Development of a smartphone based monitoring tool for people with Multiple Sclerosis
Challenges and Opportunities

Stanislas Hubeaux
Principal Statistical Scientist

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

What is Floodlight™?

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

Challenges with current assessments in MS & Floodlight™ ambition

Clinical trial endpoints

- MS is characterized by phenotypic heterogeneity
- EDSS is heavily weighted on lower limb
- Current outcome measures have limitations in precision and sensitivity to change
- Outcome measures that capture improvement are not available

Covered by Fabian Model in Session 1

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

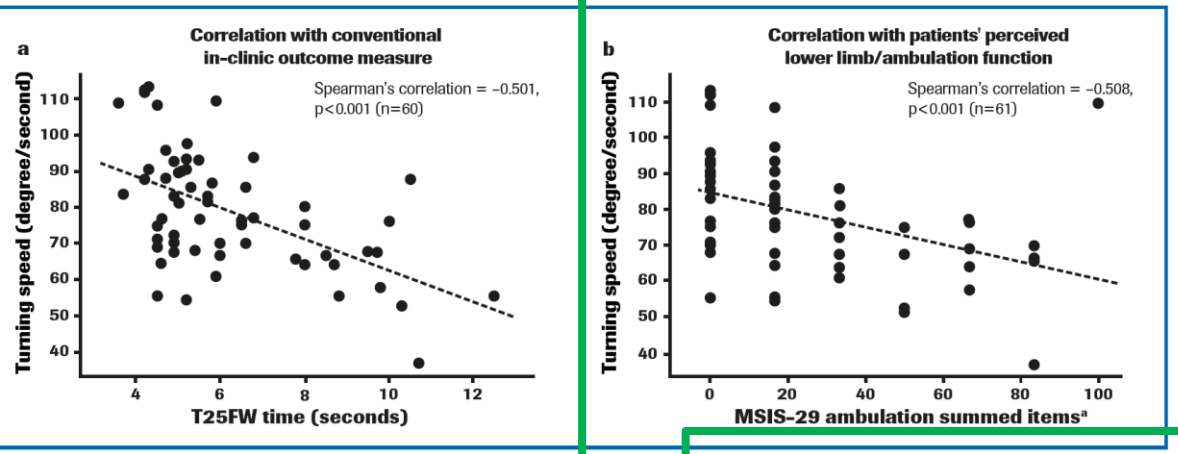
Our ambition: qualify digital measures as regulatory-grade label-enabling endpoints and make them available as measurement tools in clinical practice

Endpoint Qualification Procedure
FDA (CDER) & EMA

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

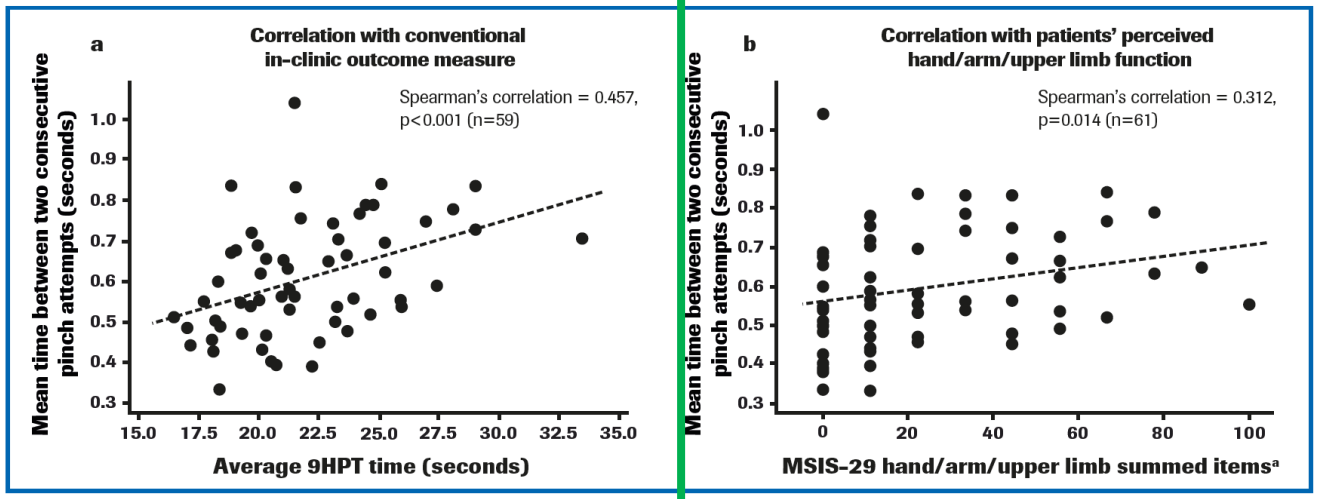
Floodlight™: Where are we today? Correlation

