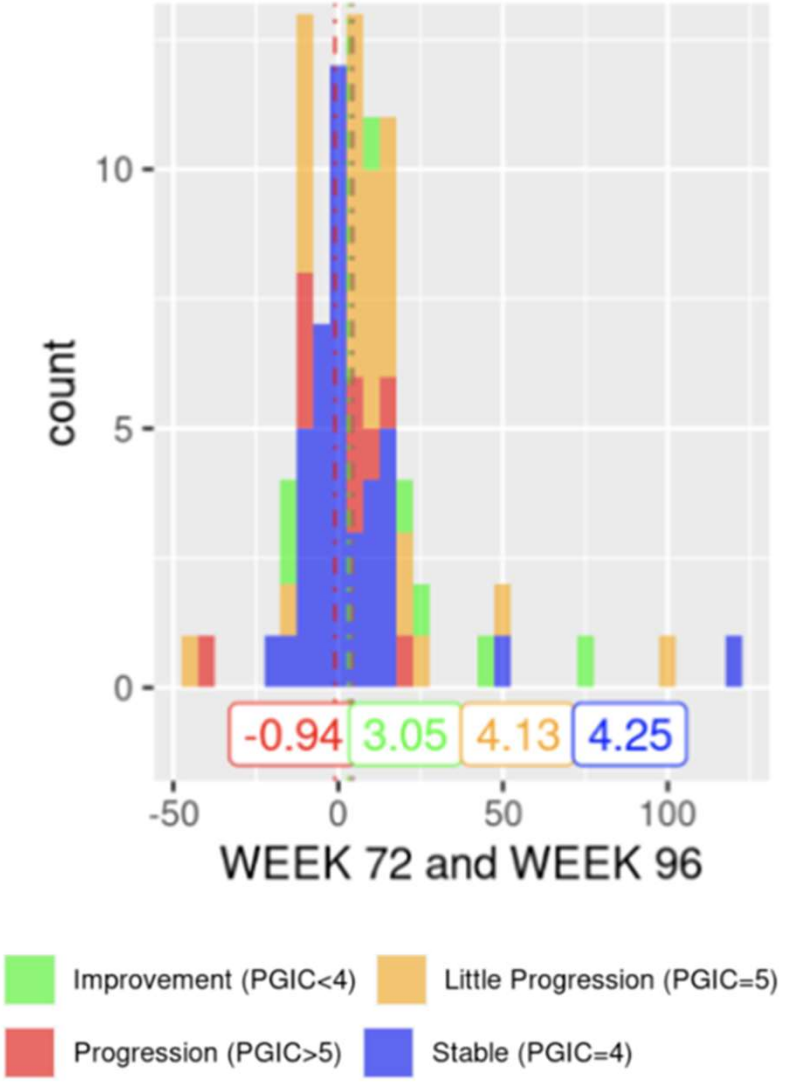
Case study: **Floodlight MS** smartphone application

Materials courtesy of Stanislas Hubeaux

- **Aim**: MCID computation for the Floodlight feature 5UTT Average Turning Speed in CONSONANCE trial

- **Key Message**
  - Strangely, mean change value for the progressors (red group) is lower than the ones for improvers (green group) and and little progressors (orange group)

  - No conclusive MCID could be calculated for the 5UTT Average Turning Speed




PGIC = Patient Global Impression of Change: questionnaire scored on 7 levels

# How to Choose Your Aggregation Window

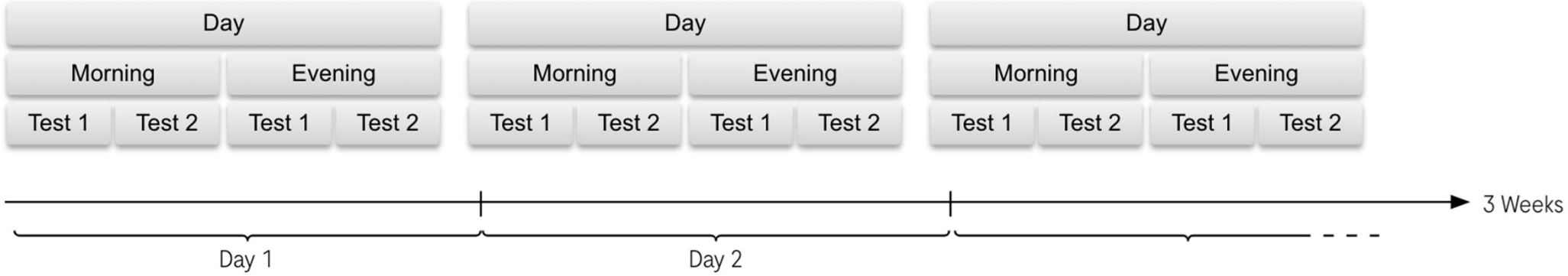
# Why aggregate and how to do it?

