MULTIPLE SCLEROSIS JOURNAL | MSJ

Original Research Paper

A smartphone sensor-based digital outcome assessment of multiple sclerosis

Xavier Montalban, Jennifer Graves, Luciana Midaglia, Patricia Mulero, Laura Julian, Michael Baker, Jan Schadrack, Christian Gossens, Marco Ganzetti, Alf Scotland, Florian Lipsmeier, Johan van Beek, Corrado Bernasconi, Shibeshih Belachew, Michael Lindemann* and Stephen L Hauser*

Multiple Sclerosis Journal

2022, Vol. 28(4) 654–664

DOI: 10.1177/13524585211028561

Abstract

Background: Sensor-based monitoring tools fill a critical gap in multiple sclerosis (MS) research and clinical care.

Objective: The aim of this study is to assess performance characteristics of the Floodlight Proof-of-Concept (PoC) app.

Methods: In a 24-week study (clinicaltrials.gov: NCT02952911), smartphone-based active tests and passive monitoring assessed cognition (electronic Symbol Digit Modalities Test), upper extremity function (Pinching Test, Draw a Shape Test), and gait and balance (Static Balance Test, U-Turn Test, Walk Test, Passive Monitoring). Intraclass correlation coefficients (ICCs) and age- or sex-adjusted Spearman's rank correlation determined test–retest reliability and correlations with clinical and magnetic resonance imaging (MRI) outcome measures, respectively.

Results: Seventy-six people with MS (PwMS) and 25 healthy controls were enrolled. In PwMS, ICCs were moderate-to-good (ICC(2,1) = 0.61–0.85) across tests. Correlations with domain-specific standard clinical disability measures were significant for all tests in the cognitive ($r = 0.82$, $p < 0.001$), upper extremity function ($|r| = 0.40$–0.64, all $p < 0.001$), and gait and balance domains ($r = -0.25$ to -0.52, all $p < 0.05$; except for Static Balance Test: $r = -0.20$, $p > 0.05$). Most tests also correlated with Expanded Disability Status Scale, 29-item Multiple Sclerosis Impact Scale items or subscales, and/or normalized brain volume.

Conclusion: The Floodlight PoC app captures reliable and clinically relevant measures of functional impairment in MS, supporting its potential use in clinical research and practice.

Keywords: Multiple sclerosis, smartphone, sensors, digital health technology, wearable electronic devices, mobile phone

Date received: 12 February 2021; revised: 1 June 2021; accepted: 3 June 2021

Correspondence to:
C Gossens
F. Hoffmann-La Roche Ltd, 124 Grenzacherstrasse, 4070 Basel, Switzerland.
christian.gossens@roche.com

Xavier Montalban
Patricia Mulero
Department of Neurology-Neuroimmunology, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Hospital Universitari Vall d'Hebron, Barcelona, Spain

Jennifer Graves
Department of Neurosciences, University of California San Diego, San Diego, CA, USA

Luciana Midaglia
Department of Neurology-Neuroimmunology, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Hospital Universitari Vall d'Hebron, Barcelona, Spain and Department of Medicine, Autonomous University of Barcelona, Barcelona, Spain

Laura Julian
Genentech, Inc., South San Francisco, CA, USA

Michael Baker
Jan Schadrack
Christian Gossens
Marco Ganzetti
Alf Scotland
Florian Lipsmeier
Johan van Beek
Corrado Bernasconi
Shibeshih Belachew
Michael Lindemann
F. Hoffmann-La Roche Ltd, Basel, Switzerland

Stephen L Hauser
UCSF Weill Institute for Neurosciences and Department of Neurology, University of California San Francisco, San Francisco, CA, USA

* Shared senior authorship.

Introduction

The course of multiple sclerosis (MS) was traditionally categorized as relapsing or progressive, with relapses associated with focal inflammation and progression with neurodegeneration. Recent advances in therapeutics have increased control of MS relapses in most patients but have also unmasked an underlying insidious progression, termed progression independent of relapses, in many patients previously thought to have relapsing–remitting disease.[1] Thus, progression can now be viewed as a universal feature of MS that is present throughout the disease course. Consequently, optimal control of MS would require that progression is minimized or eliminated in all patients. To achieve this goal, sensitive tools for monitoring disability in all people with MS (PwMS) are required, even for patients whose disease activity appears to be superficially under control in terms of relapses.[2] Furthermore, detection of progression onset or worsening is critical to optimally adapt the therapeutic strategy. Additional challenges in MS disability assessment include detection of pseudo-relapses and symptoms that often fluctuate with illness, fatigue, or changes in body temperature.[3–5] These fluctuations limit the utility of once- or twice-yearly in-clinic

monitoring; a more frequent estimate of function during daily life routine is likely to have greater value in tracking MS impairment.

