2023-01

# An observational study to assess validity and reliability of smartphone sensor-based gait and balance assessments in multiple sclerosis: Floodlight GaitLab protocol

Rinderknecht, MD



10.1177/20552076231205284
DIGITAL HEALTH
SAGE Publications

*Research Protocol*

DIGITAL HEALTH

DIGITAL HEALTH
Volume 9: 1–17

DOI: 10.1177/20552076231205284
journals.sagepub.com/home/dhj
Sage

# An observational study to assess validity and reliability of smartphone sensor-based gait and balance assessments in multiple sclerosis: Floodlight GaitLab protocol

**Mike D Rinderknecht[1,*], Mattia Zanon[1,*], Tjitske A Boonstra[2], Lorenza Angelini[1], Dimitar Stanev[1], Gabriela González Chan[3], Lisa Bunn[3], Frank Dondelinger[1], Richard Hosking[4], Jenny Freeman[3], Jeremy Hobart[5], Jonathan Marsden[3,**] and Licinio Craveiro[1,**]**

## Abstract

**Background:** Gait and balance impairments are often present in people with multiple sclerosis (PwMS) and have a significant impact on quality of life and independence. Gold-standard quantitative tools for assessing gait and balance such as motion capture systems and force plates usually require complex technical setups. Wearable sensors, including those integrated into smartphones, offer a more frequent, convenient, and minimally burdensome assessment of functional disability in a home environment. We developed a novel smartphone sensor-based application (Floodlight) that is being used in multiple research and clinical contexts, but a complete validation of this technology is still lacking.

**Methods:** This protocol describes an observational study designed to evaluate the analytical and clinical validity of Floodlight gait and balance tests. Approximately 100 PwMS and 35 healthy controls will perform multiple gait and balance tasks in both laboratory-based and real-world environments in order to explore the following properties: (a) concurrent validity of the Floodlight gait and balance tests against gold-standard assessments; (b) reliability of Floodlight digital measures derived under different controlled gait and balance conditions, and different on-body sensor locations; (c) ecological validity of the tests; and (d) construct validity compared with clinician- and patient-reported assessments.

**Conclusions:** The Floodlight GaitLab study (ISRCTN15993728) represents a critical step in the technical validation of Floodlight technology to measure gait and balance in PwMS, and will also allow the development of new test designs and algorithms.

## Keywords

Analytical validity, clinical validity, cognition, digital biomarkers, test-retest, upper limb

Submission date: 3 May 2023; Acceptance date: 18 September 2023

## Introduction

Multiple sclerosis (MS) is a chronic, immune-mediated disease of the central nervous system, characterized by inflammation, demyelination, axonal damage, and neuronal loss.[1] While the clinical presentation of MS is notably heterogeneous, walking impairment is observed in the vast majority of people with MS (PwMS),[2] and has been commonly reported as the most challenging MS symptom,[3] with a significant impact on quality of life, health status, social interactions, and productivity.[4] Impaired postural control is also often observed in MS,

[1]F. Hoffmann-La Roche Ltd, Basel, Switzerland
[2]F. Hoffmann-La Roche, Roche Nederland B.V., Woerden, Netherlands
[3]Faculty of Health, University of Plymouth, Plymouth, UK
[4]University of Plymouth Enterprise Ltd, Plymouth, UK
[5]Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK

*Shared first authors.

**Shared last authors.

**Corresponding author:**
Mattia Zanon, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, CH-4070 Basel, Switzerland.
Email: mattia.zanon@roche.com

with approximately two-thirds of PwMS reporting lack of balance and coordination as the main symptom affecting their mobility in daily living, reducing their independence, and increasing the risk of falls and injuries.[5]

Standardized measures such as the Expanded Disability Status Scale (EDSS),[6] the Timed 25-Foot Walk (T25FW),[7] the Timed Up and Go (TUG),[8] and the Berg Balance Scale (BBS)[9,10] are commonly used to assess walking, balance, and postural control in MS. However, some of these measures require a skilled examiner and can only be performed in a clinical environment, often providing a static snapshot that may have limited ecological validity.[11] Wearable sensors using inertial measurement units (IMUs) have been suggested as an alternative method to evaluate walking and postural control impairments.[11–14] Smartphone sensor-based technology in particular may offer a more frequent, cost-efficient, convenient, and minimally burdensome assessment of functional disability in a home environment, as well as providing a much larger set of measures to probe different aspects of functional ability.[15–18] Monitoring disease manifestations and trajectories across multiple parameters with frequent measurements is crucial in MS. This is particularly relevant in light of recent evidence that progression occurs across the spectrum of MS phenotypes and independent of relapse.[19–24]

One such example is Floodlight technology, which consists of a suite of smartphone sensor-based tests assessing gait (Two-Minute Walk Test (2MWT)),[25] static balance (Static Balance Test [SBT]), dynamic balance (U-Turn Test [UTT]),[26] upper extremity function, and cognition.[27,28] In a 24-week feasibility study, which investigated a precursor to Floodlight™ MS, smartphone sensor-derived digital measures (metrics) derived from the Floodlight tests showed moderate-to-good test–retest reliability and moderate-to-excellent agreement with standard clinical assessments of gait, upper extremity function, and cognition.[15,27] Additional research is currently ongoing to further explore construct validity and clinical utility,[29–31] but several questions remain unanswered. In this paper, we present a protocol of a study designed to explore as follows: (a) the concurrent validity of the Floodlight gait and balance tests against gold-standard assessments (motion capture systems and force plates) and silver-standard assessments (IMUs); (b) the reliability of Floodlight digital measures derived under different controlled gait and balance conditions, and different on-body sensor locations; (c) the ecological validity of the Floodlight gait and balance tests to estimate the extent to which they measure real-world walking and balance performance in PwMS; and (d) the construct validity compared with clinician- and patient-reported assessments (Figure 1).

## Methods

### Study objectives

This study is designed to evaluate different measurement properties of Floodlight smartphone-based digital measures of gait and balance. The main aims are as follows: (a) to determine how digital measures of gait derived from the Floodlight 2MWT compare with similar measures derived from a three-dimensional (3D) motion capture system and a foot-worn IMU system while walking straight, on a treadmill; (b) to determine how digital measures of dynamic balance derived from the Floodlight UTT compare with similar measures derived from a 3D motion capture system and a lower-back-worn IMU system; (c) to determine how the digital measures of gait and dynamic balance derived from the Floodlight 2MWT are impacted by the smartphone location (multiple waist and thigh locations), instrumentation (treadmill vs corridor walking), walking speed (self-selected comfortable vs fast-paced), cognitive-motor interference (single- vs dual-task paradigms), and time (test–retest reliability); (d) to determine how digital measures of gait and dynamic balance derived from the 2MWT and UTT collected on-site in a supervised environment compare with the 2MWT and UTT collected remotely in an unsupervised environment, and how both compare with similar measures derived from real-world daily walking; (e) to explore the association between Floodlight digital measures of gait and dynamic balance and clinical features of MS (e.g. disability level, spasticity, and fatigue); and (f) to determine how posturographic digital measures derived from the Floodlight SBT compare with similar measures derived from a force plate system, and explore their association with clinical features of MS (e.g. disability level and falls history).

*Primary objective.* In previous Floodlight studies, patients were required to carry their smartphone in a belt bag against their body near the waist.[15,27] However, in real-world environments, smartphones are typically carried off-body (e.g. in handbags), in the front/back pockets, and in the hand.[32] Recent studies have shown variable reliability and validity for most gait variables when sensors are placed at different locations on the body.[33,34] The primary objective of the study is therefore to explore the reliability of Floodlight gait and balance primary digital measures (see Table 1) when derived from smartphones placed at multiple waist and thigh positions of PwMS and healthy controls (HCs).

### Study design

In this prospective, single-center, observational study, each enrolled participant will attend two on-site visits (visit 1 and visit 2) and undergo a 2-week period of unsupervised

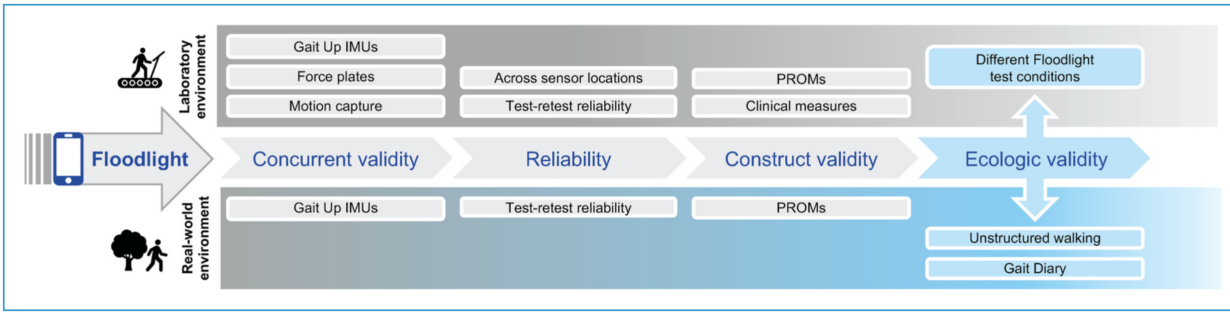


**Figure 1.** Psychometric evidence generation roadmap for the Floodlight GaitLab study. Methods and data sources used in the laboratory and real-world settings in the study are shown per the psychometric properties they help to demonstrate for Floodlight technology: concurrent validity, reliability, construct validity, and ecological validity. IMU: inertial measurement unit; PROM: patient-reported outcome measure.

remote testing in the home environment (remote testing period) (Table 2). During visit 1, an experienced neurologist (RH) with 15 years of clinical experience will collect the medical history of participants and perform a comprehensive set of standardized clinical assessments. Additionally, participants will undergo some performance-based assessments and a first set of supervised overground walking tests using Floodlight smartphone sensor-based technology and reference wearable sensors. Participants will then be given a detailed verbal explanation and an information booklet about the procedures to be performed during the remote unsupervised period. During the remote testing period, participants will perform daily tests without supervision (telephone support is available if required) using Floodlight technology and foot-worn sensors as a reference system, and will additionally collect contextual information via a Gait Diary; a comprehensive list of patient-reported outcome measures (PROMs) will also be completed (Appendix 1). Participants will then return to the site (visit 2), where gait and balance tests will be performed on a treadmill under different laboratory-controlled conditions using Floodlight technology, a motion capture system, and body-worn inertial sensors (reference systems). A detailed assessment schedule and the reference system used for each of the smartphone sensor-based tests are presented in Table 2.

