Stride-level measurement of gait as an early sensitive marker of disability progression in ambulatory patients with multiple sclerosis

Margaux Poleur,[a] Barbara Willekens,[b,c] Bertrand Degos,[d] Damien Ricard,[e] Vincent van Pesch,[f] Annick Mélin,[g] Oihana Piquet,[h] Alexis Tricot,[h] Laurie Médard,[a] Mona Michaud,[h] Emilie Lommers,[i] Anna-Victoria De Keersmaecker,[b,c] Irene Coman,[d] James Overell,[j,k] Alexandra Goodyear,[l] Paul Delmar,[j] Qing Wang,[j] Helen Hayward-Koennecke,[j] Céline Cluzeau,[h] Damien Eggenspieler,[h] Paul Strijbos,[j] and Laurent Servais[m,n,*]

[a]University Department of Neurology, Citadelle Hospital of Liège, Liège, Belgium
[b]Department of Neurology, Antwerp University Hospital, Antwerp, Belgium
[c]Faculty of Medicine and Health Sciences, Translational Neurosciences Research Group, University of Antwerp, Antwerp, Belgium
[d]Neurology Department, Avicenne Hospital, APHP, Hôpitaux Universitaires de Paris-Seine Saint Denis (HUPSSD), Bobigny, France
[e]Service de Neurologie, Service de Santé des Armées, Hôpital d'Instruction des Armées de Percy, Clamart, France
[f]Cliniques Universitaires Saint-Luc, UCLouvain, Louvain, Belgium
[g]Departement of Neurology, CHC Mont-Legia, Liège, Belgium
[h]Sysnav, Vernon, France
[i]Department of Neurology, Centre Hospitalier Universitaire de Liège, Liège, Belgium
[j]F. Hoffmann-La Roche Ltd., Basel, Switzerland
[k]Department of Clinical Neuroscience, School of Medicine, Dentistry and Nursing, University of Glasgow, UK
[l]Genentech, South San Francisco, CA, USA
[m]MDUK Oxford Neuromuscular Centre, & NIHR, University of Oxford, UK
[n]Neuromuscular Reference Center, Department of Pediatrics, University Hospital Liege & University of Liege, Belgium

eClinicalMedicine
2026;93: 103823
Published Online xxx
https://doi.org/10.1016/j.eclinm.2026.103823

Summary

Background Wearable digital health technologies offer a unique opportunity to assess gait at the stride level in real-world settings. Walking impairment is a major cause of disability in multiple sclerosis (MS), yet current clinical metrics lack sensitivity to early and progressive changes in mobility.

Methods We conducted two studies (NCT04888689/NCT04882891) using a wearable device to develop and validate digital mobility outcome measures based on individual strides in patients with MS. First, we assessed technical performance in a controlled, single-center environment between September 12 and September 18, 2021. We then conducted a 12-month longitudinal study under daily living conditions across six sites between March 2021 and January 2024. The evaluated metrics included stride velocity 95th centile, walking distance 90th centile, and strides per hour.

Findings The controlled and longitudinal studies included 21 and 78 participants, respectively. The device demonstrated high stride detection accuracy (precision: 0·99) and a mean absolute error in stride velocity of 0·019 m/s. In the longitudinal study, stride velocity 95th percentile showed excellent reliability (ICC (2,1) = 0·97, SEM = 0·06) and strong agreement with Expanded Disability Status Scale (Spearman's rho = 0·65, $p < 0·001$) and Timed 25-Foot Walk (Spearman's rho = –0·71, $p < 0·001$), sensitivity to 12-month progression in both relapsing–remitting and progressive MS ($p = 0·049$ and $p = 0·006$, respectively), outperforming the Expanded Disability Status Scale. Walking distance 90th percentile and strides per hour were reliable and valid but less sensitive to progression.

Interpretation Stride velocity 95th percentile derived from real-world, stride-level data, provides a valid, reliable, and sensitive digital outcome for detecting MS progression. It may serve as an early indicator of progression and support the accelerated evaluation of treatments targeting progression.

Funding This study was funded by F. Hoffmann-La Roche Ltd.

*Corresponding author. Neuromuscular Reference Center, Hopital de La Citadelle, 1 Bvd du XII de Ligne, 4000, Liège, Belgium.
E-mail address: laurent.servais@paediatrics.ox.ac.uk (L. Servais).

Keywords: Wearable digital health technology; Stride-level data; Multiple Sclerosis; Stride Velocity 95th centile

Research in context

Evidence before this study

We searched PubMed and ClinicalTrials.gov up to November 2025 for studies evaluating digital outcomes to monitor ambulation in people with multiple sclerosis (MS), using search terms including "digital mobility outcomes", "wearable sensors", "gait analysis", and "multiple sclerosis". Most studies reporting the use of wearable digital health technology (wDHT) to assess ambulation in MS have explored step counts and physical activity monitoring using consumer-grade actimeters or smartphone-based tools. Some digital outcomes demonstrate promising metric properties, including known-group and convergent validity, but only a few exhibit sensitivity to change. Regulatory agencies have expressed interest in digital gait-speed outcomes, but no digital endpoint has yet achieved qualification for MS trials. Stride-level metrics, such as stride-velocity percentiles, have been qualified only in neuromuscular diseases such as Duchenne muscular dystrophy and have not been systematically investigated in MS.

Added value of this study

This is the first study to evaluate stride-level metrics derived from a validated wDHT in ambulatory patients with MS across both controlled and real-world conditions. We demonstrate the analytical validity, high reliability, and sensitivity to change of stride velocity 95th centile (SV95C), walking-distance 90th centile (WD90C), and strides per hour as digital outcomes. SV95C, in particular, shows strong psychometric properties and the ability to detect 1-year ambulation decline in both progressive and relapsing-remitting MS, outperforming standard clinical measures such as the Expanded Disability Status Scale and the Timed 25-Foot Walk in this population.

Implications of all the available evidence

Digital outcomes based on individual stride analysis may represent a promising innovation for objectively monitoring MS-related disability progression. In particular, SV95C appears to be a viable candidate for use as a surrogate or exploratory endpoint in future clinical trials, potentially enabling earlier and more sensitive detection of ambulation decline than conventional assessments. These findings should, however, be interpreted with caution, as their broader applicability depends on confirmation in larger, independent cohorts.

Introduction

Over recent years, a clear trend towards the use of wearable technologies has emerged across many neurological conditions.[1] Digital outcomes allow objective, continuous, and precise quantification of disease progression.[2] The higher temporal resolution and real-world relevance of digital outcomes potentially provide significant advantages over conventional assessments, which are often limited to periodic in-clinic visits and subjective evaluations. Nevertheless, few digital outcomes have been used as primary or secondary endpoints in clinical development in neurology, and even fewer have reached the final step in regulatory qualification pathways.[3] In the context of ambulation, most outcomes are captured by consumer-available actimeters and assess the number of steps, with limited ability to reliably access stride-level data.[4] Advances in inertial sensors and signal analysis have improved the precision of human movement analysis.[5] This provides a unique opportunity for the identification and measurement of individual strides during normal daily living[6] rather than relying on estimates of patient activity. This represents significant advancement for wearable technology beyond the simple metric of "steps per day" or even walking-bouts metrics.

Shifting from basic step counting to detailed stride analysis potentially provides a deeper and more clinically valuable understanding of human ambulation. Stride velocity 95th centile (SV95C) represents the spontaneous 5% fastest stride in daily living. It was qualified in 2023 by the European Medicines Agency for use as a primary endpoint in clinical trials for Duchenne muscular dystrophy (DMD) provided it is measured using a valid and suitable wearable digital health technology (wDHT).[3] This first qualification of a digital outcome for use as a primary endpoint in clinical trials demonstrates that regulatory agencies recognize that well-validated digital endpoints can capture the clinical experience and consider their use in the drug approval process.[3] This technology opens opportunities in other neurological conditions where progressive loss of ambulation is a key component, such as multiple sclerosis (MS). Clinically relevant concepts of interest in MS, such as gait speed,[7] could be investigated more precisely, enabling earlier detection of impairment.

MS is an acquired inflammatory demyelinating disorder of the central nervous system that affects 2·9 million individuals worldwide[8] with an increasing prevalence. It is characterized by inflammation and neurodegeneration, leading to a broad spectrum of clinical manifestations and levels of disability with gait

impairment as one of the most prevalent MS symptoms.[9]

Patients with MS have traditionally been divided into three categories based on the clinical disease course[10] (relapsing-remitting, secondary-progressive, and primary-progressive) in routine care, as well as in clinical development and during drug approval. High-efficacy therapies have been shown to effectively control relapse biology, whereas treatments targeting chronic MS progression have generally yielded limited results,[11] with a few exceptions.[12] Ongoing advances in understanding MS pathophysiology have opened the way for treatments targeting chronic disability progression. MS drug development timelines remain long, with no effective predictors or surrogate markers for clinical outcomes.[13] MS progression represents both a key therapeutic and measurement need, as the development of treatments requires quantifiable, objective, and reliable outcome measures reflecting disability progression.[14] These outcome measures should accurately capture disability progression in an appropriate MS cohort within a reasonable clinical trial timeframe.