Smartphone sensor-based, remotely administered digital tests represent a promising new avenue to capture disability burden with quantitative accuracy.[6] Such tests typically allow the remote capture of ecologically valid measures in a frequent, time- and cost-efficient fashion with minimal patient burden. In addition, they quantify multiple aspects of nervous system function during a given task, in opposition to clinician-administered disability tests that only assess the capacity to complete a task with a performance summary score.[7,8] Consequently, sensor-based tests can potentially disentangle intra- and inter-individual differences in underlying patterns of overall identical levels of MS-related functional impairment and thereby augment the resolution of disability measures compared with traditional scores.[7] They are also rater-independent, offering more objective measures of functional ability.[8–10]

In recent years, sensor-based tests have been increasingly used to assess functional ability in MS,[7–14] as well as in Parkinson's disease[15] and Huntington's disease.[16,17] The Floodlight Proof-of-Concept (PoC) app was designed to remotely measure, at home in an unsupervised setting, functional ability in cognitive, upper extremity, and gait and balance domains using a smartphone device without additional hardware.[9] Consequently, assessment with the Floodlight PoC app can be performed more frequently than clinician-administered tests, which potentially allows the capture of subtle changes in function that occur in free-living environments between clinic visits, including changes that would not necessarily trigger clinical interventions. The study "Monitoring of Multiple Sclerosis Participants With the Use of Digital Technology (Smartphones and Smartwatches)—A Feasibility Study" was the first clinical trial to implement the Floodlight PoC app.[9] Using data from this study, we previously demonstrated that PwMS were engaged and highly satisfied with this app.[9] We expand on this prior work by presenting here the critical evaluation of the app's test–retest reliability and its correlations with standard clinical and magnetic resonance imaging (MRI) measures in PwMS.

Methods

Study design and participants

The present study is a 24-week, prospective study (clinicaltrials.gov: NCT02952911) aimed to assess the feasibility of remotely monitoring PwMS with the Floodlight PoC app, which was developed for use in this study on a provisioned smartphone and smartwatch. The full study design and inclusion/exclusion criteria have been previously described.[9] Both PwMS and healthy controls (HC) were 18–55 years of age. PwMS (untreated or treated) were diagnosed according to the 2010 revised McDonald criteria[18] and had an Expanded Disability Status Scale (EDSS) score at baseline between 0.0 and 5.5, inclusive. All patients provided written informed consent. At each scheduled clinic visit (baseline, Week 12, Week 24), all participants underwent a clinical evaluation. In addition, PwMS were assessed by MRI at baseline and Week 24.

Participants were provided with a preconfigured smartphone (Samsung Galaxy S7) and smartwatch (Motorola 360 Sport) with the Floodlight PoC app installed;[9] for this report, only the smartphone data are reported. The smartphone prompted all participants to perform daily or weekly remote, sensor-based tests, referred to as active tests (Table 1). In addition, for passive monitoring, sensor data were passively recorded to examine ambulation and overall mobility in daily life (Table 1). Participants were encouraged to continuously carry the device to gather data for passive monitoring.

Floodlight PoC app

The Floodlight PoC app was designed to assess functional abilities across three key domains affected by MS: cognition, upper extremity function, and gait and balance.[9] Cognition was assessed using the electronic Symbol Digit Modalities Test (e-SDMT). As with the oral Symbol Digit Modalities Test (SDMT), it assessed impairment of key neurologic functions underlying cognitive information processing speed. The aim was to correctly match a maximum number of symbols to their paired digits within 90 seconds.

The Pinching Test and Draw a Shape Test assessed upper extremity function. The Pinching Test evaluated fine pinching or grasping dexterity and instructed participants to successfully pinch as many circular tomato shapes appearing on the smartphone screen at different positions as possible within 30 seconds. The Draw a Shape Test, which required participants to draw six prewritten shapes of increasing complexity (two diagonal lines, a square, circle, a figure-of-8, and a spiral), assessed fine finger or manual dexterity

Finally, the Static Balance Test (SBT), U-Turn Test (UTT), Walk Test, and Passive Monitoring examined

Table 1. Test features from the Floodlight PoC app included in the analysis.

Functional domain	Test/feature	Definitions[a]	Disability concept assessed by the feature	Sensor	Testing frequency	Functional meaning of higher scores
Cognition	**e-SDMT**					
	Number of correct responses (*n*)	Number of correct responses given within 90 seconds	Cognitive speed to process information	Touchscreen	Weekly	Better
Upper extremity function	**Pinching Test**					
	Double touch asynchrony (s)	Gap duration between the first and the second finger touching the screen	Ability to synchronously perform fine distal upper extremity movement	Touchscreen	Daily	Worse
	Number of successful pinches (*n*)	Number of tomatoes successfully pinched, or squeezed, within 30 seconds	Ability to perform fast and accurate finger opposition movement	Touchscreen	Daily	Better
	Draw a Shape Test					
	Overall mean trace accuracy	Proportion of sample points that overlap with the reference shape	Accuracy of fine distal upper extremity movement	Touchscreen	Daily	Better
	Overall mean trace celerity (s^{-1})	Ratio of overall mean trace accuracy and the time needed to draw the shape	Speed and accuracy of fine distal upper extremity movement	Touchscreen	Daily	Better
Gait and balance	**SBT**					
	Sway path (m/s^2)	Sum of the accelerometer signals in the *x*-, *y*-, and *z*-axis.	Ability to maintain stable orthostatic posture	Accelerometer	Daily	Worse
	UTT					
	Turn speed (rad/s)	Angular velocity while performing U-turns	Ability to turn while walking	Accelerometer and gyroscope	Daily	Better
	Walk Test					
	Step power	Integral of mean-centered acceleration magnitude signal over time divided by number of steps	Energy invested per step during regular walking	Accelerometer	Daily	Better
	Passive Monitoring					
	Turn speed (rad/s)	Angular velocity while performing U-turns	Ability to turn while walking	Accelerometer and gyroscope	Continuously	Better
	Step power	Integral of mean-centered acceleration magnitude signal over time divided by number of steps	Energy invested per step during regular walking	Accelerometer	Continuously	Better