## Motivations for a “precision” study

- Aggregation of repeated measurements enables to reduce “noise” from human and device variability
- AIM: **determine the number of repeated measurements needed** to be aggregated in order to have acceptable error
- HOW: run a **precision study** to:
  - Evaluate the variability of repeated measurements
  - Identify sources of variability in repeated measurements:
    - Test-Test: Sensor noise and user-phone interactions
    - Morning-Evening: Circadian rhythm and environment
    - Day-Day: Biological variability and environment

# Full Factorial Design for Precision Study

## Application to Floodlight MS - Materials courtesy of Xavier Denos and Stanislas Hubeaux




- Study design and analysis follows CLSI guidance EP5-A3: "*Evaluation of Precision of Quantitative Measurement Procedures*"

- Test run executed 4 times a day for 3 weeks (15 healthy volunteers):
  - Twice in the morning before 12pm; twice in the afternoon after 2pm
  - Daily alternation between right and left hand

# Variance Component Analysis (VCA) for Precision Study

## Application to Floodlight MS - Materials courtesy of Xavier Denos and Stanislas Hubeaux

- Fraser (2001) defines mathematically the number of replicate measurements needed to detect MCID as:

$$N = 2\,Z^2 \left(\frac{SD}{MCID}\right)^2$$

Where:

- SD = standard deviation from VCA
- MCID assumed known or at least agreed upon (if not, MDC can be used)
- Z defines expected False Positive Rate:
  - One-sided change (only worsening >MCID), 10% FPR: Z = 1.28

- `remlVCA()` function in R package VCA

# Variance Component Analysis (VCA) for Precision Study

## Application to Floodlight MS - Materials courtesy of Xavier Denos and Stanislas Hubeaux

| Assessment in Precision Study | Total Variability* | # Replicate Measurements N |
|---|---|---|
| IPS – Total Score (SD) | 5.3 | 6 |
| IPS Baseline – Total Score (SD) | 3.5 | 3 |
| 5UTT – Average Turn Speed (CV) | 14.9% | 2 |
| PT01 - Total number of successful pinches (CV) | 19.2% | 4 |
| DAS – Overall Mean Celerity (CV) | 22.9% | 5 |

* Total variability = test-retest + day-day

$$N = 2\,Z^2 \left(\frac{SD}{MCID}\right)^2$$

**Assumed MCID values**:

- 4 points for IPS and IPS baseline
- 20% for 5UTT, PT, DAS

# How to Test for Treatment Effect on a Digital Biomarker

# Longitudinal modelling of digital biomarkers

- **Scenario**:
  - You have deployed a DHT in an interventional trial and measured digital biomarkers over time
  - You want to model them in a flexible yet robust way and **estimate treatment effect**
  - Bonus points: you have a **time-varying covariate** that strongly affects the digital outcome (e.g., symptomatic rescue medication in Parkinson's disease) and want to regress out its effect

- What modelling choices do you have?

# Longitudinal modelling of digital biomarkers

- **Scenario**:
  - You have deployed a DHT in an interventional trial and measured digital biomarkers over time
  - You want to model them in a flexible yet robust way and **estimate treatment effect**
  - Bonus points: you have a **time-varying covariate** that strongly affects the digital outcome (e.g., symptomatic rescue medication in Parkinson's disease) and want to regress out its effect

- What modelling choices do you have?