## Participants and recruitment

The study aims to collect data from approximately 100 participants, including 75 PwMS and 25 HCs. A total of 100 PwMS and 35 HCs will be considered as the enrollment target to account for a potential 10% patient drop-out during the study and missing data due to technical problems during data acquisition. Participants will be recruited from a single site in Plymouth (UK) via their clinical team or by responding to advertisements (e.g. South West Impact of Multiple Sclerosis newsletter and Clinical Research Network for MS microsite). PwMS will be stratified into three approximately equal-sized groups according to their disability level (EDSS score ≤ 4.0, EDSS score 4.5–5.5, and EDSS score 6.0–6.5). Each group will include patients with different MS clinical phenotypes (relapsing-remitting, primary progressive, and secondary progressive MS). Although MS is more prevalent in women, the aim will be to balance each group for sex, given the known sex differences in gait kinematics in PwMS.[37] HCs will be matched at a population level to the MS participants according to age and sex distributions. To ensure adequate matching, HCs will be recruited in stages based on the PwMS already recruited. Recruitment began in January 2022 and the last visit of the last participant is expected in late 2023. All participants will be required to provide written informed consent before performing any study-related procedures. Detailed inclusion and exclusion criteria for both PwMS and HCs are presented in Table 3.

*Sample size.* While there is no explicit guidance for estimating sample sizes for this type of study,[41] sample sizes of approximately 25–50 participants have been commonly used for exploratory studies investigating the psychometric properties of measurement systems.[42–46] Additionally, according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines, a sample size of 100 participants, even with a small number of repeated measurements, allows for estimations of intraclass correlation coefficients (ICCs) and standard error of measurements with minimal bias and good precision.[47] Thus, a minimum sample size of 75 PwMS and 25 HCs is considered adequate to address the primary and secondary study objectives.

## Study equipment

*Floodlight technology.* The Floodlight technology will be installed on provisioned Samsung Galaxy A40 smartphones (running Android version 9), which have an in-built IMU that includes a triaxial accelerometer, a triaxial gyroscope,

**Table 1.** Primary and secondary gait and balance Floodlight digital measures.

| Digital measure class | Digital test | Digital measure | Definition |
|---|---|---|---|
| Primary | 2MWT | Step intensity | The intensity of walking calculated based on the variation in the amplitude of the accelerometer signal for each step |
| | UTT | Turn speed | The angular velocity while performing U-turns (rad/second)[26,27] |
| | SBT | Sway path | Sum of horizontal plane acceleration segments (path traced by the acceleration on the horizontal plane of the subject) (m/second$^2$)[27] |
| Secondary | 2MWT | Step number | Number of steps counted during a 2MWT (n) |
| | | Step time | Mean duration of a step (seconds) |
| | | Step frequency | Mean frequency of steps (Hz)[35] |
| | SBT | Sway jerk | Relative smoothness of postural sway (m$^2$/second$^5$)[36] |

2MWT: Two-Minute Walk Test; SBT: Static Balance Test; UTT: U-Turn Test.

and a magnetometer. Triaxial sensors provide simultaneous measurements in three orthogonal directions. Smartphone sensor data from the accelerometer and gyroscope will be collected at a sampling frequency of 50 Hz. This will enable the detection of the majority of frequency components of human body motion, which are typically below 10 Hz.[48,49] The accelerometer has a range of ±4 g and a resolution of 0.122 mg/Least Significant Bit (LSB), whereas the gyroscope has a range of ±1000 degrees per second and a resolution of 30.5 millidegrees per second/LSB.

In contrast to previous versions of the Floodlight technology (e.g. the Floodlight Proof-of-Concept app[15] or the Floodlight Open app[50]), research versions of the Floodlight technology will be used that are specifically tailored for collecting the smartphone sensor data outlined in this protocol.

### Reference systems

*Vicon® optical motion capture system.* The on-site motion capture laboratory is equipped with 12 infrared cameras (Vicon Vero™ v1.3 cameras with a resolution of 1.3 MP at 100 frames per second) and a dual-belt treadmill (Motek, Netherlands) on which the participants will perform the gait and balance tasks. The treadmill has an in-built self-paced algorithm, which allows participants to walk at a self-selected speed. Studies have shown that walking speed during self-paced treadmill walking was more similar to overground walking than fixed-speed treadmill walking.[51,52] To capture the participants' motion, 26 reflective markers (14 mm in diameter) will be positioned on body landmarks based on the Plug-in Gait marker set,[53] using the location naming conventions from the Human Body Model II reference manual.[54] In addition, clusters of three markers will be placed on six smartphone devices simultaneously worn at the different waist and thigh locations to obtain reference data on acceleration, angular velocity, and spatial orientation of the smartphones. Clusters of three markers will also be placed on both wrists as a proxy for a smartphone carried in the hand (see Figure 2). Tridimensional trajectories from the reflective markers will be recorded via the Vicon® and D-Flow motion capture software (Motek, Netherlands) at a sampling frequency of 100 Hz.

*Force plates.* Multiaxial force plates (ForceLink R-Mill force plates, Motek, Netherlands) embedded in the treadmill will record ground reaction forces (GRFs) and the moment of forces acting on the force plate along the $x$ (sideways)-, $y$ (vertical)-, and $z$ (running direction)-axes with load capacities of 5000 N, at a sampling rate of 1000 Hz. These measurements will be used to estimate the net center-of-pressure position during the treadmill walking tasks and during the SBT battery with a center-of-pressure error of ≤5 mm. The GRF will also be processed by Gait Offline Analysis Tool software (Motek, Netherlands) for gait events detection (e.g. heel strike and toe-off), which provides optimal processing for all treadmill-based gait data including the motion capture data acquired with the Vicon® cameras. Where a walking aid is used, the gait events will be identified using the marker-based method, as described above.

*Gait Up IMU system.* The Gait Up system (Gait Up, Lausanne, Switzerland) consists of Physilog® 6S units (size: 42.2 × 31.6 × 15.0 mm), which comprise a high-quality 3D accelerometer (range: ±8 $g$; sensitivity: 0.244 mg/LSB), a 3D gyroscope (range: ±2000 degrees

**Table 2.** Assessment schedule.

| | Visit 1 | Remote testing period | Visit 2 |
|---|---|---|---|
| Context of assessment | Laboratory (overground walking) | Real-world | Laboratory (overground walking; treadmill [motion capture laboratory]) |
| Duration | ~3 hours | 10–14 days | ~3 hours |
| Number of smartphone locations | 6[a] | 1[b] | 6[a] |
| Reference systems | • Gait Up IMUs (5 × Physilog® 6S units)<br>• Video recording | • Gait Up IMUs (2 × Physilog® 6S units) | • Vicon® motion capture system, including treadmill with embedded Forcelink force plates<br>• Gait Up IMUs (5 × Physilog® 6S units)<br>• Video recording |
| Walking and balance tests | • 2MWT (fast-paced; overground walking)<br>• T25FW (overground walking) | • 2MWT (fast-paced)<br>• UTT<br>• SBT<br>• Unstructured walking | • 2MWT (fast-paced; overground walking)<br>• T25FW (overground walking)<br>• 2MWT (all four conditions; treadmill walking)<br>• UTT (over treadmill)<br>• SBT (full battery; on treadmill) |
| Other Floodlight tests[c] | | • Pinching Test<br>• Draw a Shape Test<br>• Cognitive Test | • Pinching Test<br>• Draw a Shape Test<br>• Cognitive Test |
| Patient and disease characteristics | • Demographics<br>• Anthropometric measurements<br>• Self-reported MS history[d]<br>• Self-reported medical history<br>• Falls Questionnaire | | |
| Neurologic assessment[d] | • Neurologic examination (EDSS)<br>• Modified Ashworth Scale | | |
| Performance-based tests | • Cognition: Oral SDMT, CVLT-3, BVMT-R<br>• Upper extremity function: 9HPT | | |
| Patient-reported outcome measures[e] | • Borg Rating of Perceived Exertion[f] | • Floodlight Daily Mood Questionnaire<br>• Floodlight Symptom Tracker[g]<br>• MSSS-88<br>• MSWS-12 and MSWS-41<br>• MSIS-29v2<br>• ABC | • Borg Rating of Perceived Exertion[f] |

(continued)

**Table 2.** Continued.

| | Visit 1 | Remote testing period | Visit 2 |
|---|---|---|---|
| | | • FSMC<br>• Upper Limb Function Item Bank | |
| Other questionnaires | • Smartphone Location Questionnaire | • Floodlight Gait Diary<br>• Floodlight User Feedback Questionnaire | |
| Safety assessments | • Adverse event reporting | • Adverse event reporting | • Adverse event reporting |