Figure 5. U-turn speed measured with 5UTT correlates with T25FW and patients' perceived lower limb and ambulation function



[a]Items 4 and 5 of the MSIS-29 summarized and rescaled to [0, 100].
5UTT, Five-U-Turn Test; MSIS, Multiple Sclerosis Impact Scale; T25FW, Timed 25-Foot Wal

Figure 3. Mean time between two consecutive pinch attempts correlates with average 9HPT and patients' perceived hand/arm/upper limb function

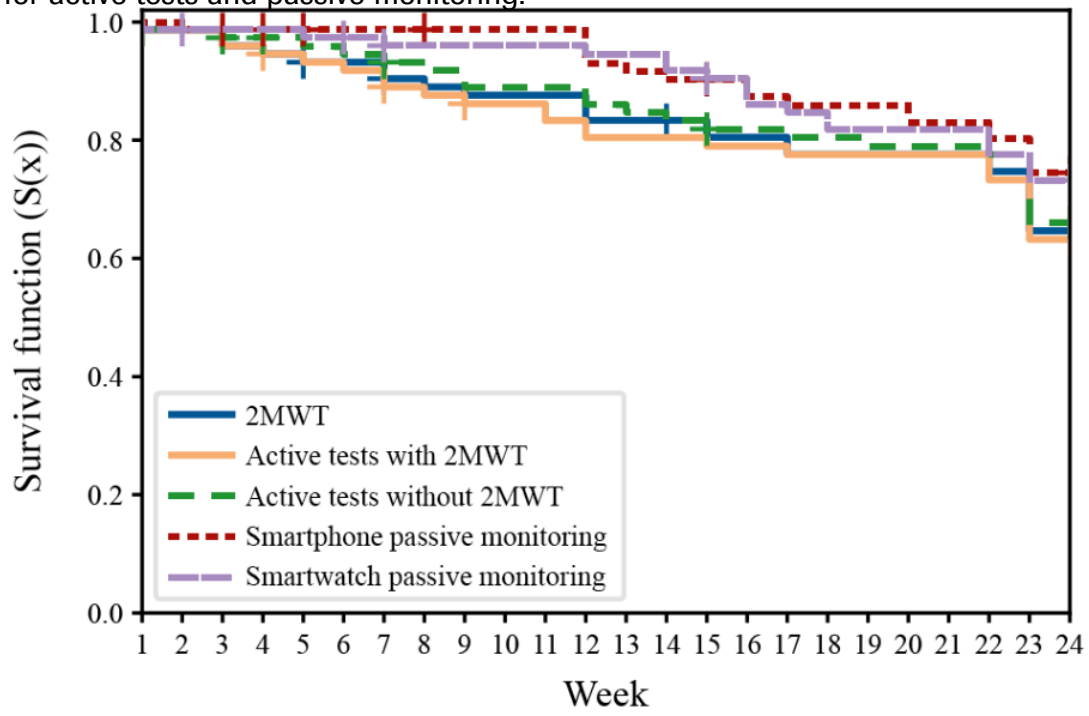


[a]Items 2, 6 and 15 of the MSIS-29 summarized and rescaled to [0, 100].
9HPT, 9-Hole Peg Test; MSIS, Multiple Sclerosis Impact Scale.

Presented at AAN 2019: *FLOODLIGHT: Smartphone-Based Self-Monitoring Is Accepted by Patients and Provides Meaningful, Continuous Digital Outcomes Augmenting Conventional In-Clinic Multiple Sclerosis Measures*

Floodlight™: Where are we today? Adherence & Conclusions

Adherence of people with multiple sclerosis to active tests and passive monitoring. The abandoning event was defined as the last week in which the participant was adherent according to the definitions for active tests and passive monitoring.