01 **Linear Mixed Effect Model (LME)**

✅ Parsimonious, powerful, interpretable in terms of progression rates
❌ Relies heavily on assumption of linearity
❌ Cannot handle time-varying covariates

# Longitudinal modelling of digital biomarkers

- **Scenario**:
  - You have deployed a DHT in an interventional trial and measured digital biomarkers over time
  - You want to model them in a flexible yet robust way and **estimate treatment effect**
  - Bonus points: you have a **time-varying covariate** that strongly affects the digital outcome (e.g., symptomatic rescue medication in Parkinson's disease) and want to regress out its effect

- What modelling choices do you have?

| | | |
|---|---|---|
| 01 | Linear Mixed Effect Model (LME) | ✅ Parsimonious, powerful, interpretable in terms of progression rates<br>❌ Relies heavily on assumption of linearity<br>❌ Cannot handle time-varying covariates |
| 02 | Mixed Model for Repeated Measures (MMRM) | ✅ Maximum flexibility as no functional form is assumed<br>✅ Can handle time-varying covariates<br>❌ Many parameters to be estimated may lead to convergence issues<br>❌ Estimates may swing a lot over time |

# Longitudinal modelling of digital biomarkers

- **Scenario**:
  - You have deployed a DHT in an interventional trial and measured digital biomarkers over time
  - You want to model them in a flexible yet robust way and **estimate treatment effect**
  - Bonus points: you have a **time-varying covariate** that strongly affects the digital outcome (e.g., symptomatic rescue medication in Parkinson's disease) and want to regress out its effect

- What modelling choices do you have?

| | | |
|---|---|---|
| 01 | Linear Mixed Effect Model (LME) | ✅ Parsimonious, powerful, interpretable in terms of progression rates<br>❌ Relies heavily on assumption of linearity<br>❌ Cannot handle time-varying covariates |
| 02 | Mixed Model for Repeated Measures (MMRM) | ✅ Maximum flexibility as no functional form is assumed<br>✅ Can handle time-varying covariates<br>❌ Many parameters to be estimated may lead to convergence issues<br>❌ Estimates may swing a lot over time |
| 03 | Generalised Additive Mixed Model (GAMM) | ✅ High flexibility as no functional form is assumed, but **SMOOTHNESS is enforced**<br>✅ Can handle time-varying covariates<br>✅ Fewer parameters than MMRM mean better convergence |

# Generalised Additive Mixed Models (GAMM)

A Generalized Additive Mixed Model (GAMM) is a Generalized Linear Mixed Model (GLMM) in which part of the linear predictor is specified in terms of smooth functions of covariates (Wood, 2017).

$$\Delta(\text{dBM}) = X\cdot\beta + f_1(C) + f_2(t \mid \text{ARM} = \text{Placebo}) + f_3(t \mid \text{ARM} = \text{Active}) + Zb + \varepsilon$$

Change from baseline in dBM

Design matrix; includes intercept, stratification factors, and baseline dBM value

**Smooth function of time-varying covariate**

**TWO smooth functions of time, one for each arm**

Random intercept and slope per subject

Residual

The smooth functions are represented as linear combinations of a suitable set of basis functions (**thin plate splines**);
the coefficients of these linear combinations are estimated with a degree of penalization to ensure control over the level of "wiggliness" of the smooths and to avoid overfitting.