[a]Smartphone sensor data will be collected simultaneously from six different locations as follows: right front pocket, left front pocket, right back pocket, left back pocket, central front, and back waist.
[b]A single smartphone will be carried during the active tests in a waist-worn belt bag, and as close to the body as possible (e.g. in a front pocket or in a waist-worn belt bag) during the unstructured walking (passive monitoring) activities.
[c]During the remote testing period, the Pinching Test and Draw a Shape Test will be taken daily alternating between hands, and the Cognitive Test once per week; during visit 2: the Pinching Test and Draw a Shape Test will be taken twice with each hand, and the Cognitive Test taken twice with the dominant hand.
[d]Completed by PwMS only.
[e]During the remote testing period, patient-reported outcome measures to be completed by PwMS only.
[f]Administered immediately before and after each 2MWT (structured overground walking and on treadmill) executed by the participant.
[g]The Floodlight Symptom Tracker will only be triggered every second week and, given that the observational period is up to 14 days, participants may complete it only once–the Floodlight Symptom Tracker can also be opened and completed on any day as needed by participants.
2MWT: Two-Minute Walk Test; 9HPT: Nine-Hole Peg Test; ABC: Activities-specific Balance Confidence scale; BVMT-R: Brief Visuospatial Memory Test–Revised; CVLT-3: California Verbal Learning Test–Third Edition; EDSS: Expanded Disability Status Scale; FSMC: Fatigue Scale for Motor and Cognitive functions; IMU: inertial measurement unit; MS: multiple sclerosis; MSIS-29vs2: 29-item Multiple Sclerosis Impact Scale version 2; MSSS-88: 88-item Multiple Sclerosis Spasticity Scale; MSWS-12: 12-item Multiple Sclerosis Walking Scale; MSWS-41: 41-item Multiple Sclerosis Walking Scale; PwMS: people with multiple sclerosis; SBT: Static Balance Test; SDMT: Symbol Digit Modalities Test; T25FW: Timed 25-Foot Walk; UTT: U-Turn Test.

per second; sensitivity: 70 millidegrees per second/LSB), a magnetometer (range: ±50 millitesla; resolution: 0.161–3.22 microtesla/LSB [along the *x*- and *y*-axes] and 0.294–5.87 microtesla/LSB [along the *z*-axis]), and a barometric pressure sensor (range: 260–1260 hPa; sensitivity: 4096 LSB/hPa). The Gait Up algorithms have been validated against motion capture systems, and their measurement accuracy is acceptable for the purposes of this study (i.e. yielding reliable results in normal and limping walking conditions, at a range of walking speeds between 0.9 and 2.0 m/second).[56] While the Gait Up IMU sensors support a range of sampling frequencies up to 512 Hz, a frequency of 128 Hz will be used in this study, as this offers a sufficient level of precision without using the additional battery life and memory required at higher sampling frequencies. These are important considerations in a real-world study in which participants may collect a large quantity of data between daily charging of devices and data download.

At the on-site visits, during the walking tasks on the treadmill and in the corridor, participants will wear five Gait Up IMU sensors placed on the right foot, left foot, lower back (near the body's center of mass), right wrist, and left wrist. Foot-worn sensors will primarily be used for gait event detection (e.g. heel strike and toe-off), whereas the sensor attached to the lower back will be used for more accurate turn detection, and wrist-worn sensors will allow measurement of inertial data comparable to those obtained with a smartwatch. During the remote testing period, only two foot-worn sensors will be used to reduce patient burden.

*Video recordings.* The participants will be videotaped while executing the different sensor-based tests during the two on-site visits. In full-body videos (i.e. those obtained from gait and balance tests), the faces of participants will be pixelated and sound will be removed with post-processing software to ensure pseudonymization of data. During upper extremity or cognition tests, participants will be videotaped from the axial plane with no recording of their faces.

*Synchronization of study equipment.* In order to annotate and synchronize data across the different measurement systems described above, a dedicated setup and methodology were developed for the on-site and remote settings. In particular, solutions were set up as follows: (a) to synchronize smartphones with each other and with the Gait Up IMU sensors;

**Table 3.** Inclusion and exclusion criteria.

| PwMS | HCs |
|---|---|
| **Inclusion criteria** | |
| • Signed informed consent form | • Signed informed consent form |
| • Ability to comply with the study protocol according to the investigator's judgment (in particular, having the ability to walk for a period of at least 6 minutes; rests are permitted as required) | • Ability to comply with the study protocol according to the investigator's judgment |
| • Age: ≥18 years (inclusive) | • Age: ≥18 years (inclusive) |
| • Body mass index:[a] <35 kg/m$^2$ | • Body mass index:[a] <35 kg/m$^2$ |
| • Confirmed diagnosis of MS, according to 2010 or 2017 McDonald criteria | |
| • Treatment with an approved or off-label disease-modifying treatment (or untreated) | |
| • EDSS of 0.0–6.5 (inclusive) | |
| **Exclusion criteria**[b] | |
| • Pregnancy[c] | • Pregnancy[c] |
| • Clinical relapse (self-reported or confirmed by their clinical team) in the past 60 days | • Ambulatory limitation, according to investigator's assessment (i.e. use of walking aids; musculoskeletal, orthopedic, vision, vestibular, cardiovascular, or neurologic deficits that could impair gait) |
| • Treatment with fampridine/dalfampridine (Fampyra®)/ Ampyra®) or other symptomatic MS treatment unless on stable dose for ≥30 days prior to screening | |
| • Self-reported change in rehabilitation protocol in the previous 60 days and during the study period | |
| • Treatment initiation with a disease-modifying therapy expected to occur in the course of the observation period for patients who are untreated at screening | |
| • Recovery from an infection or an intercurrent illness that may interfere with balance and gait according to the investigator's judgment | |
| • Uncorrected vision, musculoskeletal problems, marked vestibular deficits not caused by MS, or other non-MS neurologic problems that may interfere with balance and gait according to the investigator's judgment | |

[a]Given the known effects of higher body mass index on gait and balance biomechanics, patients with class 2 obesity and above are also excluded.[38,39]
[b]PwMS who experience acute disease activity (e.g. clinical relapse) or who are required to initiate a symptomatic or a disease-modifying treatment during the study will be discontinued.
[c]Significant changes in joint kinematics and center-of-pressure trajectories were previously reported for women during pregnancy.[40]
EDSS: Expanded Disability Status Scale; HC: healthy control; MS: multiple sclerosis; PwMS: people with multiple sclerosis.

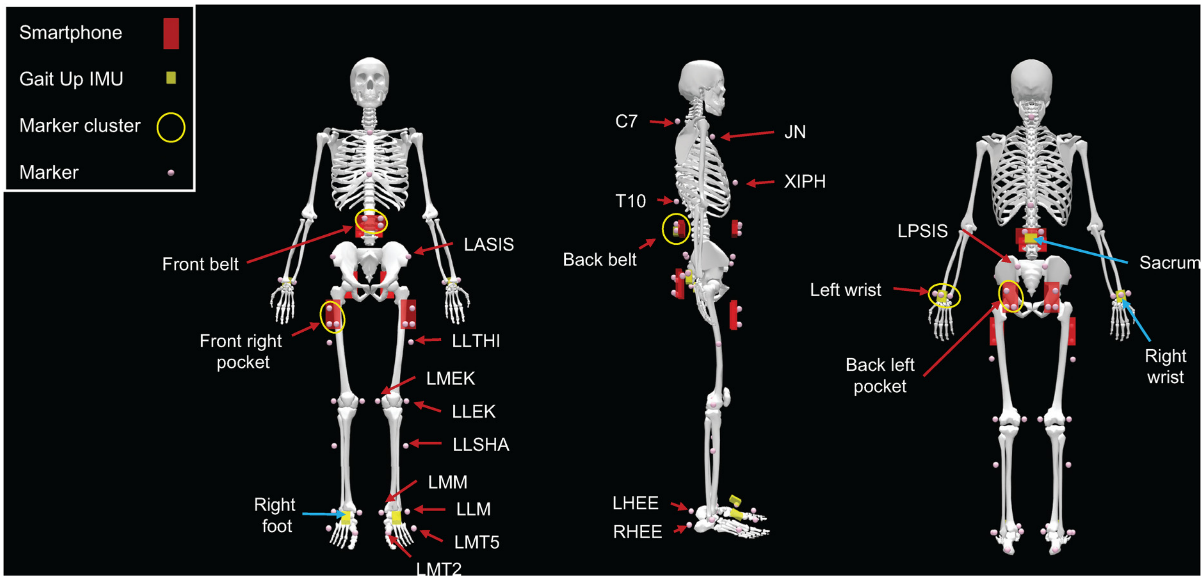


**Figure 2.** Locations of the Floodlight smartphones, Gait Up IMU sensors, and motion capture markers[53,54] used in the motion capture lab. C7: 7th cervical vertebra; IMU: inertial measurement unit; JN: jugular notch; LASIS: left anterior superior iliac spine; LHEE: left heel; LLEK: left lateral epicondyle of the knee; LLM: left lateral malleolus of the ankle; LLSHA: left shank (lateral); LLTHI: left thigh (lateral); LMEK: left medial epicondyle of the knee; LMM: left medial malleolus of the ankle; LMT2: left 2nd metatarsal; LMT5: left 5th metatarsal; RHEE: right heel; LPSIS: left posterior superior iliac spine; T10: 10th thoracic vertebra; XIPH: xiphoid process of the sternum. Right-side marker names (except for RHEE) were omitted for presentation purposes. Visualization created using OpenSim.[55]

(b) to annotate the active test type and condition; and (c) to synchronize smartphones or IMUs with the motion capture system. The technical setup was shown to align consistently with the time series obtained from the different systems and correct for clock drifts and other artifacts. For the test annotation, a pattern was embedded in the acceleration time series that is used to decode the test type and condition, and to ensure that it agrees with the annotation provided by the operator.