PoC: Proof of Concept; e-SDMT: electronic Symbol Digit Modalities Test; SBT: Static Balance Test; UTT: U-Turn Test.
[a]See Table S2 in the supplementary appendix for the mathematical definitions.

gait and balance. The SBT was designed to study balance by asking the participants to stand, unsupported, as still as possible for 30 seconds. The UTT instructed participants to perform five successive U-turns separated by at least 4 meters within 1 minute at a comfortable pace to evaluate difficulties or unusual patterns in turning while walking and dynamic balance. The Walk Test aimed to assess gait while walking as fast as possible but also safely for 2 minutes on an even ground without performing U-turns. Finally, Passive Monitoring examined both aspects of gait—turning while walking and regular, straight walking—throughout the day. While performing the gait assessments, participants carried the smartphone in a running belt or in their trouser pocket and were permitted to use an assistive device and/or orthotic as needed.

For each active test and passive monitoring, one to two test features that are illustrative of the test and probe different neurologic concepts were extracted from the raw sensor data (Table 1).

Signal processing

As the smartphone-based tests were performed without supervision by a physician or study coordinator, quality control steps were applied to identify and exclude individual assessments that were performed incorrectly. This ensures the measurements are both reliable and accurate. To exclude such incorrectly performed assessments, quality control flags were defined for each test (Table S1). In addition, only sufficiently adherent participants, that is, those who contributed at least six individual assessments in the course of the study, were included in the analysis of that particular test. Applying these two quality control steps resulted in the final dataset consisting of valid assessments.

Next, all valid assessments contributed by a participant were aggregated to study test–retest reliability and Spearman's rank correlations in a cross-sectional analysis. As the tests were performed once daily at most, the test–retest analysis was based on the median test performance on the active tests and passive monitoring during 12 two-week windows. Two-week windows were chosen to reduce variability that is independent of general disease status and might be attributed to good or bad days or to differences between weekdays and weekends. For the cross-sectional correlation analysis, the median test performance across the entire study duration as well as the mean of the three in-clinic assessments (mean of two assessments for MRI) were computed.

Statistical analysis

Test–retest reliability was evaluated with intraclass correlation coefficients (ICC[2,1]) separately in PwMS and HC, which considered all consecutive 2-week windows. Generally, at least three valid individual assessments were required for each 2-week period. An exception was made for the e-SDMT to accommodate its weekly testing schedule; only one valid e-SDMT assessment for each 2-week window was considered sufficient. Test–retest reliability was considered as poor (ICC $<$ 0.5), moderate (ICC $=$ 0.5–0.75), good (ICC $=$ 0.75–0.9), or excellent (ICC $>$ 0.9).[19]

To examine the agreement of the test features from the Floodlight PoC app with standard clinical and MRI measures in PwMS, the test features were correlated against domain-specific clinical measures (oral SDMT, Nine-Hole Peg Test [9HPT], Berg Balance Scale [BBS], Timed 25-Foot Walk [T25FW]), EDSS, 29-item Multiple Sclerosis Impact Scale (MSIS-29) items or subscales, T2 FLAIR (fluid-attenuated inversion recovery lesion volume, and normalized brain volume using Spearman's rank correlation. In addition, test features from UTT and Walk Test were correlated against corresponding features obtained from Passive Monitoring. The strength of correlation was considered as good-to-excellent ($|r|>$ 0.75), moderate-to-good ($|r|=$ 0.5–0.75), fair ($|r|=$ 0.2–0.49), or not correlated ($|r|<$ 0.25), where $|r|$represents the absolute value.[20]

Partial correlation analyses were performed for the upper extremity function tests and gait assessments to investigate the contribution of each test and test feature in predicting 9HPT times and T25FW times, respectively. All test features from the Pinching Test and the Draw a Shape Test as well as 9HPT time were included in the upper extremity function model. Each of these features was correlated against each other while controlling for the remaining features in the model. The partial correlation for the gait assessments was run in a similar fashion but included all test features from the UTT, Walk Test, and Passive Monitoring in addition to T25FW time instead.

All correlations were adjusted for age and sex with a robust linear model. Statistical significance was set at $p < 0.05$ without correction for multiple comparisons.

Results

Full baseline demographics and disease characteristics have been previously described[9] and are summarized in Table 2. In total, 76 PwMS were enrolled between

Table 2. Baseline demographics and disease characteristics.