Progressive ambulation impairment is considered a major consequence of MS. It significantly impacts the quality of life of MS patients,[15] and as such, is captured by traditional gold-standard clinical assessments such as the Expanded Disability Status Scale (EDSS) or Timed 25-Foot Walk (T25FW).[14] Most studies reporting the use of wDHT to assess ambulation in MS focused on physical activity or step count[16,17] measured by commercially available mainstream devices or smartphone sensors.[18] Some digital outcomes have shown encouraging metric properties,[16,17] reaching known-group and convergent validity but very few demonstrated sensitivity to detect change.[7] Despite increasing regulatory interest in walking speed as an outcome in MS[19] and strong evidence in the literature supporting its clinical relevance,[20] none of the currently available measures have yet been sufficiently validated to achieve regulatory approval. Beyond regulatory approval, robust evidence is needed to demonstrate the added value of digital outcome measures. To date, wDHTs have not demonstrated sufficient predictive value or undergone rigorous validation to meaningfully support drug development decisions or even drug approval.

By analyzing gait at the level of individual strides, our objective was to develop and validate digital mobility outcomes that are accurate, acceptable, reliable, and sensitive to the progressive impact of MS on ambulation. We specifically aimed to assess the analytical and clinical validity of stride-level mobility outcomes derived from daily-living ambulation.

Methods

Study design and populations

The design and objectives of both studies are presented in Fig. 1. Both study protocols are available as Supplementary Materials (Supplementary Files 1 and 2). This manuscript was prepared in accordance with the STROBE reporting guidelines.

ActiMS controlled environment study

We conducted a cross-sectional, single-center study, called "ActiMS Controlled Environment", from September 12, 2021, to September 18, 2021 to validate the performance of the wDHT in detecting strides and measuring the gait features in participants diagnosed with MS, at the Citadelle Hospital of Liège, Belgium. The inclusion criteria were a confirmed diagnosis according to the 2017 McDonald criteria, age between 18 and 65 years, an EDSS $\leq$5·5, no clinical and/or radiological relapse within the previous 3 months, and stable disease-modifying therapy for at least 2 months before inclusion. The exclusion criteria were significant cognitive disorders based on the investigator's clinical judgment, other past or current diseases that could impact the patient's motor function, and recent surgery or trauma.

ActiMS study in daily living conditions

We conducted a longitudinal, multicenter cohort study to develop a digital endpoint with robust clinical properties in people with MS at five sites in Belgium and France. Data were collected from March 2021 to January 2024. The long-term follow-up of this study is ongoing. Progression was defined according to the 2014 Lublin criteria.[10] Eligibility criteria were identical to those of the controlled environment study. Patients included in the controlled environment study were also eligible to participate in this study.

Standard protocol approvals, registrations, and patient consents

Approval of the controlled environment validation study (NCT04888689) was obtained from the Ethics Committees of the University Hospital of Liège (reference 2020/004) and the Regional Hospital of Liège (reference 1838). All participants were fully informed of the study protocol and provided written informed consent.

Approval of the protocol of the longitudinal cohort study (NCT04882891) was obtained from Ethics Committees of all participating sites. In Belgium, approvals were obtained from the central ethics committee of the Centre Hospitalier Universitaire de Liège (reference 2020/005) and from the local ethics committees of the Citadelle Hospital of Liège (reference 1838bis), Cliniques Universitaires Saint-Luc (reference 2021-485), Antwerp University Hospital (reference 2021-1720), and CHC Mont-Légia (reference 22/10/1140). In France, approval was obtained from the Committee for the Protection of Persons of Tours (reference 2·104·151·000·052). After receiving information about the study, patients provided written informed consent.

Both studies were conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki.

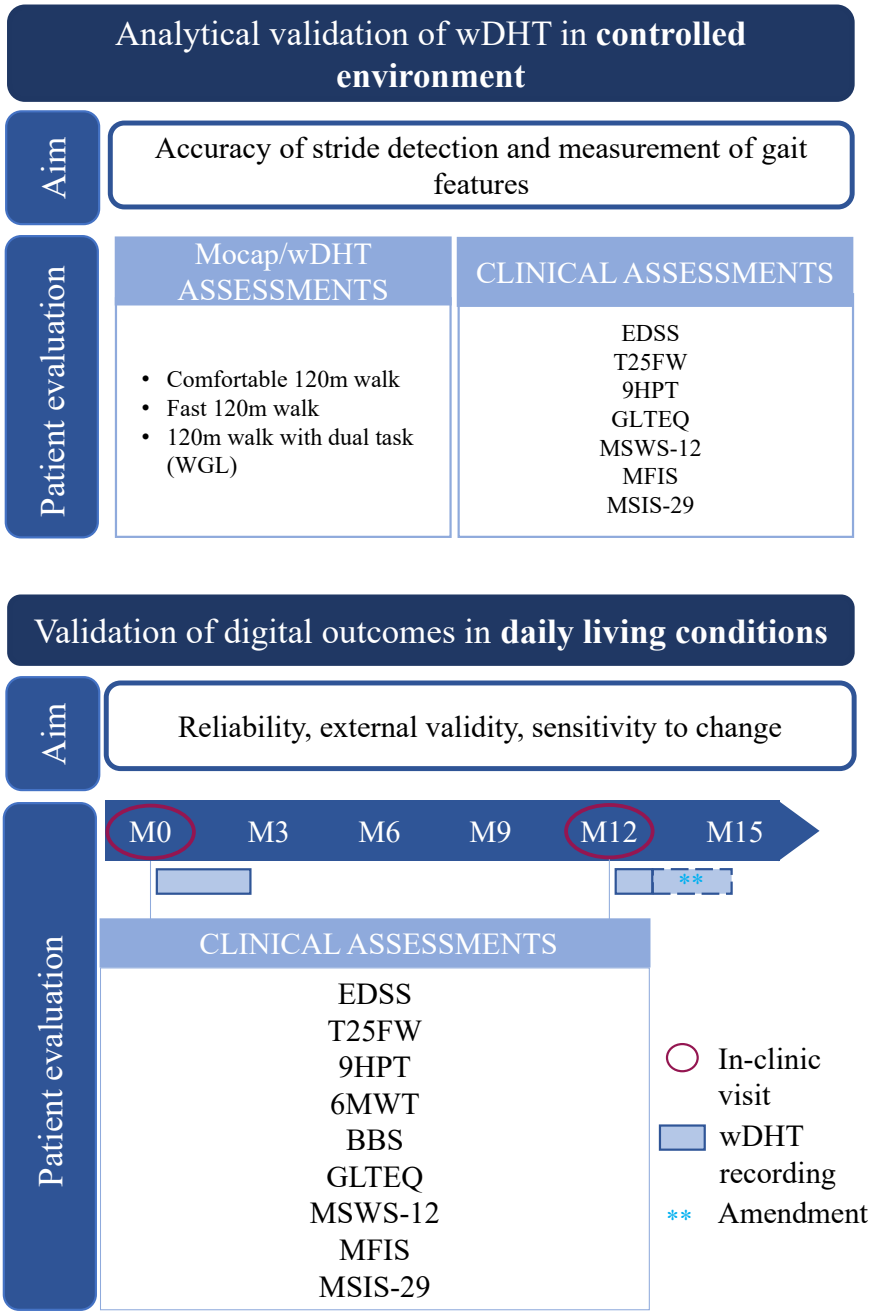

***Fig. 1:* Project overview.** Abbreviations: 6-Minute Walking Test (6MWT), 9-Hole Peg Test (9HPT), 12-item Multiple Sclerosis Walking Scale (MSWS-12), Berg Balance Scale (BBS), Expanded Disability Status Scale (EDSS), Godin Leisure Time Exercise Questionnaire (GLTEQ), Modified Fatigue Impact Scale (MFIS), Multiple Sclerosis Impact Scale (MSIS-29), Motion capture (Mocap), Timed 25-Foot Walk (T25FW), wearable digital health technology (wDHT), Word Generation List (WGL).

Data collection and devices

Clinical assessment

All clinical measures were performed at each visit in both studies except for the 6-Minute Walk Test (6MWT) and Berg Balance Scale (BBS), which were not performed in the Controlled Environment Study. The EDSS, T25FW and Nine-Hole Peg Test (9HPT)[21] were performed as gold-standard clinical assessments of patients with MS. We incorporated the 6MWT[22] and the BBS to gather data on walking endurance and ataxia,

respectively. The EDSS was performed by a physician (neurologist or neurology resident) trained in the protocol. Other clinical evaluations were conducted by physiotherapists or MS nurses trained in the protocol and its assessment procedures.

Patient-reported outcomes (PRO)

Patient-reported outcome measures, such as the Godin Leisure Time Exercise Questionnaire (GLTEQ), 12-item Multiple Sclerosis Walking Scale (MSWS-12), Modified Fatigue Impact Scale (MFIS), and Multiple Sclerosis Impact Scale (MSIS-29), were collected at each visit in both studies.

Data collection

Clinical data were collected on both paper and electronic case report forms with the latter hosted on Research Electronic Data Capture, a secure, web-based application designed for data collection.

Gait analysis in controlled environment

Patients were equipped with a wearable digital health technology (ActiMyo® device, Sysnav, Vernon, France) as well as motion-capture markers (OptiTrack®, NaturalPoint, Oregon, U.S.A.) on both ankles. Motion capture markers were fixed to the wDHT. Participants completed three walking exercises: fast walking speed, comfortable walking speed, and dual-task walking along a rectangle-shaped path to provide a wide range of strides at different gait speeds and cognitive loads. They were instructed to "walk at a comfortable pace," "walk as fast as possible", and "walk while performing the Word Generation List (WGL) test", respectively.