## Study assessments

*Demographics and medical history.* Demographics (year of birth, sex, and educational level), relevant medical history, and use of concomitant medications within 6 weeks prior to screening will be collected for all participants (PwMS and HCs). The following anthropometric measurements will also be collected: weight, height, iliac height (measured as the distance from the anterior superior iliac spine to the floor on both sides (left and right)), and foot length (measured as the distance between the posterior aspect of the heel and the longest toe measured along the foot axis). Additionally, the distance from the midpoint of each smartphone screen to the floor (smartphone-to-floor height) and the shoe sole length will be measured for calibration purposes of the smartphone and the Gait Up IMU sensors, respectively.

For PwMS, additional data will be collected, namely the date of MS onset and diagnosis, type of MS, recent relapse history, current and previous disease-modifying treatments, use of symptomatic medication that may affect ambulation (e.g. fampridine), and the use of a walking aid, orthotics, and functional electrical stimulation, including the type and frequency of use.

*Clinician- and performance-based assessments.* A complete neurologic examination will be performed in PwMS only, and will be captured systematically using an electronic version of the Neurostatus EDSS (MedAvante-ProPhase, Hamilton, NJ, USA).[57] Given the prevalence and impact of spasticity on gait and upper extremity biomechanics,[58,59] changes to muscle tone will be measured at all relevant joints using the modified Ashworth scale.[60] To overcome known inter- and intra-reliability deficiencies,[61–63] a single skilled rater will assess the participants using a standardized protocol in terms of test position, speed of movement, number of repetitions, and order of muscle groups testing.

Upper extremity function will be assessed in all participants using the Nine-Hole Peg Test, consisting of four trials, two for each hand.[64] Participants will be videotaped on the axial plane for observational motion analyses. Cognition will be evaluated using the Brief International Cognitive Assessment for Multiple Sclerosis battery,[65] which includes the oral Symbol

Digit Modalities Test,[66–68] the five immediate recall trials of the California Verbal Learning Test—Third Edition,[69] and the three immediate recall trials of the Brief Visuospatial Memory Test—Revised.[70]

*Floodlight smartphone sensor-based upper extremity and cognition active tests.* All participants will perform two active tests designed to measure upper extremity function (Pinching Test,[27] Draw a Shape Test[27,71]) and one to measure information processing speed (Cognitive Test[72]) during both the remote testing period and on-site visit 2. The tests will be performed for each hand (alternately using the dominant hand and non-dominant hand) with the smartphone placed on a table while stabilizing the device with the untested hand.

*PROMs.* PwMS will complete several PROMs once at a convenient time during the remote testing period. This includes PROMs that evaluate their ambulation (12-[73] and 41-item Multiple Sclerosis Walking Scales; the 41-item version was developed to allow consistent measurement across the different MS subtypes and overcome measurement limitations, namely range, precision, targeting, and relevance),[74] falls history (Falls Questionnaire),[75] spasticity (first three sections of the Multiple Sclerosis Spasticity Scale-88),[76] balance (Activities-specific Balance and Confidence Scale),[77–79] fatigue (Fatigue Scale for Motor and Cognitive Functions),[80] upper extremity function (upper limb function item bank),[81] and quality of life (version 2 of the 29-item Multiple Sclerosis Impact Scale).[82] PwMS will also be prompted to complete the Floodlight Daily Mood Questionnaire, which consists of two questions ("How is your mood now?" and "How are you feeling physically now?"), and the Floodlight Symptom Tracker, which will track any worsening of a predefined list of MS-related symptoms in the previous 2 weeks.

*Other questionnaires.* A number of additional questionnaires collecting contextual data, as well as information on usability and satisfaction, will be completed. The Smartphone Location Questionnaire will be used to record behavioral information on the usual/preferred location to carry their personal smartphone device at visit 1. The Floodlight Gait Diary, to be completed daily during the remote testing period, will be used to capture the wear location of the smartphone device during the gait and balance tests and the unstructured walking task, as well as the conditions in which the remote tests were taken as follows: (a) indoor or outdoor environment; (b) terrain surface and incline; (c) clothes worn tight/loose (which is relevant for understanding micro-movements of the smartphone in certain body locations); and (d) use of walking aids (Appendix 1). Finally, the Floodlight User Feedback Questionnaire will be administered to all participants and aims to collect information on their experience, satisfaction, usage, motivation, and acceptance when using the Floodlight technology.

## Experimental gait and balance protocols

*Structured and unstructured walking.* Laboratory-based gait assessments will be used to investigate the concurrent validity of the Floodlight 2MWT against gold- and silver-standard assessments (i.e. motion capture and Gait Up IMU sensors), the consistency of Floodlight digital measures of gait derived under different controlled walking conditions (e.g. walking speed and cognitive interference), and the reliability of digital measures of gait derived from smartphones placed on multiple on-body locations.

Real-world unstructured walking and real-world structured walking assessments will be used to investigate how real-world confounders (e.g. terrain surface/incline, curvilinear vs straight walking due to space limitations or obstacles) can affect the digital measures derived from Floodlight 2MWT, and to estimate the extent to which the Floodlight gait tests measure real-world walking performance in PwMS.[83,84]

In the Floodlight 2MWT,[25] participants are instructed to walk as fast and as far as they can for 2 minutes while walking safely, in a generally straight line (i.e. with minimal turns with a >90° angle). Variations of the 2MWT will be investigated, consistent with the objectives above (Table 2; Figure 3).

*Structured in-lab treadmill walking.* In the motion capture laboratory (visit 2), participants will be asked to walk for 2 minutes on a dual-belt treadmill, under the following conditions: (a) at a fixed slow pace of 2 km/hour; (b) at a comfortable self-selected pace by the participant; (c) at a self-selected fast pace; and (d) at a comfortable self-selected pace while performing a cognitive task, which consists of reporting out loud the intermediate results of serial subtractions of seven, starting from 200 (i.e. $200 - 7 = 193$, $193 - 7 = 186$, and etc.; similar tasks have been previously used in MS studies).[85–87] Each patient will start with the "fixed pace" condition (baseline), and the order of the remaining three conditions will be pseudorandomized. At the beginning of each self-paced condition, participants will have approximately 45 seconds in order to adjust to the new condition and reach their desired constant speed (Figure 3(a); Appendix 2).

Participants will be allowed, but not encouraged, to rest. If the participant stops walking during the test, the research team will say "You are doing well, you should keep walking if you are able", and the timer should not be stopped, in line with the original 2MWT instructions.[88,89] Patients are allowed to use their usual walking aid and/or orthotic if needed. Participants will be asked to rate their perceived exertion using the Borg Rating of Perceived

Exertion (RPE) scale,[90] before starting the 2MWT and then immediately after each 2MWT experimental condition. The Borg RPE has been shown to be a reliable and valid measure of perceived exertion in PwMS.[91]

Participants will carry six smartphones (weighing 140 g each) while walking, placed on six different positions in shorts and an adjustable waist belt customized by the research team as follows: right and left front pockets, central front at waist, left and right back pockets, and lower back. Five Gait Up IMU sensors (weighing 15 g each) will be attached to each foot, each wrist, and the lower back. 3D trajectories will be captured by 12 Vicon Vero™ cameras, and the piezoelectric force plates embedded onto the treadmill will also be used to capture GRF while walking (Figure 3(a)).

*Structured in-lab overground walking.* During visits 1 and 2, participants will be instructed to walk in a corridor for 2 minutes, back and forth between two marks on the floor placed 10 m apart. All participants will be instructed to walk safely at their fastest possible pace, with their walking aid and/or orthotic, if needed. The test will be performed under quiet conditions, with minimum distractions and corridor traffic, with the exception of one member of the research team who will walk behind the participant in case of loss of balance. In addition to the 2MWT, participants will perform the T25FW test on a marked 25-foot (7.62 m) course,[7] and will be instructed to walk safely at their fastest pace possible. The time (seconds) duration to complete the 25-foot course will be measured, and the average of the two trials will be computed. Participants will carry six smartphones and five Gait Up IMU sensors while walking (Figure 3(b)).

*Structured real-world overground walking.* During the remote testing period, the 2MWT is to be performed daily over a period of 10–14 days. Participants will be instructed to perform the fast-paced 2MWT with the instructions to avoid as many sharp turns (>90° angle) as possible, to safely walk on even and flat ground as quickly as possible, and, where feasible, to alternate daily between indoor and outdoor environments. Participants will be allowed to wear regular footwear and use a walking aid and/or orthotic if needed. Participants will be asked to carry one smartphone in a running belt positioned at waist level (front), and two Gait Up IMU sensors (one on each foot) while walking (Figure 3(c)). Furthermore, they will be asked to document contextual information in the Gait Diary (e.g. taken indoors or outdoors, in which location the smartphone was carried).

*Unstructured real-world walking.* Participants will be instructed to carry a smartphone as they go about their usual daily routine (any preferred body location) and will be expected to walk freely for between 15 minutes and 4 hours every day, over a period of 10–14 days. Smartphone sensor data will be passively collected during this activity (passive monitoring). Participants will also carry two Gait Up IMU sensors (one on each foot) while walking. Global Positioning System locations will be collected and anonymized by shifting the locations. Contextual information will also be collected in the Gait Diary (Figure 3(d)).