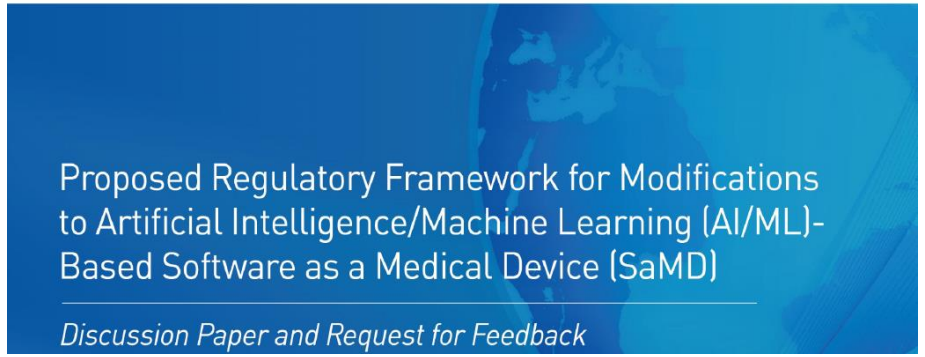
# of patients																							
2MWT	75	72	72	70	69	67	67	64	63	62	62	62	59	59	58	56	56	54	54	54	54	54	52	45
Active tests with 2MWT	75	72	72	70	68	67	66	63	62	60	60	58	56	56	56	55	55	54	54	54	54	54	51	44
Active tests without 2MWT	76	74	74	72	70	69	68	66	65	63	63	63	61	60	59	57	57	56	56	55	55	55	54	46
Smartphone passive monitoring	76	76	75	74	73	71	71	71	70	70	70	70	66	65	64	63	61	60	60	60	58	58	56	52
Smartwatch passive monitoring	76	75	74	74	74	73	72	69	69	69	69	69	68	68	66	63	60	59	57	57	57	57	54	51

- FLOODLIGHT™ outcomes correlate with in-clinic outcome measures of MS disability
- Patients are highly engaged and satisfied with smartphone-based self-assessments
- FLOODLIGHT™ outcomes may represent a promising avenue to enable precise continuous assessment of MS disease in clinical trials and real-world practice settings