Variable	$N = 76$
Age (years), mean (SD)	39.5 (7.9)
Female, *n* (%)	53 (69.7)
Diagnosis, *n* (%)	
Primary progressive multiple sclerosis	3 (3.9)
Secondary progressive multiple sclerosis	4 (5.3)
Relapsing–remitting multiple sclerosis	69 (90.8)
Expanded Disability Status Scale	
Mean (SD)	2.4 (1.4)
Median	2.0
IQR	1.5–3.5
Range	0.0–5.5
Oral Symbol Digit Modalities Test score	
Mean (SD)	53.8 (11.8)
Median	55.0
IQR	46.0–63.0
Range	26.0–77.0
Nine-Hole Peg Test time (s)	
Mean (SD)	22.2 (4.0)
Median	21.4
IQR	19.7–23.7
Range	16.4–39.6
Berg Balance Scale total score	
Mean (SD)	52.5 (5.7)
Median	56.0
IQR	51.0–56.0
Range	31.0–56.0
Timed 25-Foot Walk time (s)	
Mean (SD)	6.0 (2.1)
Median	5.3
IQR	4.6–6.8
Range	3.5–12.1
T2 FLAIR lesion volume (mL)	
Mean (SD)	6.3 (7.5)
Median	3.0
IQR	1.0–9.3
Range	0.1–31.2
Normalized brain volume (mL)	
Mean (SD)	1474.5 (75.6)
Median	1477.2
IQR	1439.6–1525.6
Range	1137.3–1628.1

SD: standard deviation; IQR: interquartile range; FLAIR: fluid-attenuated inversion recovery.

November 2016 and May 2018. Most PwMS had mild disease (mean baseline EDSS: 2.4; EDSS range: 0.0–5.5) and were diagnosed with relapsing–remitting MS. Valid data were available in 68–73 PwMS. PwMS and in 18–24 HC across the 10 tests considered (Figure S1), using test-specific quality control criteria. Similarly, 81.5–99.9% of assessments in PwMS and 76.0–99.9% of assessments in HC were considered valid (Figure S2). During the 24-week study, only eight participants exhibited a change in EDSS $\geqslant 1$, which is defined as clinically meaningful,[21] an insufficient number to enable further analysis of changes in EDSS (Figure S3).

Test–retest reliability

Test–retest reliability was assessed in PwMS and HC with valid assessments in all consecutive 2-week windows (range: 32–46 PwMS and 8–11 HC). In PwMS, ICCs(2,1) were moderate or good (Table 3), suggesting that reliable data can be captured with the Floodlight PoC app. In HC, where the group sizes were lower, ICCs(2,1) were mostly poor to good.

Correlation with clinical and MRI measures

The age- and sex-adjusted Spearman's rank correlation analysis in PwMS is summarized in Table 3 and Figure 1. All statistically significant correlations were in the expected direction. Thus, increasing levels of MS-related disability were associated with worse performance on the Floodlight PoC app.

Overall, strongest correlations of test features were observed with the respective domain-specific standard clinical disability measures. These correlations were good-to-excellent in the cognitive domain ($r = 0.82$) and fair or moderate-to-good in the upper extremity function domain ($|r| = 0.40$–0.64) and gait and balance domain ($r = -0.25$ to -0.52, all $p < 0.05$). Only the SBT did not correlate with its domain-specific standard clinical measure, the BBS ($r = -0.20$, $p > 0.05$). Most test features also correlated with EDSS (all $p < 0.05$ except for Draw a Shape Test overall mean trace celerity and Passive Monitoring step power) and their respective MSIS-29 subscale or items (all $p < 0.05$ except for Draw a Shape Test overall mean trace celerity, Passive Monitoring turn speed, and Passive Monitoring step power). Normalized brain volume correlated significantly with test features across all domains with the strongest association found with e-SDMT ($r = 0.54$, $p < 0.001$). Similar results were obtained with unadjusted measures (Table S3).

Table 3. Test–retest reliability in PwMS and HC and age- and sex-adjusted Spearman's rank correlation analysis of the Floodlight PoC app in PwMS.

Domain	Test	Feature	Test–retest reliability ICC (2,1) (95% CI)		Spearman's rank correlations for PwMS				
			HC	PwMS	Domain-specific standard clinical measure	EDSS	MSIS-29 subscale/ items[a]	T2 FLAIR lesion volume (mL)	Normalized brain volume (mL)
Cognition	e-SDMT	Number of correct responses (*n*)	0.55 (0.34–0.80)[b]	0.85 (0.76–0.91)[c]	Oral SDMT	0.82***	−0.43***	−0.52***	−0.42***	0.54***
Upper extremity function	Pinching Test	Double touch asynchrony (s)	0.72 (0.53–0.90)[d]	0.71 (0.62–0.80)[c]	9HPT	0.64***	0.30*	0.35**	0.17	−0.26*
	Pinching Test	Number of successful pinches (*n*)	0.81 (0.65–0.94)[d]	0.72 (0.61–0.82)[c]	9HPT	−0.52***	−0.26*	−0.33**	−0.12	0.32**
	Draw a Shape Test	Overall mean trace accuracy	0.53 (0.32–0.79)[b]	0.85 (0.79–0.90)[e]	9HPT	−0.48***	−0.40***	−0.40***	−0.26*	0.33**
	Draw a Shape Test	Overall mean trace celerity (s^{-1})	0.45 (0.25–0.73)[b]	0.81 (0.73–0.87)[e]	9HPT	−0.40***	−0.08	0.03	−0.26*	0.24*
Gait and balance	SBT	Sway path (m/s^2)	0.40 (0.20–0.73)[f]	0.71 (0.61–0.80)[g]	BBS	−0.20	0.24*	0.31**	0.21	−0.05
	UTT	Turn speed (rad/s)	0.45 (0.24–0.75)[d]	0.83 (0.76–0.89)[h]	T25FW	−0.52***	−0.45***	−0.39***	−0.13	0.27*
	Walk Test	Step power	0.85 (0.70–0.95)[f]	0.78 (0.70–0.86)[i]	T25FW	−0.31**	−0.28*	−0.24*	0.04	−0.02
	Passive Monitoring	Turn speed (rad/s)	0.66 (0.42–0.89)[j]	0.72 (0.61–0.82)[k]	T25FW	−0.25*	−0.27*	−0.12	−0.11	0.14
	Passive Monitoring	Step power	0.63 (0.39–0.88)[j]	0.61 (0.48–0.74)[k]	T25FW	−0.33**	−0.19	−0.22	0.09	0.00