*Turning and dynamic balance.* The UTT assesses both gait and dynamic balance.[26] The user is instructed to walk back and forth between two points that are approximately 4 m apart, making a U-turn every time they reach one of the points.

*In-lab turning and dynamic balance.* Laboratory-based gait assessments will be used to investigate the concurrent validity of digital measures derived from the Floodlight UTT[26] against gold- and silver-standard assessments (including video recording for identification of start/end of turns, and motion capture and lumbar-worn IMU for turn speed and other motion measures), and the reliability of Floodlight digital measures of turning derived from smartphones placed on multiple on-body locations.

In the motion capture laboratory, participants will be instructed to walk back and forth using the static treadmill path, which is leveled with the ground, and perform as many U-turns as possible within 60 seconds just outside of the area at each end of the treadmill (this path was chosen to allow for the 3D trajectories to be captured with the motion capture system). Participants will carry six smartphones and five Gait Up units while performing the UTT (Figure 3(a)).

*Real-world turning and dynamic balance.* Real-world turning assessments will be used to investigate how real-world confounders (e.g. uneven ground) can affect digital measures derived from the Floodlight UTT. During the remote testing period, the UTT will be performed in the home environment. Participants will carry one smartphone in a waist-level running belt (front), and two Gait Up IMU sensors (one on each foot) while performing the UTT, and will be allowed to use a walking aid and/or orthotic as needed. Furthermore, they will be asked to document contextual information in the Gait Diary (Figure 3(c)).

*Static balance.* In the SBT,[27] users are asked to stand still unsupported with feet apart, eyes open, and relaxed arms straight alongside the body for 30 seconds, while the smartphone is kept in a running belt at the waist level (front).

*In-lab static balance.* In the laboratory environment, a modified battery of the SBT will be investigated, which consists of five 30-second tasks (each to be repeated twice) of increasing difficulty as follows: (a) natural

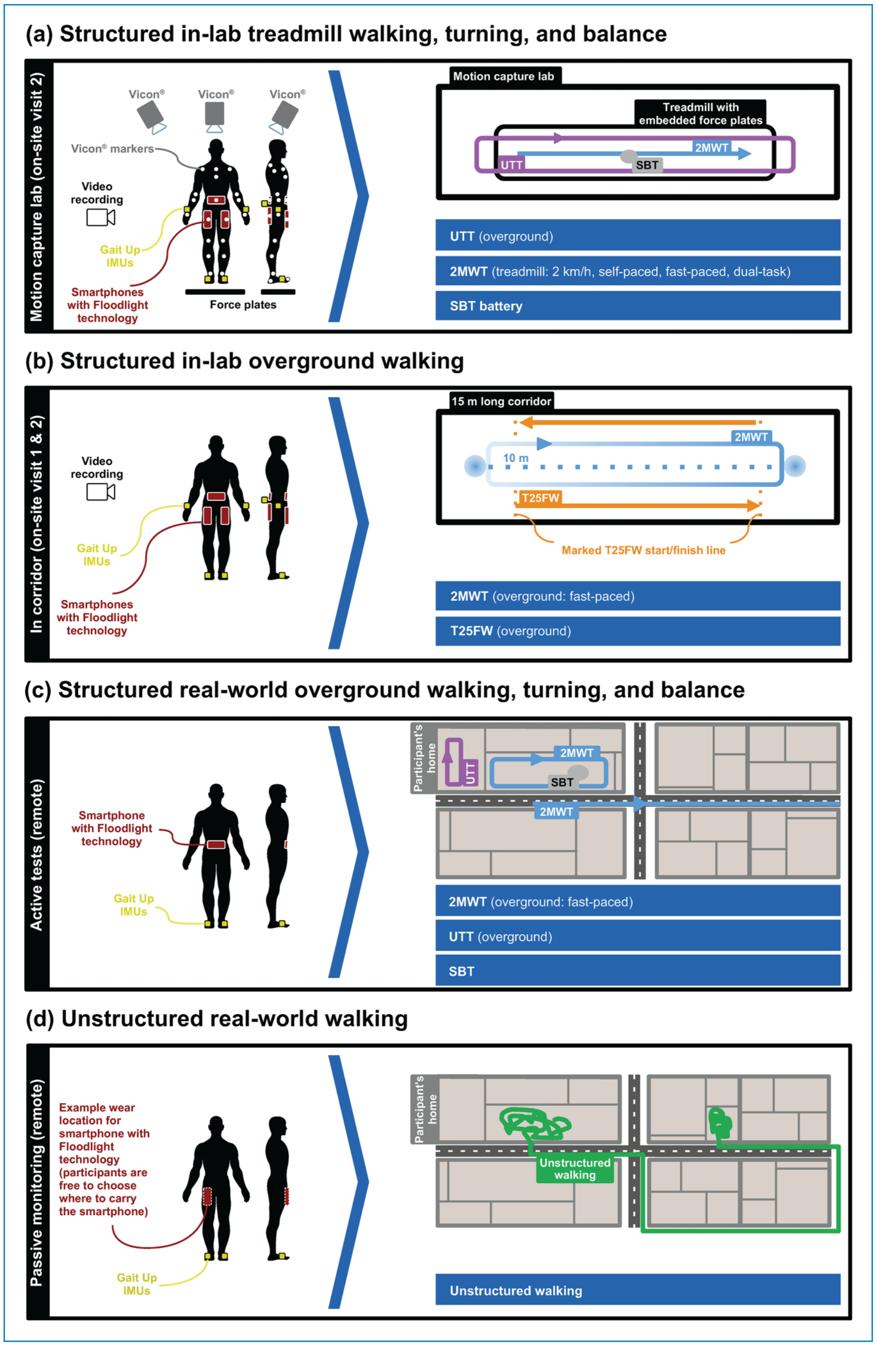


**Figure 3.** Setup for gait and balance tests administered during both on-site visits and during the remote testing period. (a) Structured in-lab treadmill walking, turning, and balance. In the motion capture laboratory (visit 2), participants will perform the 2MWT on a dual-belt treadmill, under four different conditions (here, Vicon® camera number and the marker positions are shown schematically; for full details, see section Study equipment: Reference systems). (b) Structured in-lab overground walking. During visits 1 and 2, participants will perform the 2MWT and the T25FW. Participants will carry six smartphones and five Gait Up IMU sensors while walking. (c) Structured real-world overground walking, turning, and balance. Participants will perform the fast-paced 2MWT and alternate daily between indoor and outdoor environments where feasible, as well as performing the UTT and SBT. Participants will be asked to carry one smartphone in a running belt positioned at waist level (front), and two Gait Up IMU sensors (one on each foot) while walking. Gray boxes indicate a participant's home and other buildings outside of the participant's home. (d) Unstructured real-world walking. Participants will be instructed to carry a smartphone as they go about their usual daily routine (any preferred body location). Participants will carry two Gait Up IMU sensors (one on each foot) while walking. Gray boxes indicate a participant's home and other buildings outside of the participant's home. 2MWT: Two-Minute Walk Test; IMU: inertial measurement unit; SBT: Static Balance Test; T25FW: Timed 25-Foot Walk; UTT: U-Turn Test.

stance with feet apart and eyes open; (b) natural stance with feet apart and eyes closed; (c) parallel stance with feet together and eyes open; (d) full tandem stance with eyes open; and (e) single foot stance with eyes open. For the last two tasks, the participants can choose which foot is in front or which foot they will stand on, respectively. Participants will only continue with narrower stance positions if deemed safe by the research team. This protocol recapitulates some of the stance positions of the Four-Stage Balance Test[92] and adds a condition of "eyes closed". Changes to visual guidance and reduction of the support surface size are two common strategies for making balance conditions increasingly challenging, and potentially more sensitive to detect postural control deficits.[93]

Participants will carry six smartphones and five Gait Up IMU sensors. Posturographic digital measures derived from the Floodlight SBT on each of the six smartphones will be compared with gold-standard posturographic measures derived from center-of-pressure trajectories collected with piezoelectric force plates embedded onto the treadmill (in static position), and from 3D trajectories of the motion capture markers located on the trunk and waist (Figure 3(a)).

*Real-world static balance.* During the remote testing period, participants will perform the SBT in the home environment with two Gait Up IMU sensors attached (one on each foot, to allow the recording of any steps used in recovering from imbalance). Some PwMS may be unable to safely perform this test unsupervised, a decision which will be made by the research team at visit 1 (Figure 3(c)).

## Data analysis

Continuous variables will be summarized with descriptive statistics as follows: (a) mean and standard deviation in case of normally distributed data; and (b) median, interquartile range, minimum and maximum values otherwise. Categorical variables will be summarized with frequency counts and percentages. All results will be presented separately for PwMS and HCs and for major predefined subgroups (e.g. EDSS categories).

To explore the concurrent validity of the primary gait and balance digital measures derived from Floodlight tests (see Table 1) against the reference systems, measures obtained by the different measurement devices/systems will be examined using Spearman's rank correlation. Correlations will be interpreted as very strong ($\rho \geq 0.8$), moderate ($0.6 \leq \rho < 0.8$), fair ($0.3 \leq \rho < 0.6$), and poor ($\rho < 0.3$).[94] These correlations will be separately calculated for each of the six smartphone locations and test conditions. Motion capture and foot-worn IMU sensors will be considered the gold- and silver-standard measures, respectively. Corrections for multiple testing will be applied in cases where Floodlight measures are compared with multiple measures of reference systems. For association pairs where Floodlight measures have a conceptual one-to-one match from the reference systems (e.g. turn speed during a UTT computed through Floodlight and through the motion capture system), inter-instrument agreement will be assessed by Bland–Altman plots,[95] limits of agreement, and inter-instrument ICC(2,1) for absolute agreement.[96]

Additionally, different modeling approaches, such as mixed-effect models, may be used to analyze the relationship between different reference system measures and each measure derived from the Floodlight technology. The same analysis approaches will be performed to assess the reliability and validity of the secondary and any potential exploratory digital measures that may be developed.