https://www.jmir.org/2019/8/e14863/

Potential Regulatory Framework

FDA Discussion Document



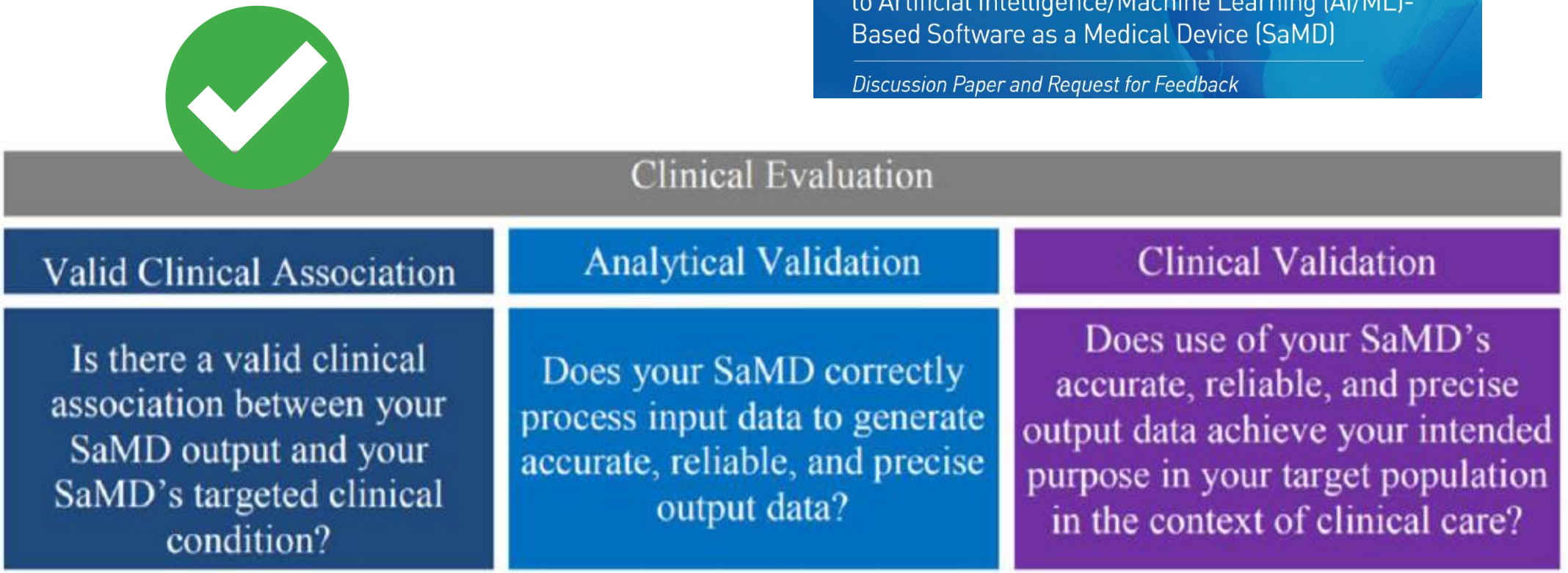


Figure 3: IMDRF description of Clinical Evaluation components

Tick Image: https://cdn3.iconfinder.com/data/icons/flat-actions-icons-9/792/Tick_Mark_Dark-512.png

Potential Regulatory Framework

FDA Discussion Document

Clinical Evaluation		
Valid Clinical Association	Analytical Validation	Clinical Validation
Is there a valid clinical association between your SaMD output and your SaMD's targeted clinical condition?	Does your SaMD correctly process input data to generate accurate, reliable, and precise output data?	Does use of your SaMD's accurate, reliable, and precise output data achieve your intended purpose in your target population in the context of clinical care?