Real-world gait analysis and selected digital metrics definitions

The wearable sensor data were collected using the ActiMyo® device (Sysnav, Vernon, France), as depicted in Supplementary File 3. Patients were asked to wear two sensors – one on each ankle – each containing high-precision triaxial accelerometers, triaxial gyrometers, a temperature sensor, and a barometer (recording data at 100 Hz), that captured linear acceleration, angular velocity (in all directions), and barometric altitude.

The sensors were individually factory-calibrated to ensure the validity of each sensor's measurements across a range of temperature, pressure, and dynamic environments.

Participants were instructed to wear the sensors during the daytime for 3 months from the baseline visit and 1–3 months after the 1-year visit (according to protocol amendment version 5·0). They were advised to remove the sensors for water-related activities such as showering or swimming.

Raw data were continuously collected throughout the day, stored in an internal memory and transferred to an internal USB drive in the docking station overnight when the patient removed and charged the device. From the docking station, encrypted and anonymized data were transferred securely to a dedicated secure cloud-based server or retained on the USB drive (capacity to store up to 3 months of data). At the end of each recording period, raw data were analyzed using Sysnav's proprietary algorithms, which were specifically developed to analyze ambulation from the daily-living data collected by the wDHT device, ActiMyo®.

The algorithms detected individual strides and reconstructed the trajectory of the ankle from raw data[6,23] from which multiple metrics are derived including stride length and stride velocity. Individual strides were grouped into walking episodes as periods of continuous walking without any time gap between two successive strides exceeding 15 s. At the walking-episode level, additional variables were computed, such as walking episode length, defined as the sum of the lengths of the individual strides composing the episode.

To establish a surrogate for maximal ambulation speed, as assessed in the T25FW, and maximal walking distance, as assessed in the EDSS, we focused on stride velocity and walking episode length, as these variables are clinically relevant in MS. Stride velocity and walking episode length were computed for each stride and each walking episode across the entire recording period (i.e., one month in this study) and then aggregated to generate distributions for each variable over the full recording period. From the distribution of these variables, we computed various percentiles (50th, 80th, 85th, 90th, 95th, 99th). We selected high percentiles to reflect maximal daily-living performance, while avoiding very high percentiles, as these are computed from a small number of strides or walking bouts, which may have introduced variability in derived metrics. Our choice of centile was guided by the combined evaluation of reliability (intra-class correlation coefficient, ICC) and responsiveness (standardized response mean, SRM) across percentiles. For maximal ambulation speed, we verified that the 95th percentile, the centile of stride velocity selected for qualification as a primary endpoint in DMD, was also a suitable choice for MS. For maximal walking distance, we selected a centile by balancing the goal of representing the top performance while limiting variability. The detailed results supporting the selection of the stride-velocity and walking-distance percentiles are provided in Supplementary Files 4.

As a result, we calculated the stride velocity 95th percentile (SV95C) as an indicator of peak ambulation ability, defined as the 95th percentile of all stride velocities recorded during one month of daily living. Similarly, we calculated the walking distance 90th percentile (WD90C), as the 90th percentile of walking episode lengths recorded over the same period. In

addition, overall physical activity was quantified as the number of strides per hour, calculated as the ratio between the total number of strides performed by the patient and the total recording duration with the wDHT over one month.

Statistical analysis

Statistical analyses were performed using Python (version 3·11). Population characteristics were analyzed using descriptive statistics. Continuous and categorical clinical variables were compared between the populations of both studies, using the Wilcoxon rank-sum test and the chi-square test of independence, respectively.

Sample size calculation

As this work represents the first application of these stride-level algorithms and digital mobility metrics in a multiple sclerosis population, no prior data were available to support a formal sample size calculation tailored to this condition. The sample sizes for both the controlled environment and longitudinal studies were therefore determined based on feasibility considerations and our prior experience conducting similar validation studies in other neurological conditions.[3]

ActiMS controlled environment study: analytical validation

Analytical validation evaluates the performance of sensor technology and the data it produces at the sample level, based on a predefined set of criteria.[24] SV95C relies on the ability of the wDHT to correctly detect strides and accurately estimate stride velocity in this population and in conditions representative of daily living. WD90C relies on the ability of the wDHT to detect strides correctly and accurately measure stride length.

The accuracy of the stride detection system was validated using manual annotations of strides. Manual labeling was performed by two skilled and trained engineers using LabelStudio, an interface displaying the wDHT sensor raw signals along axes relevant to gait detection. Joint review of the manual stride annotations was done with discussion to reach a consensus if necessary. Comparison between the manual annotations and outputs of the wDHT algorithm allowed to categorize each detected stride as a True Positive (TP; correctly detected stride), False Positive (FP; stride detected by the algorithm when none occurred), or False Negative (FN; missed stride that was present in the manual annotation). After pooling data from both the left and right sides, per exercise and across all exercises, the analytical performance of the stride detection algorithm on MS patients was evaluated using the total number of True Positive (TP), False Positive (FP), False Negative (FN), the precision, and the recall.

The precision of stride detection corresponds to the ratio of well-detected strides (total number of True Positive strides) among all the detected strides (total number of True Positive and False Positive strides). It was evaluated using the following formula:

$$P = \frac{TP}{TP+FP}$$

Recall of stride detection corresponds to the ratio of well-detected strides (total number of True Positive strides) among all the existing strides (total number of True Positive and False Negative strides). It was evaluated using the formula:

$$R = \frac{TP}{TP+FN}$$

Validation of the accuracy of the system to measure stride velocity and length was performed against a gold-standard reference (optical motion capture system, Mocap), after synchronization of the Mocap and the wDHT data. For each individual stride, the error in the wDHT-estimated stride length or velocity was computed as the difference between the reference estimate and the wDHT estimate. The error was evaluated only for strides whose trajectories were captured by the motion capture system. Errors in wDHT-estimated stride length and stride velocity were evaluated using Bland–Altman analyses adapted for repeated measures to account for the non-independence of multiple strides originating from the same participant.[25] In addition, quantitative metrics were computed, including the mean absolute error and the standard deviation of the absolute error, defined respectively as the mean and standard deviation of the absolute differences between wDHT and reference estimates. The mean relative absolute error was calculated as the mean absolute error divided by the corresponding reference estimate. Agreement between the Mocap and wDHT measurements was assessed using the intraclass correlation coefficient (ICC(1,1)). The association between clinical variables (age, EDSS, height or BMI) and the stride length or stride velocity error was graphically assessed. All analyses were performed by pooling strides from both the right and left sides.

Analyses were not stratified by disease phenotype (relapsing-remitting (RRMS) vs progressive MS (PMS)), as the study was not sufficiently powered to support reliable subgroup comparisons.

Longitudinal ActiMS study in daily living conditions

Each analysis was conducted using established statistical methods, allowing direct comparison with traditional clinical endpoints while accounting for the unique characteristics of digital outcomes.

Descriptive statistics were used to summarize population characteristics. Continuous clinical variables were compared between the baseline cohort and participants who withdrew from the study using the Wilcoxon rank-sum test. Categorical variables were

compared between groups using the chi-square test of independence.

The minimal recording duration (MRD) was defined as the minimum amount of recording time required within a recording period to reliably compute digital metrics from the raw data. The optimal recording duration (ORD) was defined to inform power calculations and support the design of future clinical trials, including decisions on recording-period length and participant compliance targets. To determine these two thresholds, we assessed the stability and reliability of the digital variables by calculating intraclass correlation coefficients (ICC [1,1]) between pairs of consecutive recording periods with varying durations. MRD and ORD were operationally defined as the recording durations required to achieve ICC (1,1) values of 0·8 and 0·9, respectively. These thresholds therefore represent the minimum amount of raw data required to obtain stable and reproducible digital variables.

After evaluating the ORD and MRD, compliance was analyzed for baseline and longitudinal data based on these definitions.

For the data analysis, we decided to evaluate the digital variables in periods of exactly one month from the start of the baseline period and the 1-year period. Moreover, to have one digital variable estimation per patient (as patients wore two sensors leading to two estimations), we chose the digital variable estimation from the ankle with the most strides during the baseline period.

Reliability was assessed at baseline using the intraclass correlation coefficient (ICC [2,1]), which quantifies the consistency of measurements across different time periods. ICC (2,1) was calculated based on two consecutive 1-month measurement periods of the baseline recording period. This approach allowed us to determine the reliability of a single measurement over a month-long period. ICC values were interpreted as follows:

<0·50: poor reliability
0·50–0·79: moderate reliability
0·80–0·89: good reliability
>0·90: excellent reliability

The Standard Error of Measurement (SEM), which reflects intra-patient variability, was calculated.

The distribution of all outcome variables was assessed using the Kolmogorov–Smirnov test for normality. This evaluation revealed heterogeneous distributional properties across outcomes. Specifically, the number of strides per hour, SV95C, EDSS, MSWS-12, MFIS, and MSIS-29 did not significantly deviate from normality, whereas WD90C, 9HPT, 25FW, the Berg Balance Scale, FAB, MMSE, and GLTEQ showed significant deviation from normality. Given this heterogeneity and the limited sample size for several outcomes, apparent normality was interpreted with caution. To ensure a conservative, consistent, and robust analytical strategy across all outcomes, non-parametric methods were therefore retained for all analyses.