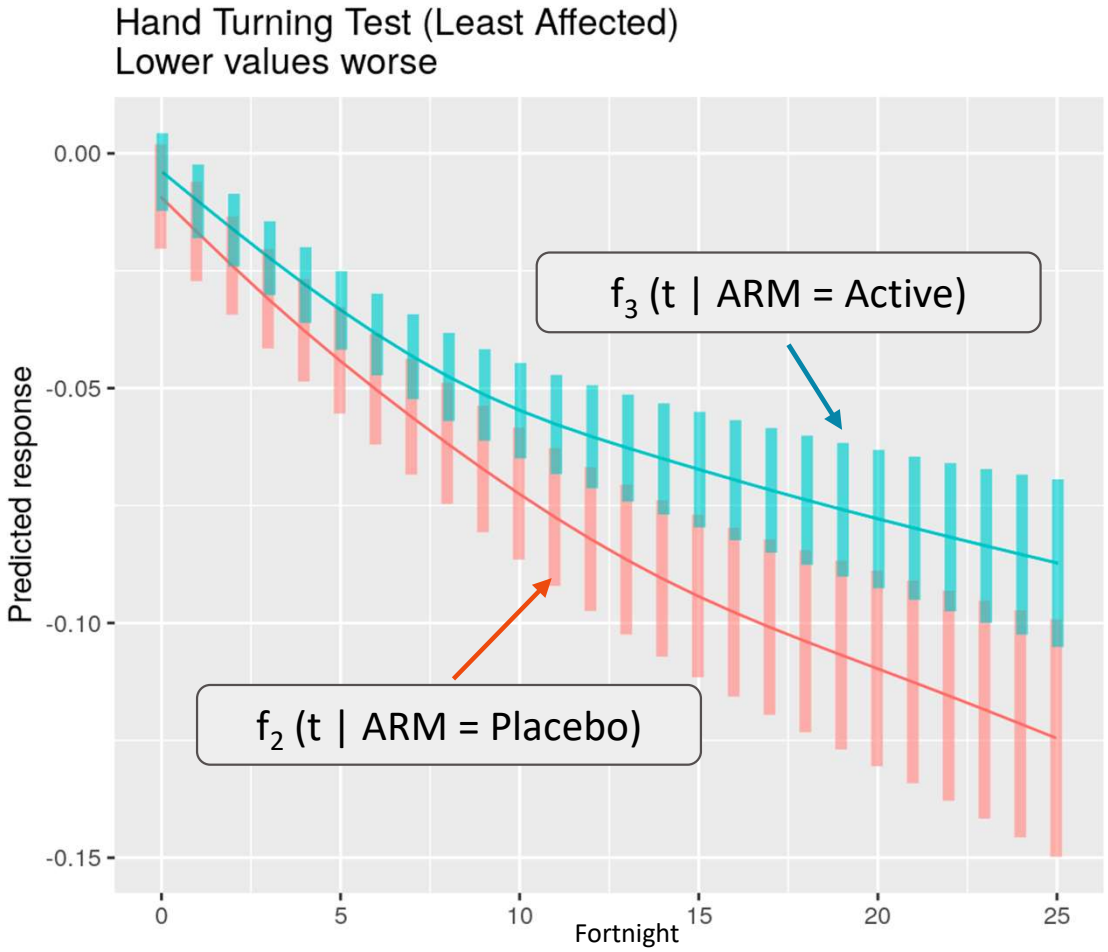
Wood, S.N. (2017). "Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.)". *Chapman and Hall/CRC*. https://doi.org/10.1201/9781315370279

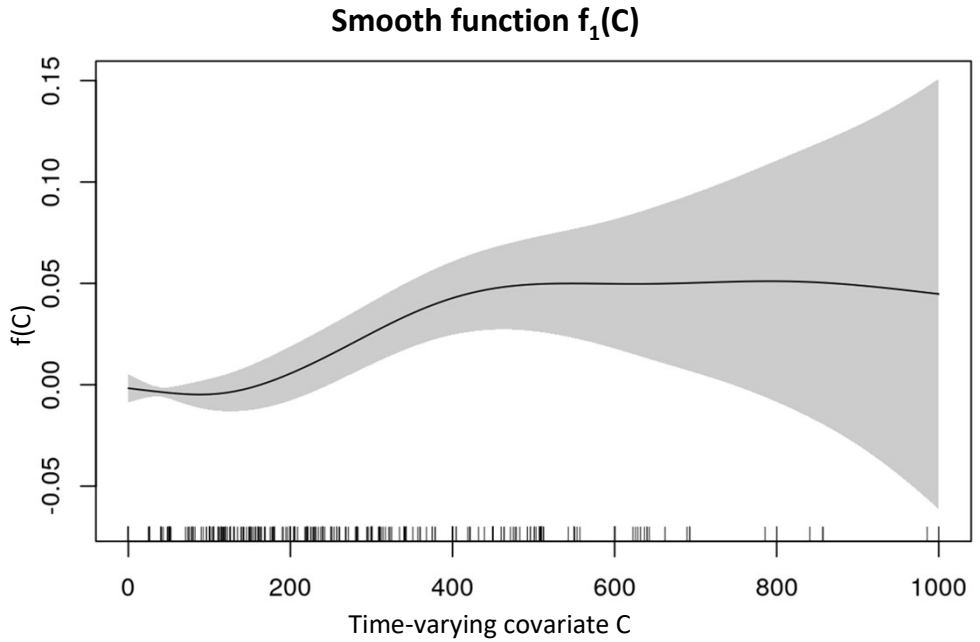
# Example of GAMM results

## Case study: **Roche PD Mobile App v2**

## PASADENA trial in early Parkinson's

Fitted with `gamm()` function in R package [mgcv]