PwMS: people with multiple sclerosis; HC: healthy controls; PoC: Proof of Concept; ICC: intraclass correlation coefficient; CI: confidence interval; EDSS: Expanded Disability Status Scale; MSIS-29: 29-item Multiple Sclerosis Impact Scale; FLAIR: fluid-attenuated inversion recovery; e-SDMT: electronic Symbol Digit Modalities Test; 9HPT: Nine-Hole Peg Test; SBT: Static Balance Test; BBS: Berg Balance Scale; UTT, U-Turn Test; T25FW Timed 25-Foot Walk.
Colored background indicates significant correlation in expected direction.
[a]The e-SDMT was correlated against the psychological subscale, the Pinching and Draw a Shape Tests against the arm-related items (items 2, 6 and 15), and all other tests against the physical subscale.
[b]$n = 11$.
[c]$n = 46$.
[d]$n = 10$.
[e]$n = 44$.
[f]$n = 9$.
[g]$n = 42$.
[h]$n = 41$.
[i]$n = 39$.
[J]$n = 8$.
[k]$n = 32$.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

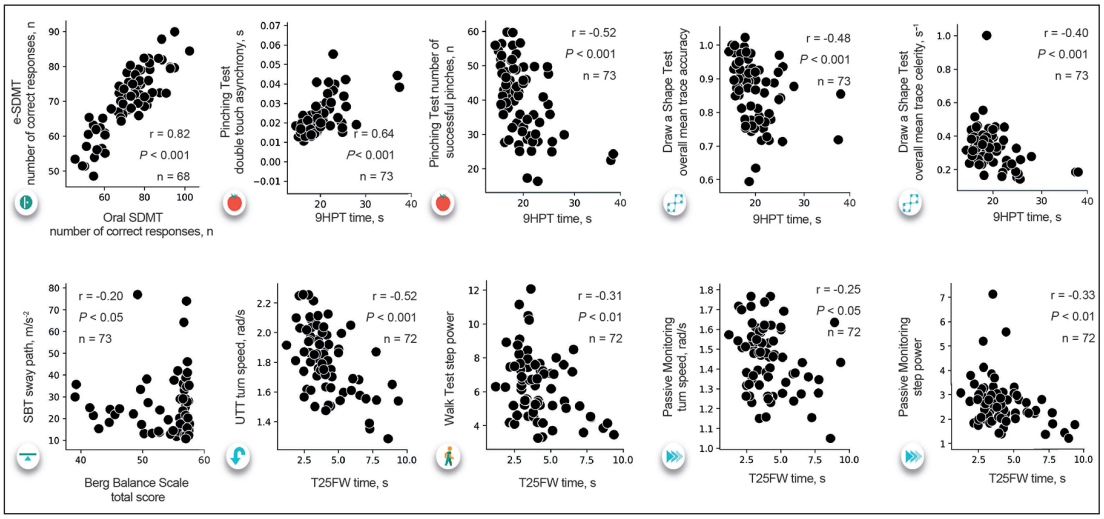

Figure 1. Age- and sex-adjusted Spearman's rank correlations between active tests and passive monitoring (vertical axis) and their respective domain-specific standard clinical measures (horizontal axis).
e-SDMT: electronic Symbol Digit Modalities Test; SDMT: Symbol Digit Modalities Test; 9HPT: Nine-Hole Peg Test; SBT: Static Balance Test; UTT: U-Turn Test; T25FW: Timed 25-Foot Walk.

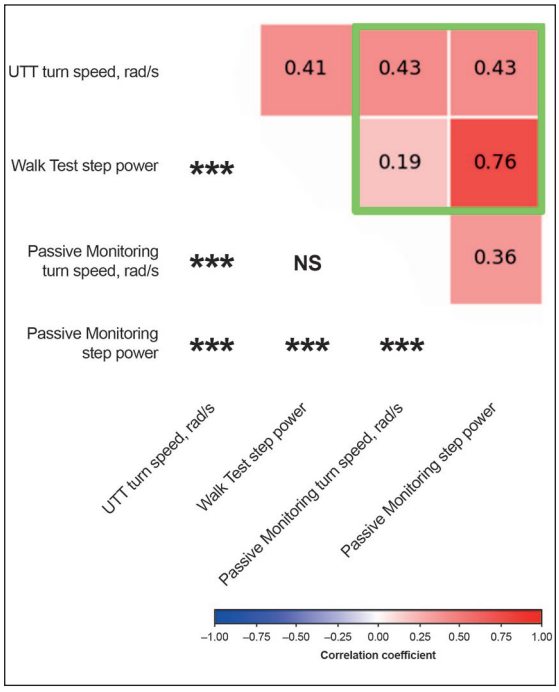

Figure 2. Age- and sex-adjusted Spearman's rank correlations between passive monitoring and active gait features.
UTT: U-Turn Test; NS: not significant.
***$p < 0.001$.