Convergent and divergent validity of the digital metrics was evaluated by examining their correlation with traditional clinical endpoints assessing related and unrelated constructs at baseline using Spearman's rank correlation coefficient. Correlation coefficients were interpreted as follows:

<0·20: very weak correlation
0·20–0·40: weak correlation
0·40–0·60: moderate correlation
0·60–0·80: strong correlation
>0·80: very strong correlation

To account for multiple comparisons and reduce the risk of Type I errors, we applied a Bonferroni correction ($\alpha' = \alpha/m$). Given that multiple statistical tests were conducted, the standard significance level ($\alpha = 0{\cdot}05$) was adjusted by dividing it by the number of tests performed ($m = 30$). This narrower threshold ($\alpha' = 0{\cdot}002$) reduces the likelihood that significant correlations represent random findings.

For known-group validity, we assessed whether the digital metrics could differentiate between groups with varying levels of disability at baseline. The study population was divided into three subgroups based on the EDSS:

EDSS $\leq$2: patients with no deficits
2<EDSS<4: patients with deficits but no gait disorder
EDSS $\geq$4: patients able to walk with limitations

Differences between groups were analyzed using the Wilcoxon rank-sum test, a non-parametric alternative to the t-test that does not assume normality.

Longitudinal changes in digital and clinical metrics were assessed over a 1-year period. To account for potential non-normality and small sample sizes, we used the Wilcoxon signed-rank test, which is a non-parametric test for paired data. This analysis was conducted across the entire population and within the PMS and RRMS subgroups. As a post hoc analysis, we studied the correlation between SV95C and the MSWS-12, a patient-reported outcome that evaluates walking ability. Additionally, we investigated whether SV95C could distinguish between patients who experienced significant worsening, defined as a $\geq$ 4-point decline in MSWS-12, and those who did not.

Role of the funding source

F. Hoffmann-La Roche Ltd. contributed to the refinement of the study protocol, the interpretation of the data, and the revision of the manuscript.

Results

Population

Participant flow through the study is summarized in Fig. 2. The baseline demographic and disease-related information of the study population are shown in Table 1. A total of 21 and 78 participants were recruited into the controlled environment and longitudinal studies, respectively. Nineteen patients who participated in the controlled environment study were also included in the daily living study. Seventeen patients with RRMS (aged 23–64 years), including 12 women, did not attend a follow-up visit. Reasons for withdrawal are detailed in Fig. 2. As shown in Table 1, there were no significant differences in baseline characteristics between the populations of the two studies. In the daily living study, there were no significant differences between participants who withdrew and the overall study population with respect to age, gender, height, body mass index (BMI), age at first clinical manifestation, disease course, EDSS scores, T25FW, 9HPT (right and left), BBS, GLTEQ, MSWS-12, MFIS, or MSIS-29 ($p = 0{\cdot}52$, $0{\cdot}37$, $0{\cdot}87$, $0{\cdot}36$, $0{\cdot}3$, $0{\cdot}07$, $0{\cdot}73$, $0{\cdot}93$, $0{\cdot}28$, $0{\cdot}7$, $0{\cdot}83$, $0{\cdot}21$, $0{\cdot}58$, $0{\cdot}64$, and $0{\cdot}99$ respectively).

ActiMS controlled environment study

Analytical validation

The results of the analytical validation of stride detection and stride velocity and length estimation are listed in Table 2. The performance of stride detection was excellent in patients with MS, as indicated by precision (near 1 for all exercises) and recall (0·96 for all exercises).

Due to the setup of the controlled environment (limited coverage area of the camera), the availability of strides was lower for the Mocap than for the wDHT. Results of the analytical validation of stride velocity and length are presented in Figs. 3 and 4. The wDHT-derived stride velocity and length were accurate (mean absolute error 0·019 m/s and 0·022 m, respectively) and precise (standard deviation of the absolute error of 0·019 m/s and 0·023 m respectively) on both ankles and across all exercises (Table 2).

The mean relative absolute error was low for both stride velocity and stride length (2·1–2·7%), indicating a small proportional error over the full range of measured values. Agreement between wDHT and Mocap was excellent across all exercises. ICC values ranged from 0·996 to 0·998 for stride velocity estimation and from 0·992 to 0·996 for stride length estimation, demonstrating near-perfect agreement (Table 2).

As demonstrated in Fig. 3 by the lines at ±1·96 SD, the length and velocity of 95% of the strides are quantified with errors below 7 cm and 6 cm/s, respectively. Based on visual inspection, no systematic relationship was observed between stride velocity or stride length errors and their corresponding measured values.

As illustrated in Fig. 4, the mean errors in stride velocity and length were similarly distributed across all exercises. No clear trends or dependencies were observed between clinical variables (age, EDSS, height, and BMI) and the mean errors in stride velocity and stride length per patient.

Longitudinal study in daily living conditions

Clinical data availability

At baseline, MSIS-29 assessments were missing because this measure was added later through a protocol amendment (n = 16), with an additional four cases resulting from site oversight. The 6MWT was also missing for 16 participants for the same amendment-related reason. At the 12-month visit, missing data were limited to four participants who did not complete

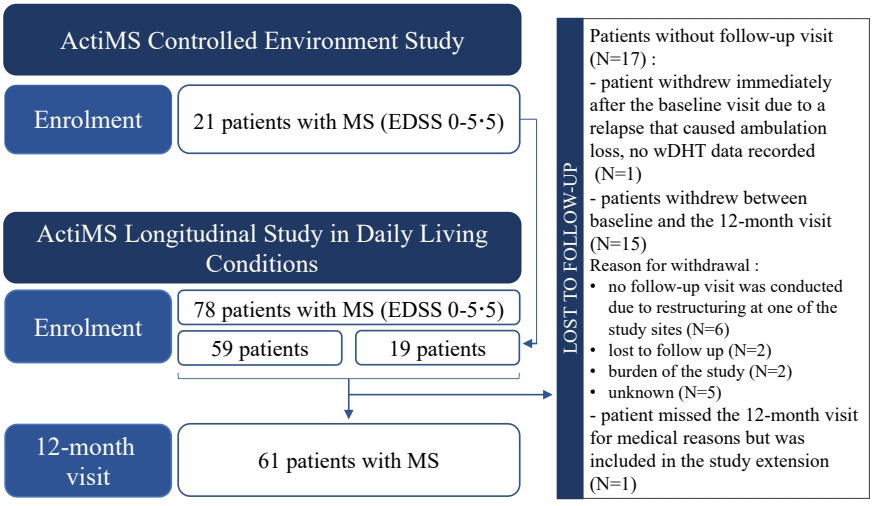

Fig. 2: **Diagram illustrating the flow of participants in both studies.**

	Controlled environment study			Longitudinal daily living study			Overall population comparison (p-value)
	Total	RRMS	PMS	Total	RRMS	PMS (primary/ secondary)	
Number of patients	21	18	3	78	61	17 (9/8)	NA
Age (years): median (IQR) [range]	39 (34–54) [23–62]	38·5 (34·5–52·7) [23–62]	51 (42–56) [33–61]	48·5 (38·2–55) [23–64]	43 (37–54) [23–64]	55 (51–58) [33–62]	0·16 ns
Height (centimeters): mean ± SD [range]	172·4 ± 11·7 [152·5–189·8]	172·9 ± 9·7 [152·5–189·8]	169·5 ± 16 [153·2–185·2]	170·9 ± 9·8 [151–189·8]	170·8 ± 9·7 [151–189·8]	171·6 ± 10·7 [151–185·2]	0·50 ns
BMI (kg/m^2): mean ± SD [range]	27·3 ± 4·2 [21·4–36·2]	27·1 ± 3·9 [21·4–36·2]	28·3 ± 6·3 [21·8–34·5]	26·6 ± 4·8 [18·6–45]	26·7 ± 4·9 [18·7–45]	26·3 ± 4·9 [18·6–35·2]	0·43 ns
Sex: female (%)	12 (54·5)	11 (61·1)	1 (33·3)	43 (55·1)	37 (60·7)	6 (35·3)	1 ns
EDSS: mean ± SD [range]	2·6 ± 1·3 [1·5-5·5]	2·2 ± 1 [1·5-5·5]	4·7 ± 1 [3·5-5·5]	3 ± 1·4 [0–5·5]	2·7 ± 1·3 [0–5·5]	4·1 ± 1·1 [1·5-5·5]	0·25 ns
T25FW (s): mean ± SD [range]	5·3 ± 2·3 [3·1–13·7]	4·7 ± 1·4 [3·1–9·6]	5·4 ± 4·8 [5·4–13·7]	6·5 ± 6·7 [2·8–60]	5·2 ± 2·2 [2·8–17·5]	11·2 ± 13 [4·3–60]	0·28 ns
9HPT right side (s): mean ± SD [range]	21·8 ± 6·0 [13·8–37·3]	20·4 ± 3·7 [13·8–27·3]	30·4 ± 10·8 [18–37·3]	24·8 ± 14·9 [15·5–135·2]	23·4 ± 15·5 [15·5–135·2]	29·6 ± 11·4 [18·3–59·9]	0·41 ns
9HPT left side (s): mean ± SD [range]	22·8 ± 5·8 [15·4–35·0]	21·5 ± 4·5 [15·4–32·5]	30·8 ± 7·3 [22·4–35·0]	24·1 ± 7·4 [16·0–60·6]	23·0 ± 6·1 [16·0–46·5]	27·8 ± 10·2 [19·8–60·6]	0·50 ns
6MWT (meters): mean ± SD [range]	NA	NA	NA	448·8 ± 139·5 [147·5–774]	478·2 ± 130·1 [200–774]	329 ± 112·9 [147·5–479]	NA
BBS: mean ± SD [range]	NA	NA	NA	51·9 ± 6·3 [18–56]	53·4 ± 4·0 [36–56]	46·2 ± 9·5 [18–56]	NA
GLTEQ: mean ± SD [range]	20·5 ± 18·2 [0–63]	21·7 ± 17·7 [0–63]	13·3 ± 23·1 [0–40]	25·0 ± 26·3 [0–140·0]	27·8 ± 28·4 [0–140·0]	14·8 ± 12·6 [0–40]	0·71 ns
MSWS-12: mean ± SD [range]	24·4 ± 14·1 [12–59]	20·8 ± 11·1 [12–44]	46·0 ± 11·3 [39–59]	27·8 ± 13·2 [12–60]	24·9 ± 12·4 [12–60]	37·9 ± 11·2 [17–59]	0·20 ns
MFIS: mean ± SD [range]	32·9 ± 21·8 [0–65]	30·7 ± 21·1 [0–65]	46·3 ± 25·7 [17–65]	37 ± 21·9 [0–80]	33·8 ± 21·9 [0–80]	48·5 ± 19·0 [10–74]	0·47 ns
MSIS-29: mean ± SD [range]	59·8 ± 20·4 [30–94]	57·3 ± 20·1 [30–90]	74·7 ± 19·0 [56–94]	62·0 ± 22·0 [30–129]	61·3 ± 21·8 [30–129]	64·8 ± 23·5 [41–111]	0·77 ns