Test–retest reliability of the overground 2MWT performed in the corridor during on-site visits 1 and 2 will be assessed through ICC(2,1) for absolute agreement, and both systematic bias and outliers will be assessed through Bland–Altman plots.[95] Test–retest reliability of real-world assessments, performed during the remote testing period, will also be assessed through ICC(2,1).

Convergent and discriminant construct validity will be examined by computing Spearman's rank correlation coefficients between Floodlight digital measures and clinical assessment scores, performance-based assessments, patient-reported outcomes (PROs), and independent ratings obtained from the video recordings. The ability of Floodlight digital measures to differentiate between PwMS and HCs, or other predefined subgroups (known-groups construct validity) will be calculated using independent samples t-tests for continuous variables or appropriate non-parametric tests (e.g. Chi-square) for categorical/ordinal outcomes. PRO data will be analyzed through different approaches including Rasch Measurement Theory methods.[97,98]

Ecological validity will be analyzed through correlations between Floodlight measures obtained from the gait and balance assessments performed on-site versus remotely, and from the walking tests performed in the real world (structured walking) versus unstructured walking.

Other exploratory analyses may include exploring the proportion of individuals with the lowest and the highest possible scores for each testing paradigm of increasing difficulty (e.g. different stances of the SBT). Floor or ceiling effects imply that a measure cannot discriminate between subjects at either end of the scale.[99] These effects will be classified as significant if ≥15%, moderate if 10% to <15%, minor if 5% to <10%, and negligible if <5%.[100]

Finally, insights gained from this study will be used to conceptualize and implement algorithm improvements to further increase the reliability and validity of gait measurements obtained using Floodlight technology. The gait analysis algorithms will be assessed by comparing the times for the initial contact (heel strike) and final contact (toe-off) events with those obtained from the reference systems. The following metrics will be used: (a) precision (or "positive predictive value"; the number of true detected steps divided by the total number of detected steps); (b) recall (or "sensitivity"; the number of true detected steps divided by the total number of true steps); and (c) F-measure (or "F1 score"; the harmonic mean of precision and recall, thus defined as follows: F-measure $= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}$). Algorithm improvements will be assessed using the statistical measures mentioned above.

### Ethics, safety, and dissemination

The study is registered on the ISRCTN registry (ISRCTN15993728) and is funded by F. Hoffmann-La Roche Ltd, Basel, Switzerland. Ethical approval has been obtained from the UK Health Research Authority prior to study initiation (IRAS Project ID: 302099). Participant representatives were also included in the design of the study and reviewing of patient-facing documentation.

Several measures will be in place to ensure the safety of the participants performing the different tests and tasks. Participants will wear a safety harness to prevent falls during the treadmill tasks. During the gait tests performed in the corridor, if necessary, a member of the site staff will walk behind the participants to mitigate the risk of falls and resulting injuries. Furthermore, participants will be allowed to use their usual walking aid and/or orthotic. All participants will start with an unrecorded acclimatization trial on the treadmill, when they have the chance to get used to walking at a fixed speed and in a self-paced mode before the official trials commence. This also allows the research team to assess participants' safety and advise them on safe use of the treadmill aids/bars. In case their walking aid is not compatible with the treadmill, walking sticks will be provided if needed. Alternatively, they will be able to hold on to the sidebars of the treadmill, but this is the least preferred option. Finally, as the study may be impacted by the coronavirus disease 2019 (COVID-19) pandemic, any public health measures mandated by local authorities will be followed. Any decisions regarding participant enrollment in the study will be made with consideration of the potential impact of the COVID-19 public health emergency on participant safety.

While adverse events will not be actively solicited during this observational study, physicians and study participants will be reminded to report any adverse events, harm, or injury that might happen during the study period for which they suspect a direct causal role of the technology used in the study, or a specific study procedure or assessment performed by participants. To report any adverse events, study participants will have the ability to contact the study team either via phone during the study, or at the second on-site visit.

To ensure that participants are adequately supported, throughout the study they will have access to a telephone helpline and will be provided with a booklet containing instructions on how to perform all study-related tasks. Participants' adherence will also be monitored by the study team via the Floodlight study monitoring dashboard.

Confidentiality standards will be maintained by assigning each participant enrolled in the study a unique identification number. Participant data, including PROs and adverse events, will be recorded in electronic case report forms (eCRF) that will be periodically transferred to the study sponsor. Paper-based PROs will be digitalized and destroyed at the end of the study after the digital copy has been entered in the eCRF. All eCRF data will be stored on secure servers. Other data (e.g. Floodlight, motion capture, force place, and Gait Up) will be periodically transmitted to centralized platform(s) hosted and maintained by the study sponsor for processing, analysis, and storage, with only identified and trained users having access to the data.

Dissemination and analyses of study results and further details on technological solutions such as the synchronization of the several systems involved during the on-site and remote sensor-based assessments is expected in 2023 after study closure and will be published in peer reviewed journals and at congresses.

For up-to-date details on Roche's Global Policy on Sharing of Clinical Study Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing. Request for the data underlying this publication requires a detailed, hypothesis-driven, statistical analysis plan that is collaboratively developed by the requestor and company subject matter experts. Such requests should be directed to dbm.datarequest@roche.com for consideration. Anonymized records for individual patients across more than one data source external to Roche cannot, and should not, be linked due to a potential increase in risk of patient re-identification.

## Limitations

There are some foreseen limitations to this study. For instance, recall bias for the self-reported PRO data cannot be excluded. Another limitation is the fact that patients lost to follow-up and missing/incomplete Floodlight test data may impact sample representativeness. Limitations in terms of the study equipment and setup include that, for technical reasons, the motion capture system will only be available for the treadmill but not for the in-corridor environment. However, the impact of walking on a treadmill on Floodlight digital measures is currently unknown. Thus, to understand and compare the impact of treadmill walking versus overground walking, gait and balance tests will be performed both on the treadmill (with motion capture) and in the corridor (without motion capture), which will add to the complexity of the setup and the burden on the participants. Additionally, video recordings will not be available to assess the quality of the data or the correctness of the raw data during the remote testing period. Moreover, the gold-standard reference systems that the Floodlight digital measures will be assessed against will also be prone to error in certain instances. The use of multiple reference systems simultaneously (e.g. motion capture, Gait Up IMU sensors, and video recordings) aims to account for this, to allow for detection of any errors so that they can be factored in when analyzing the Floodlight test data for validity.

## Conclusions

The Floodlight GaitLab study (ISRCTN15993728) represents a critical step in the technical validation of Floodlight technology to measure gait and balance in PwMS. It aims to address several pertinent research questions as follows: ascertaining which disease concepts digital measures capture; how novel digital measures obtained in a laboratory setting translate to an unsupervised, real-world setting; and if these digital measures are sufficiently robust across different on-body sensor locations to allow users to choose where to carry the digital health technology tool (e.g. a smartphone) according to their preference. Thus, the insights generated by this study will help to support real-world use of the Floodlight technology and further reduce patient burden. Moreover, the study will investigate the concurrent validity, construct validity, ecological validity, and reliability of the digital measures.

The design of this study provides an example of incorporating both breadth of objectives (e.g. from analytical to ecological validity) and depth of assessment of a novel smartphone sensor-based application against multiple reference tools, into a short duration study that aims to minimize patient burden. The rich dataset that will be collected will enable the investigation of the impact of real-world confounders on Floodlight digital measures; allowing for the exploration of their validity, robustness, and consistency; and will aid in interpretation of these measures. This study will allow for the development of new test designs and algorithms and could help to inform future studies in the field to conduct and evaluate new digital health technology tools and digital measures.

**Acknowledgements:** The authors would like to thank all study participants and their families. We would also like to thank the following colleagues at F. Hoffmann-La Roche Ltd for their contributions to and support of the study: Matthias Bobst, Alan K. Bourke, Adrian Derungs, Sandro Fritz, Francisco Gavidia, Thomas Hahn, Petra Hauser, Sven Holm, Wiktoria Kasprzyk, Dominik Kedziora, Rafał Klimas, Kostas Kritsas, Hugo Le Gall, Dominik Lenart, Michael Lindemann, Florian Lipsmeier, Arnaud Mousley, Kathrin Müsch, Natan Napiorkowski, Madalina Ogica, Emanuele Passerini, Marta Płonka, Grégoire H. S. Pointeau, and Elena Spyridou. The authors would also like to thank colleagues at Plymouth for their contributions and support to the study: Michael Paisey, Tanya King, Joanne Lind, Carol Lunn, and Maureen Pedder. Writing and editorial assistance for this manuscript was provided by Frank Biegun, MSc, and Terri Desmonds, PhD, of Articulate Science, UK, and funded by F. Hoffmann-La Roche Ltd.

**Contributorship:** Study design: MDR, MZ, TAB, LA, DS, GGC, LB, JF, JH, JM, LC. Statistical analysis design: MDR, MZ, FD, LA, DS, LC. Data collection: JM, GGC, LB, RH. All authors reviewed and edited the manuscript drafts, and approved the final version of the manuscript.