Next, we assessed correlations between active gait tests and passive monitoring. UTT turn speed showed moderate-to-good correlation with Passive Monitoring turn speed ($r = 0.43$, $p < 0.001$). Stronger, good-to-excellent positive correlations were observed between Walk Test step power and Passive Monitoring step power ($r = 0.76$, $p < 0.001$; Figure 2).

The partial correlation analysis revealed that both the Pinching Test (double touch asynchrony: partial $r = 0.37$, $p < 0.001$) and Draw a Shape Test (overall mean trace accuracy: partial $r = -0.40$, $p < 0.001$; overall mean trace celerity: partial $r = -0.30$, $p < 0.01$) contain independent information in predicting 9HPT time (Figure S4). Similarly, the UTT carries unique information in predicting T25FW time when correcting for the other gait features (partial $r = -0.31$, $p < 0.01$; Figure S5).

Comparing the performance of the e-SDMT and oral SDMT in PwMS and HC showed that both HC and PwMS had fewer correct responses on the e-SDMT compared with the oral SDMT (Figure S6).

Discussion

Here, we provide the first evidence that the Floodlight PoC app can reliably capture clinically relevant data measures of functional impairment in PwMS. By leveraging smartphone-based consumer technology for clinical research, the Floodlight PoC app assesses key neurological domains affected by MS and provides a more objective and detailed picture of the disease than is possible with standard "point of care" clinical assessments. Results from this study indicate that test features derived from the Floodlight PoC app hold potential for use in clinical research and practice.

Test–retest reliability was consistent with ICCs reported for standard clinical measures in PwMS.[22] As

anticipated, statistically significant correlations were observed between test features from the Floodlight PoC app and related standard clinical and MRI measures. Correlations of similar strength between comparable smartphone sensor-based remote monitoring tests and clinician-administered tests have been reported in the cognitive,[11,12] upper extremity function,[13] and gait domains,[10] despite differences in test design and features. The active tests and passive monitoring also showed mostly fair correlations with EDSS and respective MSIS-29 subscales and items, indicating that measurements obtained with the Floodlight PoC app agree both with the overall level of MS-related disability and participant's perception of the impact of their disease. T2 FLAIR lesion volume did not show any strong correlation with either active tests or passive monitoring. This is not surprising given the clinico-radiological paradox,[23] which describes the mismatch between the white matter lesion volume and the clinical outcomes in MS, and the subsequent poor cross-sectional correlation between T2-weighted imaging and MS disability measures in a relatively mild MS population.[24] Normalized brain volume correlated with test features from all assessed domains. This is in line with previously reported correlations between normalized brain volume and measures of MS-related disability.[25]

Not surprisingly, the e-SDMT most closely resembles its domain-specific standard clinical measure. The Spearman rho of 0.82 is comparable to the previously reported correlation between other smartphone-based versions of the SDMT and the pen-and-paper version of the SDMT ($r = 0.71–0.85$).[11,12] In our study, we noted that the e-SDMT scores tended to be lower than the oral SDMT scores in PwMS and HC. This is likely due to the longer time required to select the correct response on a smartphone display compared with saying the correct response out loud. Another possible reason is that the e-SDMT displays only one symbol at a time. The oral SDMT, on the contrary, provides participants the entire symbol sequence printed on a sheet of paper,[26] thus allowing them to work ahead and use their working memory to a greater extent. This difference also makes the oral SDMT more dependent on eye tracking than the e-SDMT.

Given the different concepts assessed by the Floodlight PoC app versus clinical measures, 1:1 correlations were not necessarily expected. For example, mean trace celerity did not correlate with EDSS or the arm-related MSIS-29 items. This is likely because these clinical measures do not capture the time component as overall mean trace accuracy correlated significantly with both clinical measures.

The partial correlation analysis presented here revealed that test features from both the Pinching Test and the Draw a Shape Test independently correlate with 9HPT time. This supports the concept that specific sensor-based test features can capture performance outcome information currently not recorded with commonly used in-clinic assessments. This exemplifies the potential of sensor data to characterize functional impairment beyond a single summary score that is typically recorded for in-clinic performance outcome measures. Future work should explore the use of this technology in broader clinical applications and focus on establishing the clinical relevance for the additional information it can provide. It is possible that richer information can be extracted by incorporating additional test features. Initial results on a more comprehensive multidimensional feature space have been previously reported for the Draw a Shape Test[27] and Walk Test.[28,29]

In addition to the active tests, the Floodlight PoC app also assesses gait in a free-living situation, or in daily life, through passive monitoring. It has been suggested that signs of gait alteration may be more pronounced during daily life than in conventional in-clinic metrics,[30] thereby highlighting the importance of capturing out-of-clinic performance through passive monitoring. As such, passive monitoring may improve the translation of clinical findings to meaningful care as it informs on the patients' true abilities during daily life activities.[9] A recent study demonstrated the feasibility of passively monitoring gait in PwMS using three biosensors attached to the wrist, ankle, and sternum.[8] Moreover, our study revealed significant correlations between gait features from Passive Monitoring and T25FW, highlighting the feasibility of capturing more ecological measures of everyday functional ability through Passive Monitoring using a single smartphone device.