Figure 3: IMDRF description of Clinical Evaluation components

Leverage Similar Regulatory Framework from IVD

Precision (CLSI EP5)

High Level Description of Concept:

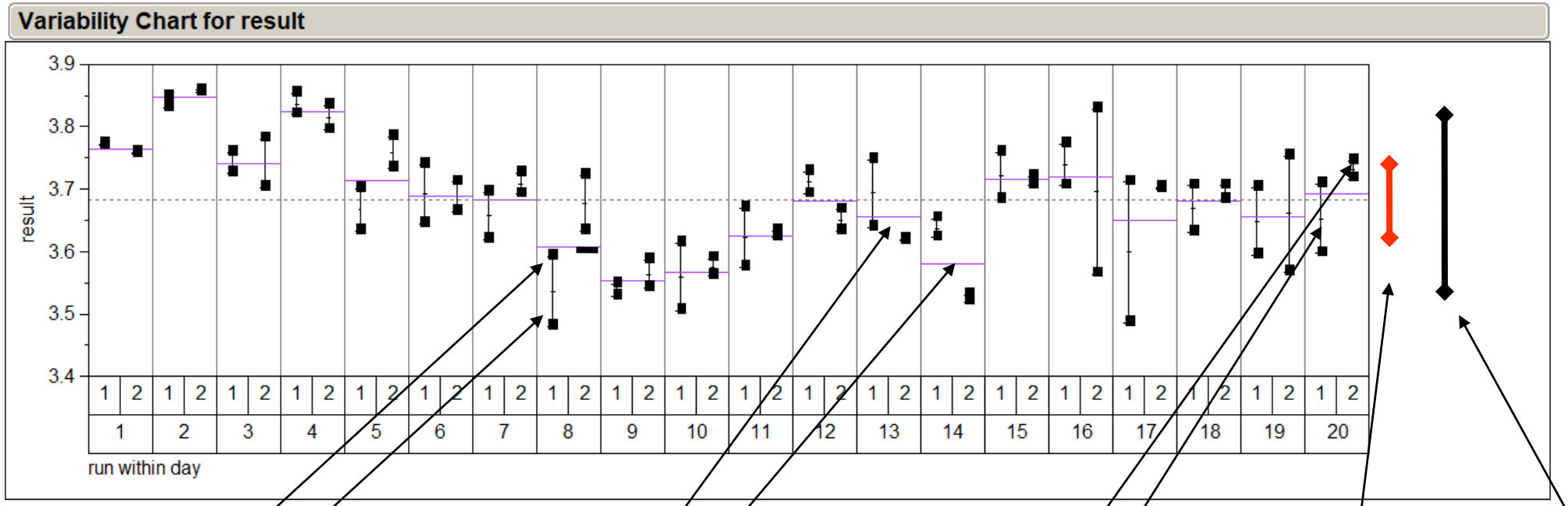
- Evaluates the random measurement error characteristics of a diagnostic test
- Multiple measurements on the same sample – fluctuations, variation of these measurements is important, not the absolute concentration value
- There is not one precision / variability, but variability components
  - Immediate repetition of measurement
  - Run – Run influences
  - Day - Day influences
  - Unit - Unit influences
  - Lot - Lot influences …….
- CLSI EP5 guideline provides standardized experiment designs and statistical analyses to quantify precision