Abbreviations: Relapsing-Remitting Multiple Sclerosis (RRMS), Progressive Multiple Sclerosis (PMS), Body Mass Index (BMI), 6-Minute Walking Test (6MWT), 9-Hole Peg Test (9HPT), 12-item Multiple Sclerosis Walking Scale (MSWS12), Berg Balance Scale (BBS), Expanded Disability Status Scale (EDSS), Godin Leisure Time Exercise Questionnaire (GLTEQ), Interquartile Range (IQR), Modified Fatigue Impact Scale (MFIS), Multiple Sclerosis Impact Scale (MSIS-29), Standard Deviation (SD), Timed 25 Foot Walk (T25FW), Not Applicable (NA).

Table 1: **Population characteristics.**

	All exercises	Fast walk	Comfortable walk	Dual-task walk
Stride detection				
TP (number of strides)	10448	3160	3629	3659
FP (number of strides)	6	2	2	2
FN (number of strides)	434	120	162	152
Precision	0·999	0·999	0·999	0·999
Recall	0·960	0·963	0·957	0·960
Stride velocity estimation				
Stride count (common to Mocap and wDHT)	3808	1065	1329	1414
Mean absolute (abs) error (m/s)	0·019	0·020	0·022	0·016
Mean relative error (%)	2·305	2·103	2·660	2·123
Std· abs· error (m/s)	0·019	0·020	0·022	0·014
ICC (1,1) agreement	0·998	0·998	0·996	0·998
Stride length estimation				
Stride count (common to Mocap and wDHT)	3808	1065	1329	1414
Mean absolute (abs) error (m)	0·022	0·022	0·026	0·020
Mean relative error (%)	2·305	2·103	2·660	2·123
Std· abs· error (m)	0·023	0·021	0·028	0·019
ICC (1,1) agreement	0·995	0·996	0·992	0·996

Abbreviations: True Positive (TP), False Positive (FP), False Negative (FN), Motion capture (Mocap), wearable Digital Health Technology (wDHT), absolute (abs), Standard deviation (std), intraclass correlation coefficient (ICC).

Table 2: **Metrics for stride detection, stride velocity and length estimation per exercise.**

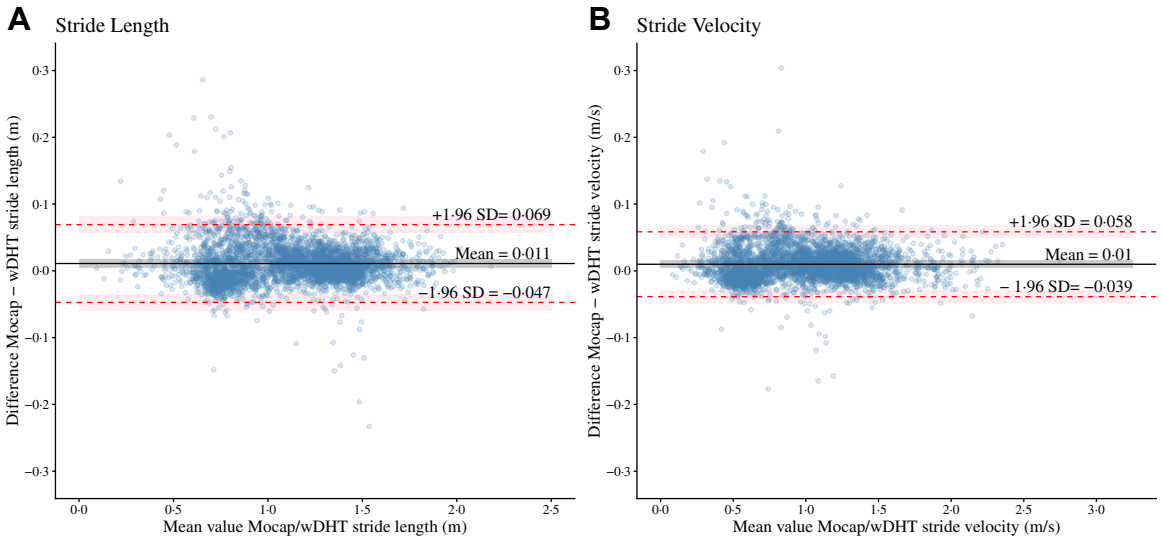

Fig. 3: **Bland-Altman plot showing the error in stride length (A) and velocity (B) for individual strides, as a function of the mean value of each variable.** Black line is the mean error for all strides, and red dashed lines delineate the 95% confidence interval. Grey shaded area indicates the 95% confidence interval around the mean error and red shaded areas represent the 95% confidence intervals of the limits of agreement.

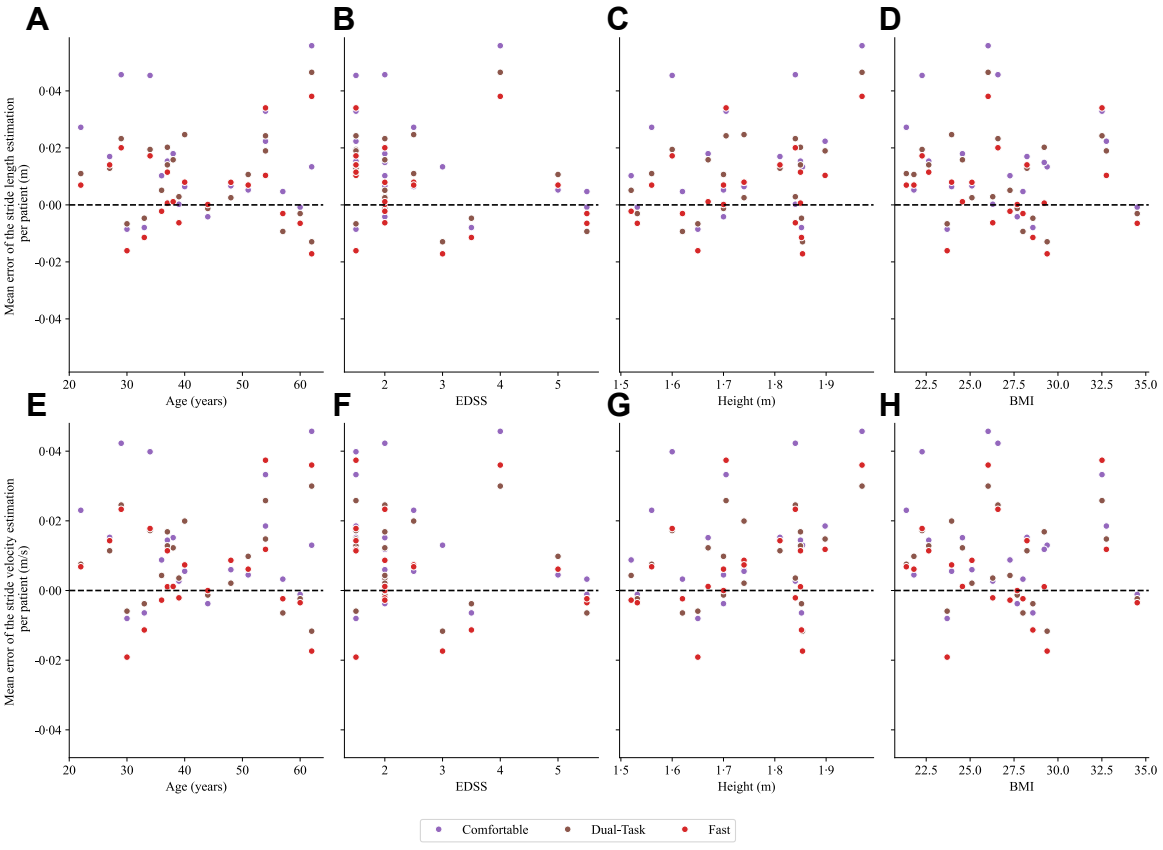

Fig. 4: **Impact of age, EDSS, height, and Body Mass Index (BMI) on the mean error of stride length (A, B, C, and D) and velocity (E, F, G, and H).**

any follow-up clinical assessments due to site restructuring, along with a small number of isolated missing values for the MSWS-12 (n = 3) and MFIS total score (n = 1) related to site oversight. Overall, only a small proportion of clinical outcome measures were missing across both visits.

wDHT wear time

The comparison of multiple ICCs, computed using two consecutive recording periods of variable lengths, enabled the validation of the MRD and ORD in the MS population and is detailed in Supplementary File 5. This demonstrates that MRD and ORD can be defined as 50 h and 180 h, respectively.