Several limitations to this study exist. Most enrolled PwMS had mild disease with limited MS-related disability; the mean EDSS at baseline was 2.4. The current analysis assessed the performance characteristics of the Floodlight PoC app in PwMS with EDSS scores in the range of 0.0–5.5. However, previously it has been shown that wearable sensors might not accurately capture step detection at slow walking speeds, particularly at EDSS score 6.5. This feature should be considered when assessing any wearable monitoring technology.[31] In addition, the analyses presented here were cross-sectional. Due to the relative short duration of the study (24 weeks), a longitudinal analysis on change in functional ability, disease progression, and relapses was not possible. Future studies will lend greater clarity into the use in a broader patient population, including people with more advanced disease, and the test performance over time.

Furthermore, test–retest reliability analysis was conducted in 2-week windows, in which no disease progression was assumed, as each assessment was done at most once per day. Same-day test–retest reliability analysis will be addressed in future work using data from subsequent studies. Further work will also be needed on the development of domain-specific and overall MS outcome measures based on digital health technology.[32]

Conclusion

Using a consumer smartphone device, self-administered at-home active tests and passive monitoring assessed the functional ability across three key domains affected by MS: cognition, upper extremity function, and gait and balance. This study demonstrated that the Floodlight PoC app provides reliable measures that align with standard clinical and MRI measures used to quantify MS functional impairment and overall disability. Test–retest reliability was moderate-to-good, and significant correlations in the expected direction were observed between the test features from the Floodlight PoC app and standard clinical and MRI measures. While active tests were conducted daily or weekly, passive monitoring permitted the continuous assessment of gait during daily life activities. The higher temporal resolution and multidimensional feature space of functional data collected by this platform hold the potential to capture subtle, potentially disease-related information which are not readily discriminated by clinician-administered assessments. It also has the potential to improve and standardize assessment of MS disease over time, provide PwMS and health care professionals in both specialty and primary care environments a better understanding of disease progression, change the way MS is monitored in clinical trials and daily practice, and ultimately improve patient care. The current iteration of the app, Floodlight™ MS, is available for public use in selected countries, and a rolling release schedule is now in process to provide access in the near future to the wider MS community across the world.

Acknowledgements

We would like to thank all patients, their families, and the investigators who participated in this trial. We would also like to thank the following colleagues at F. Hoffmann-La Roche Ltd for their contributions and support to the study: Jan Beckmann, Frank Dondelinger, Sandro Fritz, Nicholas Pierce Heinemeier, Sven Holm, Timothy Kilchenmann, Lito Kriara, Bernd Laub, Grégoire Pointeau, Caroline Polakowska, Cedric Simillion, Jens Schjodt-Eriksen, Jörg Sprengel, Ralf Stubner, Krzysztof Trybus, and Mattia Zanon. Writing and editorial assistance for this manuscript was provided by Samuel Bestall, PhD, and Marina Dragovic from Articulate Science, United Kingdom, and funded by F. Hoffmann-La Roche Ltd.

Declaration of Conflicting Interests

X.M. has received speaking honoraria and travel expenses for participation in scientific meetings, has been a steering committee member of clinical trials, or participated in advisory boards of clinical trials in the past years with Actelion, Alexion, Bayer, Biogen, Celgene, EMD Serono, Genzyme, Immunic, Medday, Merck, Mylan, Nervgen, Novartis, Roche, Sanofi Genzyme, Teva Pharmaceutical, TG Therapeutics, Excemed, MSIF, and NMSS. L.M. and P.M. have nothing to disclose. J.G. over the past year has received grant/contract research support from the National MS Society, Biogen, and Octave Biosciences. She serves on a steering committee for a trial supported by Novartis. She has received honoraria for a non-promotional, educational activity for Sanofi Genzyme. She has received speaker fees from Alexion and BMS and served on an advisory board for Genentech. L.J. is an employee of Genentech, Inc., and a shareholder of F. Hoffmann-La Roche Ltd. M.B., J.S., and C.G. are employees and shareholders of F. Hoffmann-La Roche Ltd. M.G. and C.B. are contractors for F. Hoffmann-La Roche Ltd. A.S. was a consultant to F. Hoffmann-La Roche Ltd via Inovigate during the completion of the work related to this manuscript. A.S. is now an employee of Inovigate (Basel, Switzerland). F.L. is an employee of F. Hoffmann-La Roche Ltd. J.v.B. and S.B. were employees of F. Hoffmann-La Roche Ltd during the completion of the work related to this manuscript. Both are now employees of Biogen (Cambridge, MA), which was not in any way associated with this study. M.L. is a consultant to F. Hoffmann-La Roche Ltd via Inovigate. S.L.H. serves on the scientific advisory boards for Alector, Annexon, Bionure, and Molecular Stethoscope; is on the Board of Directors for Neurona Therapeutics; and has received travel reimbursement and writing assistance from F. Hoffmann-La Roche Ltd and Novartis for CD20-related meetings and presentations.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: F. Hoffmann-La Roche Ltd, Basel, Switzerland, provided financial support for the study and publication of this manuscript.

Data availability

Qualified researchers may request access to individual patient-level data through the clinical study data

request platform (https://vivli.org). Further details on Roche's criteria for eligible studies are available at https://vivli.org/members/ourmembers. For further details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm.

ORCID iD

Shibeshih Belachew https://orcid.org/0000-0003-3976-1950

Supplemental material

Supplemental material for this article is available online.