| Contrast | Fortnight | Estimate | SE | p-value |
|---|---|---|---|---|
| Active - Placebo | 6 | 0.01162 | 0.01078 | 0.28 |
| Active - Placebo | 11 | 0.0197 | 0.01361 | 0.14 |
| Active - Placebo | 25 | 0.03725 | 0.02379 | 0.12 |

# Conclusions

- We have covered, in logical order, statistical methods that can be used to investigate the properties of digital biomarkers;
- However, field still evolving means no general consensus in the industry nor detailed guidance from regulators.

- Aspects that will always be important for adoption of digital biomarkers in clinical trials:
  a. Establish if measurements are repeatable (ICC);
  b. Determine what is "real" change (MDC) and what is clinically meaningful change (MCID);
  c. How to choose an aggregation window to reduce noise (VCA);
  d. How to model clinical trial data and test for the presence of a treatment effect (GAMM).

- The voice of the **statistician remains critical** to ensure that digital biomarkers are developed following all the logical steps in the right order from inception to deployment, with the necessary rigour and backed by quantitative evidence, so that the **tool is fit for drug-development purposes**.

# Acknowledgments


- Stanislas Hubeaux
- Xavier Denos
- Giuseppe Palermo
- Bernhard Fehlmann
- Daria Rukina
- Markus Abt
- Laurent Essioux
- … and many more!

Doing now what patients need next

# Time for a fireside chat!

## Biomarkers European Special Interest Group

June 17, 2024

*"We are a community dedicated to leading and promoting the use of statistics within the healthcare industry for the benefit of patients."*

# Fireside chat with 4 biomarker experts!

**Gaëlle Saint-Hilary**
CEO
&
Statistical Methodologist

Saryga

**Juan José Abellán Andrés**
Biostatistics and real-world evidence senior specialist

European Medicines Agency

**Kostas Sechidis**
Data Science Associate Director

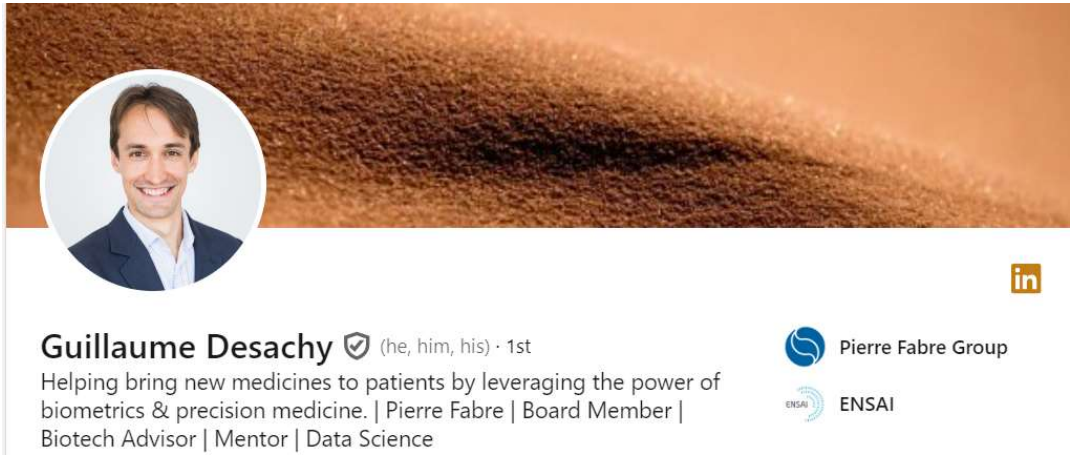
Novartis

**Marzia Antonella Scelsi**
Senior Statistical Scientist

Roche

# Biomarkers ESIG Co-Leads


https://www.linkedin.com/in/dr-nicole-kraemer/




https://www.linkedin.com/in/guillaume-desachy/