The baseline and 1-year wear-time data (<MRD, MDR-ORD, and >ORD) are presented in Table 3. In total, 98·7% and 91·5% of patients provided sufficient data for analysis (>50h) at baseline and 1 year, respectively.

Reliability

Cross-sectional analysis demonstrated the high reliability of SV95C (ICC (2,1) = 0·97, SEM = 0·06, n = 71), number of strides per hour (ICC (2,1) = 0·91, SEM = 25·92, n = 75) and WD90C (ICC (2,1) = 0·90, SEM = 8·02, n = 73). The reliability of the different digital metrics was not (for SV95C) or moderately (for WD90C) dependent on their magnitude, as shown in Fig. 5. Reliability remained good to excellent across all digital metrics and when RRMS and PMS groups were analyzed separately, with ICC (2,1) values ranging from 0·86 to 0·98. The highest reliability was observed for SV95C, with ICC (2,1) values of 0·96 and 0·98 in the RRMS and PMS groups, respectively.

Known group validity

SV95C and WD90C were able to differentiate between all three EDSS subgroups. Fig. 6 presents box plots for each digital metric with three EDSS subgroups.

Convergent and divergent validity

All digital metrics demonstrated moderate to strong correlations with EDSS, T25FW, BBS, MSWS-12, and GLTEQ (Fig. 7). These relationships remained significant after Bonferroni correction ($p < 0·003$). Correlations with outcomes related to motor function but not directly reflecting lower-limb function (MFIS and 9HPT) were generally weaker but remained statistically significant for WD95C ($\rho = 0·285$, $p < 0·05$ and $\rho = 0·327$, $p < 0·01$, respectively) and SV95C ($\rho = 0·326$, $p < 0·01$ and $\rho = 0·521$, $p < 0·01$, respectively). The digital mobility metrics showed no meaningful correlation with measures not primarily related to mobility (MMSE and FAB), supporting the divergent validity of the digital mobility metrics (Fig. 7). The results of the correlations between digital metrics and traditional clinical outcomes stratified by disease group (RRMS and PMS) are provided in Supplementary File 6.

Sensitivity to change

SV95C demonstrated a statistically significant decline over 12 months in both the PMS and RRMS populations ($p = 0·006$ and $p = 0·049$, respectively), whereas EDSS showed a significant change only in the PMS population ($p = 0·01$), as illustrated in Fig. 8. T25FW showed a statistically significant improvement in the RRMS group over 1 year (mean change: 0·63 s, $p = 0·03$); however, the magnitude of this improvement was minimal and may not be clinically meaningful. In the PMS group, no significant change was observed in T25FW. We did not observe any changes in pharmacological treatment or physical activity levels (as measured by the GLTEQ) that could account for the increases in T25FW performance in RRMS. For SV95C, the standardized response means (SRMs) for changes at 12 months were –0·936 for PMS and –0·316 for RRMS. For EDSS, the SRM for changes at 12 months in the PMS group was 0·716.

In a post-hoc analysis, the correlations between changes in MSWS-12 and SV95C did not reach statistical significance in the overall population ($R = –0·27$, $p = 0·07$, n = 47), nor in the PMS ($R = –0·29$, $p = 0·31$, n = 14) or RRMS ($R = –0·24$, $p = 0·18$, n = 33) subpopulations. SV95C did not differentiate between patients who exhibited a significant decline in MSWS-12 (defined as an increase of 4 points) and those who did not.

In the two patients who experienced a relapse during the 1-year follow-up, SV95C decreased by 7 cm/s and 8 cm/s. WD90C and strides per hour also declined (–1·9 and –10·7 m, –39 and –78 strides, respectively). In these patients, EDSS increased by 0·5 and 1·0 point, MSWS-12 scores worsened significantly (+6 and + 23 points, respectively), while T25FW showed no meaningful change.

Period	Baseline	1 year
<50 h: n (%)	1 (1·3)	5 (8·5)
50–180 h: n (%)	13 (16·9)	6 (10·2)
>180 h: n (%)	63 (81·8)	48 (81·4)

Table 3: **Sensor wear time (number and percentage of participants by recording duration).**

Discussion

We demonstrate that a digital outcome based on individual stride-level data shows strong potential as a sensitive and reliable early measure of ambulation decline experienced by people with MS. In contrast to traditional approaches that rely on aggregated measures, such as mean step counts, average daily walking

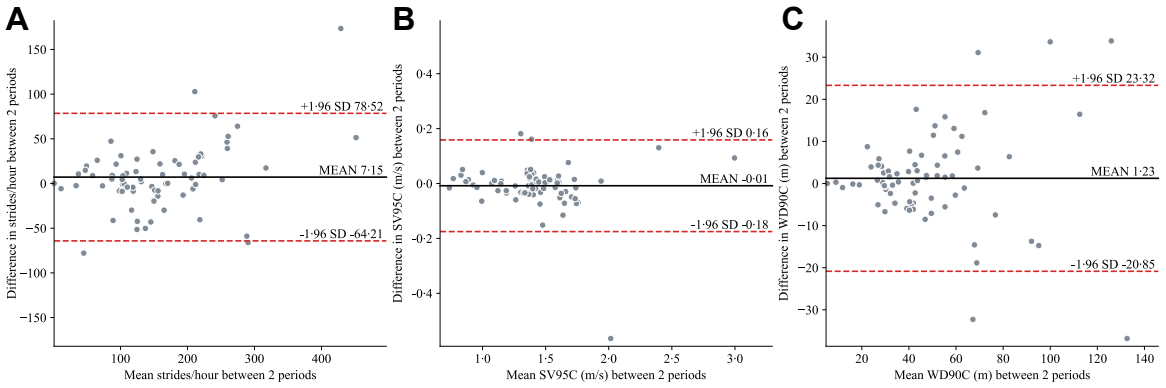

Fig. 5: **Bland-Altman plot showing the variation between 2 consecutive 1-month periods for each digital variable as function of their mean values: strides/hour (A), SV95C (B), and WD90C (C).**

speed, or walking-bout-level metrics, our method identifies and segments every stride to compute stride-specific velocity and stride length. This enables a percentile-based analysis of stride distributions. Some of the proposed digital metrics, such as SV95C, are based on stride-level detection, measurement, and on analysis of stride-level distributions. By contrast, metrics based on stride counts or walking-bout characteristics rely on aggregation approaches similar to those used in previous studies. By leveraging the full distribution of stride characteristics rather than average-based measures typically obtained from lower-back-mounted wearable sensors,[26] our approach offers a novel opportunity to detect subtle mobility impairments at an earlier stage.

The stride detection, as well as stride velocity and length measurements, of the wDHT demonstrated high accuracy, with significantly reduced errors compared with levels typically reported in publicly available data from other research-grade sensors.[26,27] The excellent reliability of SV95C (ICC of 0·97) observed in our study surpasses the reliability of outcomes that simply reflect overall patient activity, such as the number of strides per hour and WD90C. Additionally, this level of reliability exceeded that of other gait-related outcome measures reported in the literature,[28] and is comparable to the reliability demonstrated during the EMA qualification of the same outcome in DMD[3] or in other neuromuscular diseases.[29] We were able to quantify the annual decline in both RRMS and PMS, whereas EDSS showed a statistically significant decline only in the PMS subgroup. Although these results require confirmation in larger, independent cohorts, they suggest that SV95C could be used to detect disability progression in PMS and RRMS populations. This potential is particularly important, as sensitivity is a key determinant of statistical power in efficacy trials.

We computed three digital outcomes representing distinct aspects of gait impairment, all of which showed good metric properties in MS. The number of strides per hour and maximal walking distance (WD90C) provide a continuous and granular measure of mobility, as suggested by the wide variability observed within the same EDSS level. In MS,[18] few studies have

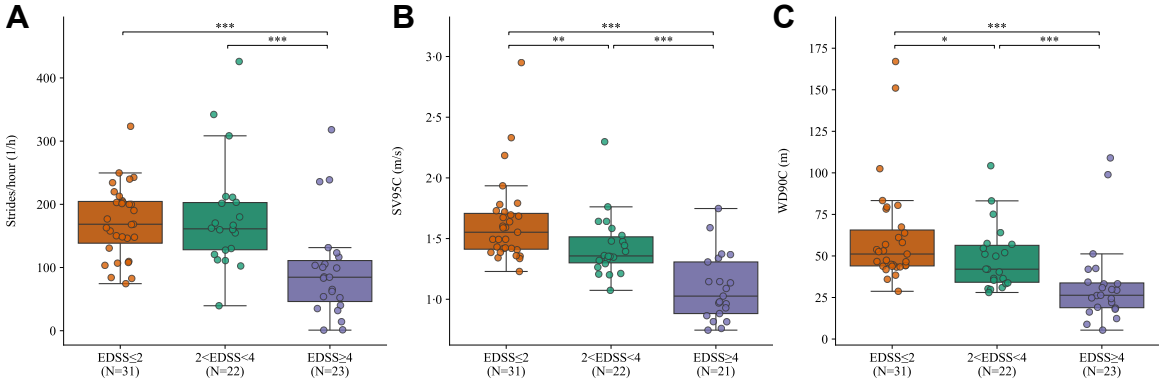

Fig. 6: **Digital outcome measures by level of disability: strides/hour (A), SV95C (B), WD90C (C).** (p < 0·05*, p < 0·01**, p < 0·001***).