References

1. Kappos L, Wolinsky JS, Giovannoni G, et al. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurol* 2020; 77: 1132–1140.
2. Hobart J, Bowen A, Pepper G, et al. International consensus on quality standards for brain health-focused care in multiple sclerosis. *Mult Scler* 2019; 25: 1809–1818.
3. Rae-Grant A, Bennett A, Sanders AE, et al. Quality improvement in neurology: Multiple sclerosis quality measures: Executive summary. *Neurology* 2015; 85: 1904–1908.
4. Steelman AJ. Infection as an environmental trigger of multiple sclerosis disease exacerbation. *Front Immunol* 2015; 6: 520–520.
5. Mills EA, Mirza A and Mao-Draayer Y. Emerging approaches for validating and managing multiple sclerosis relapse. *Front Neurol* 2017; 8: 116–116.
6. Graves JS and Montalban X. Biosensors to monitor MS activity. *Mult Scler* 2020; 26: 605–608.
7. Boukhvalova AK, Fan O, Weideman AM, et al. Smartphone level test measures disability in several neurological domains for patients with multiple sclerosis. *Front Neurol* 2019; 10: 358–358.
8. Chitnis T, Glanz BI, Gonzalez C, et al. Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis. *NPJ Digit Med* 2019; 2: 123–123.
9. Midaglia L, Mulero P, Montalban X, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: Nonrandomized interventional feasibility study. *J Med Internet Res* 2019; 21: e14863.
10. Block VJ, Lizée A, Crabtree-Hartman E, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *J Neurol* 2017; 264: 316–326.
11. Pham L, Harris T, Varosanec M, et al. Smartphone-based symbol-digit modalities test reliably captures brain damage in multiple sclerosis. *NPJ Digit Med* 2021; 4: 36.
12. van Oirschot P, Heerings M, Wendrich K, et al. Symbol Digit Modalities Test variant in a smartphone app for persons with multiple sclerosis: Validation study. *JMIR Mhealth Uhealth* 2020; 8: e18160.
13. Akhbardeh A, Arjona JK, Krysko KM, et al. Novel MS vital sign: Multi-sensor captures upper and lower limb dysfunction. *Ann Clin Transl Neurol* 2020; 7: 288–295.
14. Angelini L, Hodgkinson W, Smith C, et al. Wearable sensors can reliably quantify gait alterations associated with disability in people with progressive multiple sclerosis in a clinical setting. *J Neurol* 2020; 267: 2897–2909.
15. Lipsmeier F, Taylor KI, Kilchenmann T, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Mov Disord* 2018; 33: 1287–1297.
16. Tortelli R, Simillion C, Lipsmeier F, et al. The digital-HD study: Smartphone-based remote testing to assess cognitive and motor symptoms in Huntington's Disease (1816). *Neurology* 2020; 94(Suppl. 15): 1816.
17. McLaren B, Andrews SC, Glikmann-Johnston Y, et al. Feasibility and initial validation of "HD-Mobile," a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *J Neurol* 2021; 268: 590–601.
18. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011; 69: 292–302.
19. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.
20. Portney LG and Watkins MP. *Foundations of clinical practice research: Applications to practice*. London: Pearson, 2009.
21. Kragt JJ, Nielsen JM, van der Linden FA, et al. How similar are commonly combined criteria for EDSS progression in multiple sclerosis. *Mult Scler* 2006; 12: 782–786.
22. Goldman MD, LaRocca NG, Rudick RA, et al. Evaluation of multiple sclerosis disability outcome

measures using pooled clinical trial data. *Neurology* 2019; 93: e1921–e1931.

23. Chard D and Trip SA. Resolving the clinico-radiological paradox in multiple sclerosis. *F1000Res* 2017; 6: 1828.
24. Miller DH, Grossman RI, Reingold SC, et al. The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain* 1998; 121(Pt1): 3–24.
25. Radue EW, Barkhof F, Kappos L, et al. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. *Neurology* 2015; 84: 784–793.
26. Smith A. *Symbol Digit Modalities Test: Manual*. Los Angeles, CA: Western Psychological Services, 1982.
27. Creagh AP, Simillion C, Scotland A, et al. Smartphone-based remote assessment of upper extremity function for multiple sclerosis using the Draw a Shape Test. *Physiol Meas* 2020; 41: 054002.
28. Creagh AP, Simillion C, Bourke A, et al. Smartphone- and smartwatch-based remote characterisation of ambulation in multiple sclerosis during the two-minute walk test. *IEEE J Biomed Health Inform* 2020; 25: 838–849.
29. Bourke AK, Scotland A, Lipsmeier F, et al. Gait Characteristics harvested during a smartphone-based self-administered 2-minute walk test in people with multiple sclerosis: Test–retest reliability and minimum detectable change. *Sensors* 2020; 20: 5906.
30. Broen MP, Marsman VA, Kuijf ML, et al. Unraveling the relationship between motor symptoms, affective states and contextual factors in Parkinson's disease: A feasibility study of the experience sampling method. *PLoS ONE* 2016; 11: e0151195.
31. Storm FA, Nair KPS, Clarke AJ, et al. Free-living and laboratory gait characteristics in patients with multiple sclerosis. *PLoS ONE* 2018; 13: e0196463.
32. Taylor KI, Staunton H, Lipsmeier F, et al. Outcome measures based on digital health technology sensor data: Data- and patient-centric approaches. *NPJ Digit Med* 2020; 3: 97.