Based on a slide from Fabian Model

Leverage Similar Regulatory Framework from IVD

Precision (CLSI EP5) – Example of CLSI precision experiment

3 Variance components in an experiment – *21 Days, 2 runs per day, 2 aliquots per run*



Mean variability within a run
= Repeatability

Mean variability of
daily mean values
= day - day variation

Mean variability of mean
values of each run
= run - run variation

Repeatability

Intermediate
Precision –
within-lab
precision

Sum of the components of variation
= Intermediate Precision /Within-lab Precision

Based on a slide from Fabian Model

Potential Regulatory Framework

FDA Discussion Document (cont.)

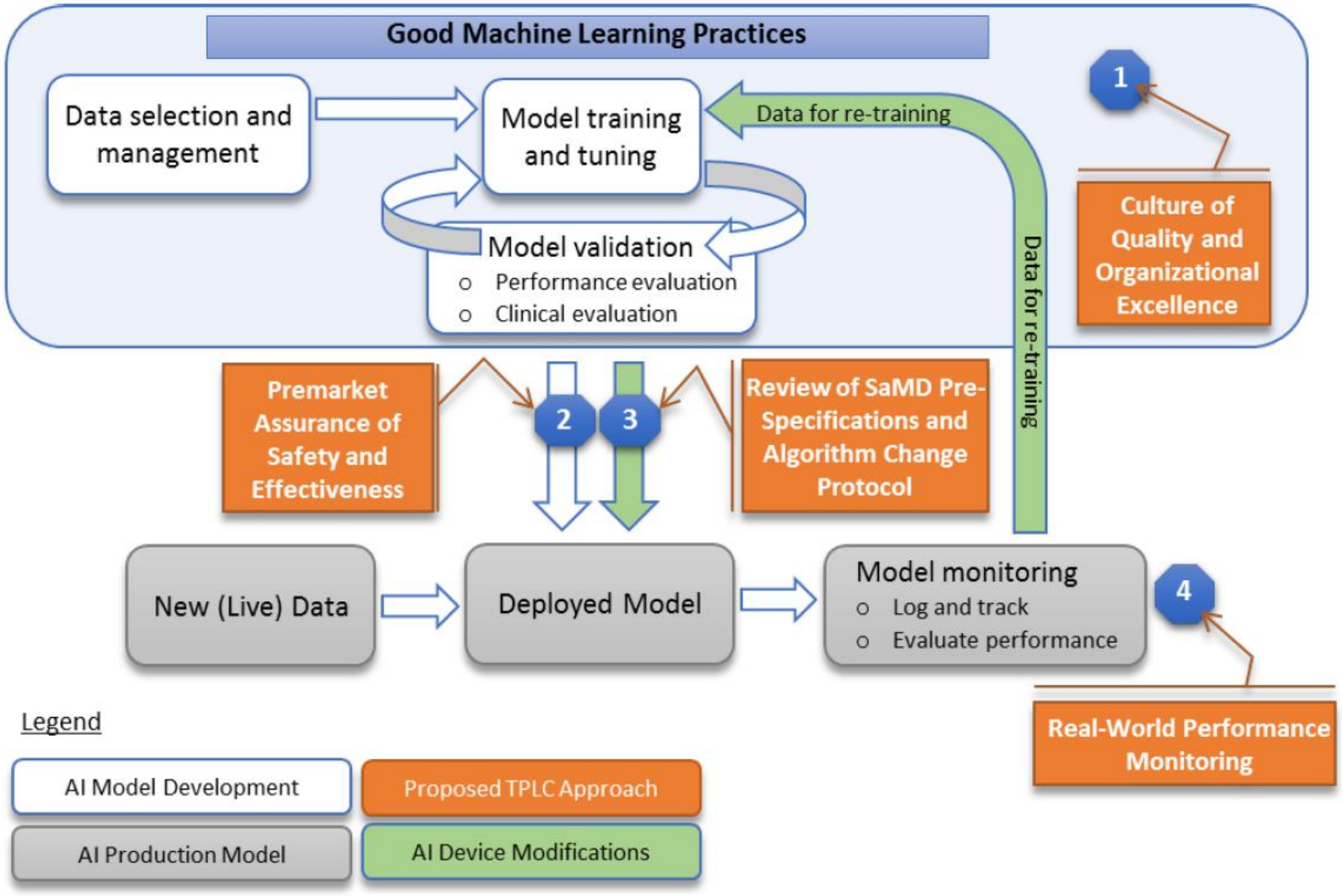


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

VI. Appendix B: Proposed Content for an Algorithm Change Protocol (ACP)

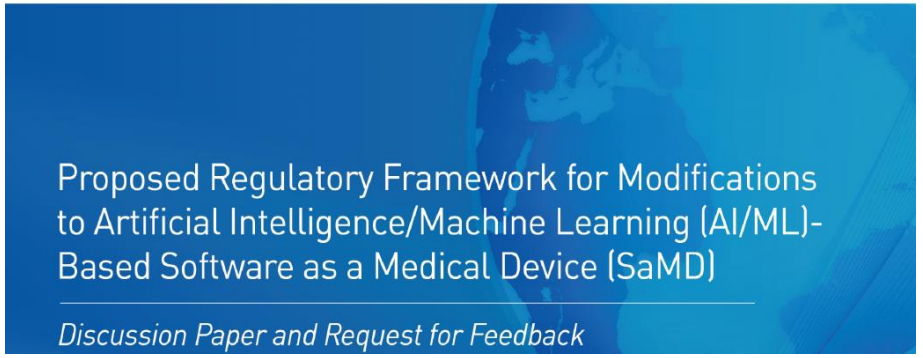
[…]

Performance evaluation protocols: These protocols may include a description of the intervals of when a new algorithm may be trained and evaluated to consider updating the medical device algorithm; the delineation of appropriate metrics and analysis procedures; statistical analysis plans; appropriate measures to minimize information leakage about the test data set if part of it is re-used in multiple evaluations; […]

Update procedures that describe how updated medical device algorithms will be tested, distributed, and communicated when released: […]