**Declaration of Conflicting Interests:** The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: MDR is a contractor for F. Hoffmann-La Roche Ltd. MZ is an employee of and shareholder in F. Hoffmann-La Roche Ltd. TAB was an employee of Roche Netherlands B.V. during her contribution to this work and is now an employee of a Dutch nonprofit organization. She is a shareholder in F. Hoffmann-La Roche Ltd. LA is a consultant for F. Hoffmann-La Roche Ltd via Capgemini Engineering. DS is a consultant for F. Hoffmann-La Roche Ltd via Capgemini Engineering. GGC is an employee of the University of Plymouth, which received research funding from F. Hoffman-La Roche Ltd. LB is an employee of the University of Plymouth, which received research funding for this study from F. Hoffmann-La Roche Ltd. FD was an employee of F. Hoffmann-La Roche Ltd during his contribution to this work, and is now an employee of Novartis International AG. FD is a shareholder in F. Hoffmann-La Roche Ltd and Novartis International AG. RH has received indirect funding from F. Hoffmann-La Roche Ltd via a University of Plymouth Enterprise Ltd contract. JF is an employee of the University of Plymouth, which received research funding for this study from F. Hoffmann-La Roche Ltd. JH or affiliated institutions have received either consulting fees, honoraria, support to attend meetings, clinical service support, or research support from Acorda, Bayer Schering Pharma, Biogen Idec., Brickell Biotech, F. Hoffmann-La Roche Ltd, Global Blood Therapeutics, Sanofi Genzyme, Merck Serono, Novartis, Oxford Health Policy Forum, Teva, and Vantia. JM is an employee of the University of Plymouth, which received research funding for this study from F. Hoffmann-La Roche Ltd. LC is an employee of and shareholder in F. Hoffmann-La Roche Ltd.

**Ethical Approval:** Ethical approval was obtained from the UK Health Research Authority (IRAS Project ID: 302099).

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the F. Hoffmann-La Roche Ltd, Basel, Switzerland.

**Guarantor:** MZ.

**ORCID iDs:** Mike D Rinderknecht https://orcid.org/0000-0002-9825-8776
Mattia Zanon https://orcid.org/0000-0002-9875-6329
Dimitar Stanev https://orcid.org/0000-0001-8545-3110
Lisa Bunn https://orcid.org/0000-0002-7263-3048
Jeremy Hobart https://orcid.org/0000-0002-2114-7920
Licinio Craveiro https://orcid.org/0000-0002-8841-8035

**Supplemental Material:** Supplemental material for this article is available online.

## References

1. Reich DS, Lucchinetti CF and Calabresi PA. Multiple sclerosis. *N Engl J Med* 2018; 378: 169–180.
2. Comber L, Galvin R and Coote S. Gait deficits in people with multiple sclerosis: a systematic review and meta-analysis. *Gait Posture* 2017; 51: 25–35.
3. LaRocca NG. Impact of walking impairment in multiple sclerosis: perspectives of patients and care partners. *Patient* 2011; 4: 189–201.
4. Yildiz M. The impact of slower walking speed on activities of daily living in patients with multiple sclerosis. *Int J Clin Pract* 2012; 66: 1088–1094.
5. Prosperini L and Castelli L. Spotlight on postural control in patients with multiple sclerosis. *Degener Neurol Neuromuscul Dis* 2018; 8: 25–34.
6. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.
7. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 704–710.
8. Sebastião E, Sandroff BM, Learmonth YC, et al. Validity of the timed up and go test as a measure of functional mobility in persons with multiple sclerosis. *Arch Phys Med Rehabil* 2016; 97: 1072–1077.
9. Berg K. *Measuring balance in the elderly: development and validation of an instrument*. PhD Thesis, McGill University, Canada, 1992.
10. Cattaneo D, Regola A and Meotti M. Validity of six balance disorders scales in persons with multiple sclerosis. *Disabil Rehabil* 2006; 28: 789–795.
11. Shema-Shiratzky S, Hillel I, Mirelman A, et al. A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity. *J Neurol* 2020; 267: 1912–1921.
12. 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.
13. Frechette ML, Meyer BM, Tulipani LJ, et al. Next steps in wearable technology and community ambulation in multiple sclerosis. *Curr Neurol Neurosci Rep* 2019; 19: 80.
14. Tulipani LJ, Meyer B, Larie D, et al. Metrics extracted from a single wearable sensor during sit-stand transitions relate to mobility impairment and fall risk in people with multiple sclerosis. *Gait Posture* 2020; 80: 361–366.
15. 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.
16. Boukhvalova AK, Kowalczyk E, Harris T, et al. Identifying and quantifying neurological disability via smartphone. *Front Neurol* 2018; 9: 740.
17. Arteaga-Bracho EE, Dai Y, Drory V, et al. Konectom™ smartphone-based digital outcome assessments for adults living with spinal muscular atrophy (SMA): a conceptual framework. *Neurology* 2021; 96(15 Supplement): 2701.
18. Abou L, Wong E, Peters J, et al. Smartphone applications to assess gait and postural control in people with multiple sclerosis: a systematic review. *Mult Scler Relat Disord* 2021; 51: 102943.
19. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014; 83: 278–286.
20. Lublin FD and Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology* 1996; 46: 907–911.
21. Lassmann H, Van Horssen J and Mahad D. Progressive multiple sclerosis: pathology and pathogenesis. *Nat Rev Neurol* 2012; 8: 647–656.
22. University of California San Francisco MS-EPIC Team, Cree BA, Hollenbach JA, et al. Silent progression in disease activity–free relapsing multiple sclerosis. *Ann Neurol* 2019; 85: 653–666.
23. Elliott C, Wolinsky JS, Hauser SL, et al. Slowly expanding/evolving lesions as a magnetic resonance imaging marker of chronic active multiple sclerosis lesions. *Mult Scler* 2019; 25: 1915–1925.
24. 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.
25. 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 (Basel)* 2020; 20: 5906.
26. Cheng W-Y, Bourke AK, Lipsmeier F, et al. U-turn speed is a valid and reliable smartphone-based measure of multiple sclerosis-related gait and balance impairment. *Gait Posture* 2021; 84: 120–126.
27. Montalban X, Graves J, Midaglia L, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. *Mult Scler* 2022; 28: 654–664.
28. van der Walt A, Butzkueven H, Shin RK, et al. Developing a digital solution for remote assessment in multiple sclerosis: from concept to software as a medical device. *Brain Sci* 2021; 11: 1247.
29. van der Walt A, Butzkueven H, Barnett M, et al. A prospective study of the feasibility of smartphone-based self-monitoring to characterise cognitive and neurological impairment in people with multiple sclerosis: Floodlight MS MoreActive. *Mult Scler* 2022; 28(3 Supplement): EP0886.