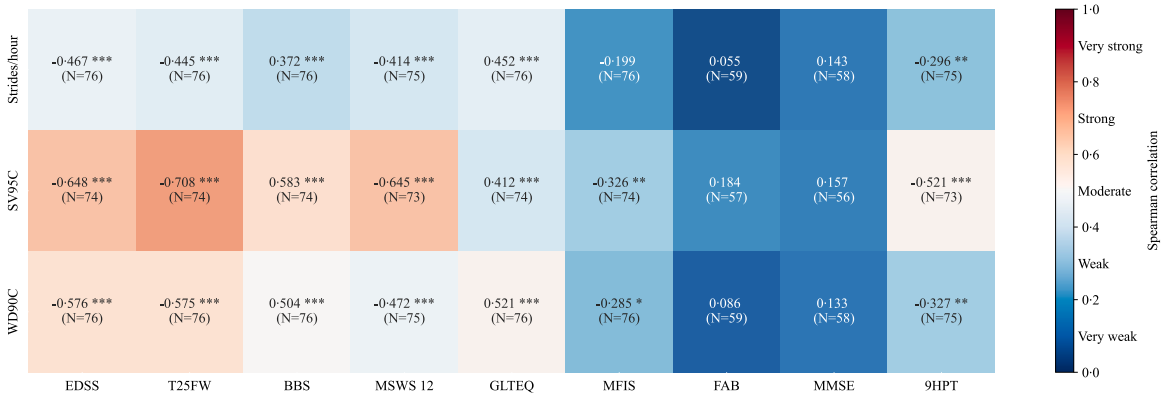

Fig. 7: **Correlations between digital and clinical variables** (p < 0·05*, p < 0·01**, p < 0·001***).

demonstrated that step count may be partially sensitive to change. However, interpreting changes in these metrics remains challenging, as they can be significantly influenced by factors beyond disease progression, such as lifestyle, motivation, comorbidities, fatigue, and environmental conditions. Gait parameters, such as SV95C, may show lower variability driven by external factors and greater variability attributable to disease progression. This is reflected in its excellent reliability and strong sensitivity to change. This is consistent with observations in other diseases, where step count and cumulative activity metrics are less reliable and less sensitive to changes[3,29] than outcomes based on individual stride metrics such as SV95C.

Overall, these three digital outcomes (strides per hour, WD90C and SV95C) accurately reflect daily walking ability, correlate with gold-standard measures, and distinguish between patients with different levels

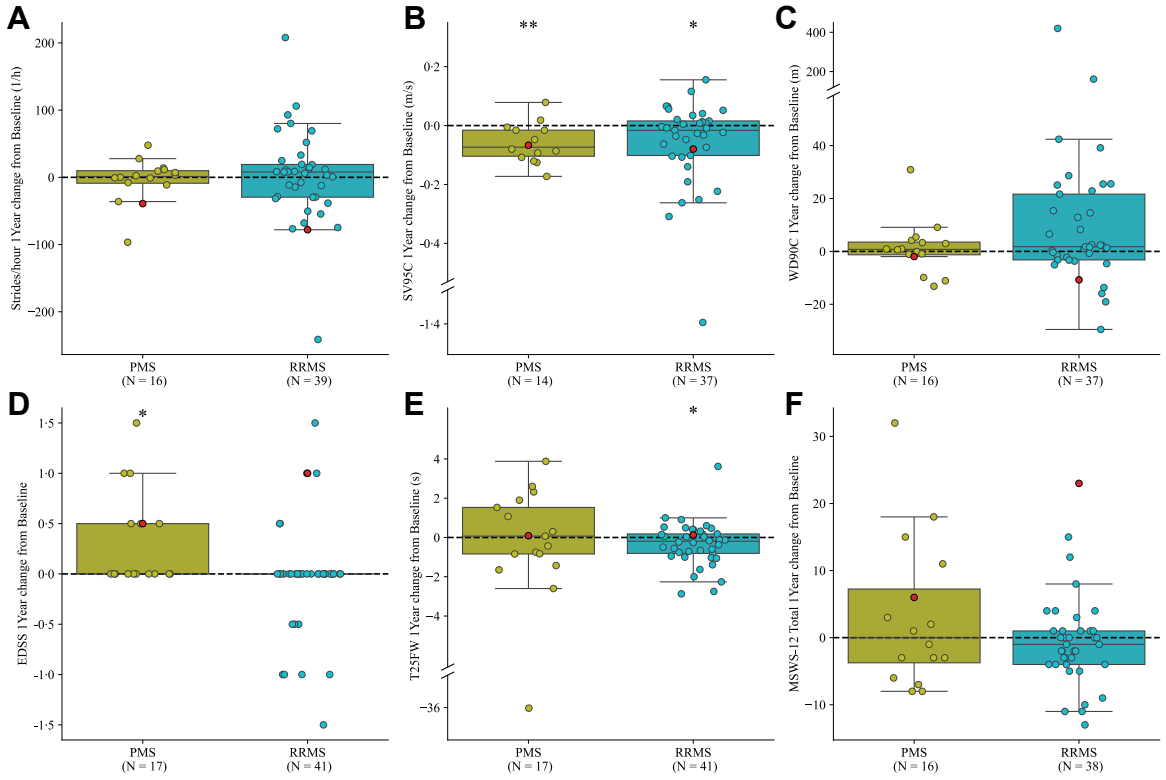

Fig. 8: **Box plots with change from baseline by subgroup (RRMS/PMS) for each digital and clinical outcome: strides/hour (A), SV95C (B), WD90C (C), EDSS (D), T25FW (E), and MSWS-12 (F).** (p < 0·05*, p < 0·01**). The red dots represent patients who experienced a relapse during the 1-year follow-up.

of disability. The most promising outcome (SV95C) appears to be more reliable, more discriminant, and more sensitive to decline than other digital outcomes.

Although promising, wDHT-based gait assessments in daily living are not yet the gold standard nor widely adopted. Some clinical trials in MS have already begun to use wDHT to provide outcome measures,[30] but the need for further validation of these outcomes and wDHT is urgent. Emerging guidance for evaluation of digital assessments[24] is essential to address the current lack of standardization and result comparability, which currently prevents the integration of these innovative solutions into clinical trials. The implementation of such technology – beyond consumer-grade devices – in clinical practice is even more distant. Key challenges include addressing the burden on patients, as well as the accessibility and interpretability of the data for both clinicians and patients.

This study has several limitations that should be acknowledged. Despite the large volume of recorded data, the sample size was relatively limited, particularly within subgroups (RRMS/PMS), and particularly in the analytical validation cohort. As a result, the study was not powered to support formal subgroup analyses by sex or age, or site-stratified analyses. However, the data generated here provide a basis for future sample size calculations in studies designed to confirm these findings and assess their generalizability. Future studies should include a larger representation of patients with PMS. In addition, recruitment was restricted to European centers and ambulant individuals, with limited ethnic diversity and representation of patients with high disability levels, which may further limit the generalizability of the findings. Longitudinal data are currently being collected in an independent cohort of similar size from a non-European population to address these limitations. Moreover, the study was initially designed as a 1-year investigation and was extended to a 3-year study, with recording periods every 6 months, to provide a more comprehensive assessment of disease progression over shorter intervals. Additionally, we observed a high dropout rate, along with a decline in adherence between the first and second recording periods. It is important to note that six of the 17 withdrawals were due to restructuring at one of the study sites. Although device-related withdrawals have rarely been reported, our data collection did not allow systematic determination of the exact reasons. To address potential device-related withdrawals and improve adherence, a new version of the wDHT has been developed to better meet patients' expectations regarding aesthetics and comfort. Additionally, comparison with age-matched healthy controls is essential to distinguish MS-related decline from normal aging. Some normative data have been published[31] but expanding the coverage of the adult population is necessary. Finally, we selected a short list of digital outcomes based on well-known concepts of interest in MS, a post hoc analysis could extend this list to explore new metrics and other concepts of interest. Future analyses may explore digital metrics to assess fatigability and upper limb function.

Nevertheless, this small-scale study demonstrated that individual stride detection using a valid and suitable wDHT in patients with MS exhibits key features, highlighting its practical potential as an early surrogate for ambulatory decline, a critical component of MS progression. Further research is essential to confirm these preliminary findings and to expand the potential applications of this technology. Efforts should prioritize the rigorous validation of digital outcome measures to assess ambulation in patients with MS, integrating these into regulatory frameworks. The next steps include validating these findings in an independent population and collecting age-matched normative data. Expanding data acquisition will help to establish meaningful thresholds for a decline in walking ability, whereas determining a meaningful threshold for improvement will represent a more challenging but crucial step toward validating a digital outcome: its integration into interventional trials as an exploratory endpoint.

Contributors

MP and LS conceived, designed and oversaw the studies.

EL, DE, AG and PS reviewed the study protocols for important intellectual content.

MP, BW, BD, DR, VV, AM, LM, IC and AD were responsible for data acquisition.

MP, OP, AT, CC, and LS accessed, reviewed, and verified the data.

MP, OP, AT, CC, DE and LS conducted the data analysis, while all authors contributed to data interpretation.

OP and AT designed the algorithms used for sensor data processing.

MP drafted the manuscript.

All authors reviewed the manuscript for important intellectual content and approved the final version.