Current Regulatory Framework

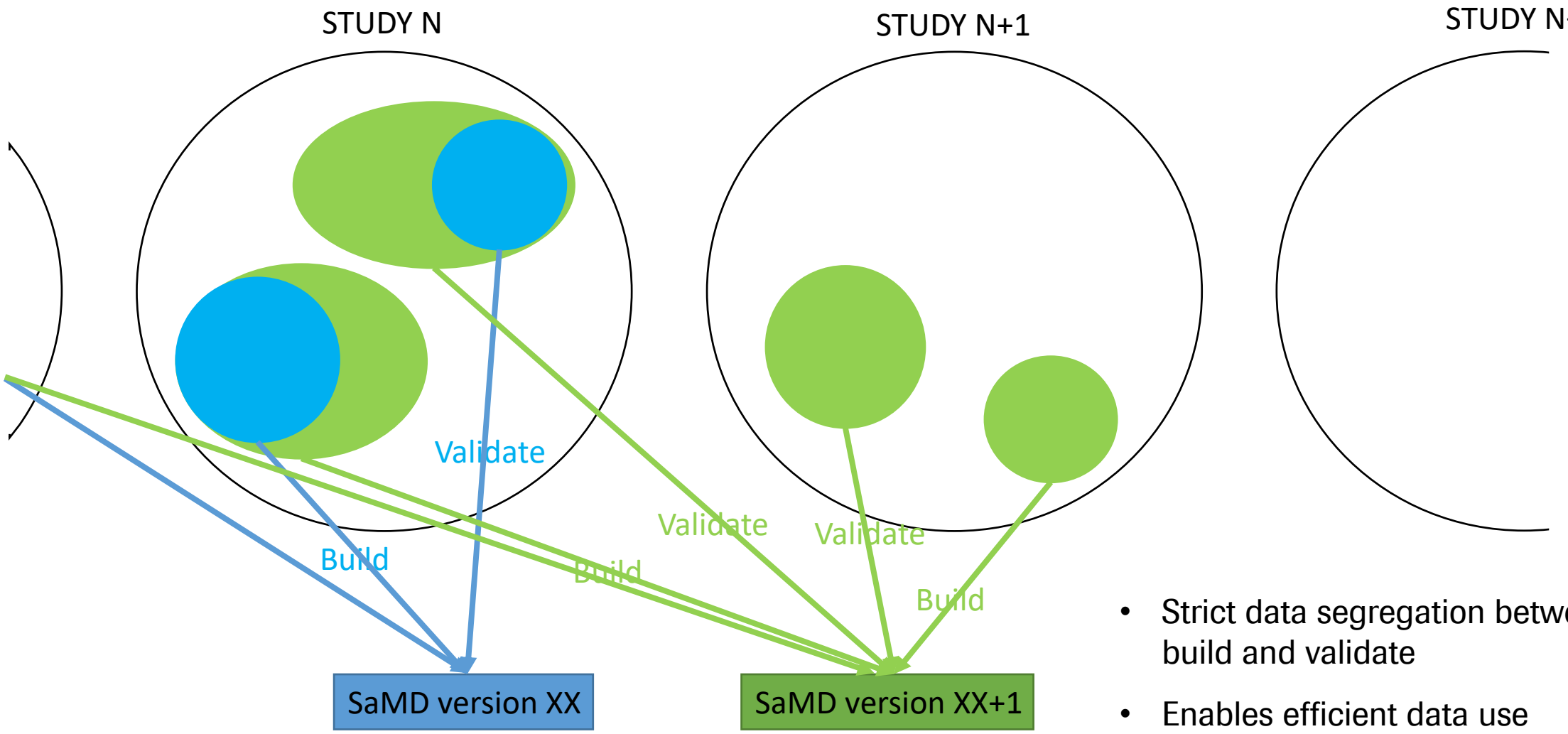
FDA Discussion Document



Clinical Evaluation		
Valid Clinical Association	Analytical Validation	Clinical Validation
Is there a valid clinical association between your SaMD output and your SaMD's targeted clinical condition?	Does your SaMD correctly process input data to generate accurate, reliable, and precise output data?	Does use of your SaMD's accurate, reliable, and precise output data achieve your intended purpose in your target population in the context of clinical care?

Figure 3: IMDRF description of Clinical Evaluation components

Proposal of Build & Validate Data Strategy



- Strict data segregation between build and validate
- Enables efficient data use
- Fast development timelines

Statistical Validation of Key Clinical Performance Measures

Cross-section Correlation at Week xx (SaMD version 2)

- Week XX digital outcome measure: aggregation (e.g. median) into a single value of all data available in a short time period around Week XX clinical visit.
- Clinical performance evaluation: compute Spearman correlation at Week XX between digital outcome measure and respective clinical outcome measure.

Longitudinal Correlation (SaMD version 3)

- Longitudinal digital outcome measure: change from baseline to Week YY of digital outcome measure.
- Longitudinal digital outcome measures clinical performance evaluation: correlate (e.g. Hazard Ratio and AUC (Harrel's C-index)) longitudinal digital outcome measure with relevant clinical time-to-event endpoints.

Prediction (SaMD version 4)

- Predictive digital outcome measure: X year follow-up of digital data can predict Y years of clinical measures. The Y years of clinical follow-up will be considered after the X year digital one.
- Predictive digital outcome measure clinical performance evaluation: perform AUC (Harrel's C-index).

Early Lessons Learned & Opportunities for Statisticians

Early Lessons Learned

- Merging sensor and non-sensor might not be obvious.
- Strict segregation and blinding of sensor data require new processes.
- Evolving regulatory frameworks drive statistics and data management requirements.

Opportunities

- How to ensure that people's smartphones meet the minimum requirements (e.g. accuracy)?
- How to avoid having all data collected at the same time (e.g. the week before visiting the physician)?
- Is Missingness informative?
- How to ensure that people continue to use the app?

Image: https://pmstudycircle.com/2012/01/lesson-learned/

Doing now what patients need next