30. Rukina D, Bogaarts G, Dondelinger F, et al. Novel smartphone sensor-based scores for remote measurement of gait and hand function impairment in people with MS. *Mult Scler* 2021; 27(2 Supplement): P306.
31. ISRCTN registry. ISRCTN11088592. A study to investigate the concordance of smartphone-based self-monitoring, imaging, and blood-based biomarkers with clinical disability in participants with multiple sclerosis, https://doi.org/10.1186/ISRCTN11088592 (2021, accessed January 2023).
32. Redmayne M. Where's your phone? A survey of where women aged 15–40 carry their smartphone and related risk perception: a survey and pilot study. *PLoS One* 2017; 12: e0167996.
33. Vienne-Jumeau A, Quijoux F, Vidal P-P, et al. Wearable inertial sensors provide reliable biomarkers of disease severity in multiple sclerosis: a systematic review and meta-analysis. *Ann Phys Rehabil Med* 2020; 63: 138–147.
34. Tietsch M, Muaremi A, Clay I, et al. Robust step detection from different waist-worn sensor positions: implications for clinical studies. *Digit Biomark* 2020; 4: 50–58.
35. Ganzetti M, Graves JS, Holm SP, et al. Neural correlates of digital measures shown by structural MRI: a post-hoc analysis of a smartphone-based remote assessment feasibility study in multiple sclerosis. *J Neurol* 2023; 270: 1624–1636.
36. Mancini M, Salarian A, Carlson-Kuhta P, et al. ISway: a sensitive, valid and reliable measure of postural control. *J Neuroeng Rehabil* 2012; 9: 59.
37. Pau M, Corona F, Pilloni G, et al. Do gait patterns differ in men and women with multiple sclerosis? *Mult Scler Relat Disord* 2017; 18: 202–208.
38. Silva FR, de Souza Muniz AM, Cerqueira LS, et al. Biomechanical alterations of gait on overweight subjects. *Res Biomed Eng* 2018; 34: 291–298.
39. Del Porto H, Pechak C, Smith D, et al. Biomechanical effects of obesity on balance. *Int J Exerc Sci* 2012; 5: 301–320.
40. Mei Q, Gu Y and Fernandez J. Alterations of pregnant gait during pregnancy and post-partum. *Sci Rep* 2018; 8: 2217.
41. Hobart J, Cano S, Warner T, et al. What sample sizes for reliability and validity studies in neurology? *J Neurol* 2012; 259: 2681–2694.
42. Lefeber N, Degelaen M, Truyers C, et al. Validity and reproducibility of inertial Physilog sensors for spatiotemporal gait analysis in patients with stroke. *IEEE Trans Neural Syst Rehabil Eng* 2019; 27: 1865–1874.
43. Keppler AM, Nuritidinow T, Mueller A, et al. Validity of accelerometry in step detection and gait speed measurement in orthogeriatric patients. *PLoS One* 2019; 14: e0221732.
44. Romijnders R, Warmerdam E, Hansen C, et al. Validation of IMU-based gait event detection during curved walking and turning in older adults and Parkinson's disease patients. *J Neuroeng Rehabil* 2021; 18: 28.
45. Werner C, Chalvatzaki G, Papageorgiou XS, et al. Assessing the concurrent validity of a gait analysis system integrated into a smart walker in older adults with gait impairments. *Clin Rehabil* 2019; 33: 1682–1687.
46. Werner C, Heldmann P, Hummel S, et al. Concurrent validity, test–retest reliability, and sensitivity to change of a single body-fixed sensor for gait analysis during rollator-assisted walking in acute geriatric patients. *Sensors (Basel)* 2020; 20: 4866.
47. Mokkink LB, de Vet H, Diemeer S, et al. Sample size recommendations for studies on reliability and measurement error: an online application based on simulation studies. *Health Serv Outcomes Res Methodol* 2023; 23: 241–265.
48. Antonsson EK and Mann RW. The frequency content of gait. *J Biomech* 1985; 18: 39–47.
49. Zeng H and Zhao Y. Sensing movement: microsensors for body motion measurement. *Sensors (Basel)* 2011; 11: 638–660.
50. van Beek J, Freitas R, Bernasconi C, et al. Floodlight open—a global, prospective, open-access study to better understand multiple sclerosis using smartphone technology. In: *Annual Meeting of the Consortium of Multiple Sclerosis Centers (CMSC)*, Seattle, Washington, USA, 28 May–1 June 2019.
51. Sloot LH, van der Krogt MM and Harlaar J. Self-paced versus fixed speed treadmill walking. *Gait Posture* 2014; 39: 478–484.
52. Ibala E, Coupaud S and Kerr A. Comparison of the muscle pattern variability during treadmill walking (fixed and self-pace) and overground walking of able-bodied adults. *J Ann Bioeng* 2019; 1: 1–11.
53. Vicon Motion Systems Limited. Full body modeling with Plug-in Gait - Nexus 2.13 documentation, https://docs.vicon.com/display/Nexus213/Full+body+modeling+with+Plug-in+Gait (2023, accessed January 2023).
54. Motek Medical BV. HBM2 reference manual, https://knowledge.motekmedical.com/wp-content/uploads/2019/07/HBM2-Reference-Manual-Full-Body.pdf (2019, accessed January 2023).
55. Delp SL, Anderson FC, Arnold AS, et al. Opensim: open-source software to create and analyze dynamic simulations of movement. *IEEE Trans Biomed Eng* 2007; 54: 1940–1950.
56. Schwameder H, Andress M, Graf E, et al. Validation of an IMU-System (Gait-up) to identify gait parameters in normal and induced limping walking conditions. In: *International Conference on Biomechanics in Sports (ISBS)*, Poitiers, France, 29 June–3 July 2015, pp.707–710. San Diego: ISBS.
57. Kappos L. Neurostatus scoring definitions: Version 04/10.2, https://www.neurostatus.net/media/specimen/Definitions_0410-2_s.pdf (2011, accessed January 2023).
58. Milinis K, Tennant A, Young C, et al. Spasticity in multiple sclerosis: associations with impairments and overall quality of life. *Mult Scler Relat Disord* 2016; 5: 34–39.
59. Pau M, Coghe G, Corona F, et al. Effect of spasticity on kinematics of gait and muscular activation in people with multiple sclerosis. *J Neurol Sci* 2015; 358: 339–344.
60. Bohannon RW and Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987; 67: 206–207.
61. Baunsgaard CB, Nissen UV, Christensen KB, et al. Modified Ashworth scale and spasm frequency score in spinal cord injury: reliability and correlation. *Spinal Cord* 2016; 54: 702–708.
62. Craven BC and Morris AR. Modified Ashworth scale reliability for measurement of lower extremity spasticity among patients with SCI. *Spinal Cord* 2010; 48: 207–213.
63. Nuyens G, De Weerdt W, Ketelaer P, et al. Inter-rater reliability of the Ashworth scale in multiple sclerosis. *Clin Rehabil* 1994; 8: 286–292.

64. Feys P, Lamers I, Francis G, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler* 2017; 23: 711–720.
65. Benedict RH, Amato MP, Boringa J, et al. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. *BMC Neurol* 2012; 12: 55.
66. Smith A. The Symbol-Digit Modalities Test: a neuropsychologic test of learning and other cerebral disorders. In: Helmuth J (eds) *Learning disorders*. Seattle, WA: Special Child Publications, 1968, pp.83–91.
67. Smith A. *Symbol Digit Modalities Test: manual*. Los Angeles, CA: Western Psychological Services, 1982.
68. Benedict RH, DeLuca J, Phillips G, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 721–733.
69. Delis DC, Kramer JH, Kaplan E, et al. *California Verbal Learning Test (CVLT-3)*. London: Pearson, 2017.
70. Benedict RHB. *Brief visuospatial memory test—revised: professional manual*. Lutz, FL: Psychological Assessment Resources, Inc, 1997.
71. Graves JS, Ganzetti M, Dondelinger F, et al. Preliminary validity of the Draw a Shape Test for upper extremity assessment in multiple sclerosis. *Ann Clin Trans Neurol* 2023; 10: 166–180.
72. Dondelinger F, Thomann AE, Volkova-Volkmar E, et al. A digital remote assessment for measuring impairment in information processing speed in people with MS. *Mult Scler* 2021; 27(2 Supplement): P303.
73. Hobart JC, Riazi A, Lamping DL, et al. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). *Neurology* 2003; 60: 31–36.
74. Hobart J and Burke L. Advancing walking measurement in multiple sclerosis clinical trials: a new patient-reported outcome measure. *Mult Scler* 2019; 25: 216.
75. Hoang PD, Cameron MH, Gandevia SC, et al. Neuropsychological, balance, and mobility risk factors for falls in people with multiple sclerosis: a prospective cohort study. *Arch Phys Med Rehabil* 2014; 95: 480–486.
76. Hobart JC, Riazi A, Thompson AJ, et al. Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain* 2006; 129: 224–234.
77. Myers AM, Powell LE, Maki BE, et al. Psychological indicators of balance confidence: relationship to actual and perceived abilities. *J Gerontol A Biol Sci Med Sci* 1996; 51: M37–M43.
78. Myers AM, Fletcher PC, Myers AH, et al. Discriminative and evaluative properties of the activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci* 1998; 53: M287–M294.
79. Powell LE and Myers AM. The activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci* 1995; 50A: M28–M34.
80. Penner IK, Raselli C, Stöcklin M, et al. The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. *Mult Scler* 2009; 15: 1509–1517.
81. Hobart J, King T, Close J, et al. Measuring patient-reported upper limb function in multiple sclerosis clinical trials: can item banks achieve measurement nirvana? *Mult Scler* 2021; 27: 123.
82. Hobart J, Lamping D, Fitzpatrick R, et al. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain* 2001; 124: 962–973.
83. Mikolaizak AS, Rochester L, Maetzler W, et al. Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement—the Mobilise-D study protocol. *PLoS One* 2022; 17: e0269615.
84. Mate KK and Mayo NE. Clinically assessed walking capacity versus real-world walking performance in people with multiple sclerosis. *Int J MS Care* 2020; 22: 143–150.
85. Sosnoff JJ, Boes MK, Sandroff BM, et al. Walking and thinking in persons with multiple sclerosis who vary in disability. *Arch Phys Med Rehabil* 2011; 92: 2028–2033.
86. Wajda DA, Motl RW and Sosnoff JJ. Dual task cost of walking is related to fall risk in persons with multiple sclerosis. *J Neurol Sci* 2013; 335: 160–163.
87. Postigo-Alonso B, Galvao-Carmona A, Benitez I, et al. Cognitive-motor interference during gait in patients with multiple sclerosis: a mixed methods systematic review. *Neurosci Biobehav Rev* 2018; 94: 126–148.
88. Goldman MD, Marrie RA and Cohen JA. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. *Mult Scler* 2008; 14: 383–390.
89. Gijbels D, Eijnde B and Feys P. Comparison of the 2-and 6-minute walk test in multiple sclerosis. *Mult Scler* 2011; 17: 1269–1272.
90. Williams N. The Borg Rating of Perceived Exertion (RPE) scale. *Occup Med* 2017; 67: 404–405.
91. Cleland BT, Ingraham BA, Pitluck MC, et al. Reliability and validity of ratings of perceived exertion in persons with multiple sclerosis. *Arch Phys Med Rehabil* 2016; 97: 974–982.
92. Centers for Disease Control and Prevention. Assessment The 4-Stage Balance Test, https://www.cdc.gov/steadi/pdf/4-Stage_Balance_Test-print.pdf (2017, accessed January 2023).
93. Nejc S, Jernej R, Loefler S, et al. Sensitivity of body sway parameters during quiet standing to manipulation of support surface size. *J Sports Sci Med* 2010; 9: 431–438.
94. Chan Y. Biostatistics 104: correlational analysis. *Singapore Med J* 2003; 44: 614–619.
95. Bland JM and Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327: 307–310.
96. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.
97. Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007; 6: 1094–1105.
98. Wright BD and Masters GN. *Rating scale analysis*. Chicago, IL: MESA Press, 1982.
99. Lim CR, Harris K, Dawson J, et al. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015; 5: e007765.
100. Gulledge CM, Smith DG, Ziedas A, et al. Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains among shoulder and knee patients undergoing nonoperative and operative treatment. *JB JS Open Access* 2019; 4: e0015.1–7.