Data sharing statement

The data supporting the findings of this study (ie published clinical data, SV95C, WD90C, and number of strides per hour) may be made available for academic and non-commercial purposes upon request to the corresponding author. Access may be granted after review and approval of a written research proposal compatible with patient consents and legitimate use, and contingent upon the signing of a data access agreement. The study protocols are provided as supplementary materials.

Declaration of interests

Damien Eggenspieler and Laurent Servais reports consulting fees from Sysnav. Helen Hayward-Koennecke reports being an employee of F. Hoffmann-La Roche Ltd. at the time the manuscript was drafted and reports a patent for a new variable of the Syde device, which is the intellectual property of F. Hoffmann-La Roche Ltd.

Emilie Lommers reports grant support from the F.R.S.–FNRS and reports honoraria, travel support, and participation on advisory boards from Sanofi, Merck, and Biogen. Barbara Willekens reports institutional support for the present manuscript from CHR-La Citadelle and reports grants, consulting fees, honoraria, and travel support from multiple organizations; all payments were made to her institution, including from Alexion, Almirall, Biogen, Janssen, Merck, Novartis, Roche, and Sanofi-Genzyme. Vincent van Pesch reports consulting

fees, honoraria, and travel support from multiple pharmaceutical companies, with all payments made to his institution, and reports unpaid participation in advisory and leadership roles. Bertrand Degos reports honoraria from Ipsen and Merz, travel support from Merz, and advisory board participation with Orion and Merz.

Anna-Victoria De Keersmaecker reports research funding from the Belgian Charcot Foundation. Margaux Poleur reports personal funding from the Belgian Fund for Scientific Research and grant support from Sysnav for the Acti-MS study. Mona Michaud reports receipt of study materials from Sysnav and reports grant support involving Sysnav and École Normale Supérieure Paris-Saclay.

Irène Coman and Damien Ricard declare no financial or non-financial interests related to this work.

Acknowledgements

This study was funded by F. Hoffmann-La Roche Ltd. Sysnav and F. Hoffmann-La Roche Ltd. collaborate on endpoint development in multiple sclerosis. Alexis Tricot, Oihana Piquet, and Celine Cluzeau report being employees of Sysnav. Alexandra Goodyear reports support for the present manuscript from Roche Genentech and states that she is an employee of Genentech. James Overell, Paul Strijbos, Paul Delmar and Qing Wang report being an employee of F. Hoffmann-La Roche Ltd. Helen Hayward-Koennecke reports being an employee of F. Hoffmann-La Roche Ltd. at the time the manuscript was drafted. Laurent Servais and Margaux Poleur report grant support from Sysnav for the Acti-MS study. Barbara Willekens, Vincent van Pesch, Bertrand Degos, Damien Ricard, and Annick Mélin report payments to their institution for study-related tasks from CHR–La Citadelle (Liège), the study sponsor.

During the preparation of this work the authors used DeepL Translate to translate the protocol of the controlled environment study from French into English (supplementary file 1). After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

We thank Georgia Stimpson for statistical support and Fardowsi Kazi for assistance with English language editing.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2026.103823.

References

1 Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of Digital Health Technologies in neurology trials. *NPJ Digit Med*. 2023;6(1):23. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2021407100&from=export.

2 Taylor KI, Staunton H, Lipsmeier F, Nobbs D, Lindemann M. Outcome measures based on digital health technology sensor data: data- and patient-centric approaches. *NPJ Digit Med*. 2020;3(1):97. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2005662785&from=export.

3 Servais L, Eggenspieler D, Poleur M, et al. First regulatory qualification of a digital primary endpoint to measure treatment efficacy in DMD. *Nat Med*. 2023;29(10):2391–2392. https://doi.org/10.1038/s41591-023-02459-5.

4 Micó-Amigo ME, Bonci T, Paraschiv-Ionescu A, et al. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J Neuroeng Rehabil*. 2023;20(1):78. https://doi.org/10.1186/s12984-023-01198-5.

5 van Unnik JWJ, Meyjes M, Janse van Mantgem MR, van den Berg LH, van Eijk RPA. Remote monitoring of amyotrophic lateral sclerosis using wearable sensors detects differences in disease progression and survival: a prospective cohort study. *eBioMedicine*. 2024;103:105104. Available from: https://doi.org/10.1016/j.ebiom.2024.105104.

6 Michaud M, Guérin A, Dejean de La Bâtie M, Bancel L, Oudre L, Tricot A. The analytical validity of stride detection and gait parameters reconstruction using the ankle-mounted inertial measurement unit syde®. *Sensors (Basel)*. 2024;24(8):2413. Available from https://www.embase.com/search/results?subaction=viewrecord&id=L644110634&from=export.

7 European Medicines Agency. EMA qualification opinion for SV95C as primary endpoint in studies in ambulatory Duchenne Muscular Dystrophy studies [cited 2025 Sep 15]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-primary-endpoint-studies-ambulatory-duchenne-muscular-dystrophy-studies_en.pdf.

8 Polhemus A, Ortiz LD, Brittain G, et al. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. *NPJ Digit Med*. 2021;4(1). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2013950595&from=export.

9 Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS. *Mult Scler*. 2020;26(14):1816–1821. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2007250194&from=export.

10 Attfield KE, Jensen LT, Kaufmann M, Friese MA, Fugger L. The immunology of multiple sclerosis. *Nat Rev Immunol*. 2022;22(12):734–750. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2016510512&from=export.

11 Lublin FD. New multiple sclerosis phenotypic classification. *Eur Neurol*. 2014;72:1–5. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L604018319&from=export.

12 Sriwastava S, Elkhooly M, Amatya S, et al. Recent advances in the treatment of primary and secondary progressive Multiple Sclerosis. *J Neuroimmunol*. 2024;390:578315. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2031409438&from=export.

13 Fox RJ, Bar-Or A, Traboulsee A, et al. Tolebrutinib in nonrelapsing secondary progressive multiple sclerosis. *N Engl J Med*. 2025;392(19):1883–1892. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2039077868&from=export.

14 Ontaneda D, Fox RJ, Chataway J. Clinical trials in progressive multiple sclerosis: lessons learned and future perspectives. *Lancet Neurol*. 2015;14(2):208–223. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L602195187&from=export.

15 European Medicines Agency. Draft qualification opinion of multiple sclerosis clinical 4 outcome assessment [cited 2025 Sep 15]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-qualification-opinion-multiple-sclerosis-clinical-outcome-assessment-mscoa_en.pdf.

16 Heesen C, Haase R, Melzig S, et al. Perceptions on the value of bodily functions in multiple sclerosis. *Acta Neurol Scand*. 2018;137(3):356–362. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L619612388&from=export.

17 Dalla Costa G, Nos C, Zabalza A, et al. A wearable device perspective on the standard definitions of disability progression in multiple sclerosis. *Mult Scler*. 2024;30(1):103–112. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2027947053&from=export.

18 Goldman MD, Chen S, Kunisetty B, Gelfand JM, Cree BAC, Block VJ. Validating wearable step counts in multiple sclerosis research: a replication study. *Front Neurol*. 2025;16:1709389. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2041176441&from=export.

19 Woelfle T, Bourguignon L, Lorscheider J, Kappos L, Naegelin Y, Jutzeler CR. Wearable sensor technologies to assess motor functions in people with multiple sclerosis: systematic scoping review and perspective. *J Med Internet Res*. 2023;25:e44428. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2026637499&from=export.

20 Food and Drug Administration. *Letter of intent regarding the use of Actibelt in multiple sclerosis*. 2018.

21 Karatsidis A, Angelini L, Scaramozza M, et al. Characterizing gait in people with multiple sclerosis using digital data from smartphone sensors: a proposed framework. *Mult Scler*. 2025;31(5):512–528. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2033358815&from=export.

22 Inojosa H, Schriefer D, Ziemssen T. Clinical outcome measures in multiple sclerosis: a review. *Autoimmun Rev*. 2020;19(5):102512.

Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2005258700&from=export.
23 Cederberg KLJ, Sikes EM, Bartolucci AA, Motl RW. Walking endurance in multiple sclerosis: meta-analysis of six-minute walk test performance. *Gait Posture*. 2019;73:147–153. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2002326145&from=export.
24 Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med*. 2020;3(1):55. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2004718071&from=export.
25 Parker RA, Scott C, Inácio V, Stevens NT. Using multiple agreement methods for continuous repeated measures data: a tutorial for practitioners. *BMC Med Res Methodol*. 2020;20(1):154.
26 Kirk C, Küderle A, Micó-Amigo ME, et al. Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. *Sci Rep*. 2024;14(1):1754. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L643289295&from=export.
27 Banks SA, Mazurek KA, Barnard L, Tobin WO, Ali F, Howe CL. Concordance of wearable sensors and depth cameras for quantitative gait assessment in multiple sclerosis. *Mult Scler Relat Disord*. 2025;104:106801. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2041330172&from=export.
28 Montalban X, Graves J, Midaglia L, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. *Mult Scler*. 2022;28(4):654–664. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2013153980&from=export.
29 Gidaro T, Gasnier E, Annoussamy M, et al. Home-based gait analysis as an exploratory endpoint during a multicenter phase 1 trial in limb girdle muscular dystrophy type R2 and facioscapulohumeral muscular dystrophy. *Muscle Nerve*. 2022;65(2):237–242. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2014305924&from=export.
30 Alexander S, Peryer G, Gray E, Barkhof F, Chataway J. Wearable technologies to measure clinical outcomes in multiple sclerosis: a scoping review. *Mult Scler*. 2021;27(11):1643–1656. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2005770090&from=export.
31 Poleur M, Ulinici A, Daron A, et al. Normative data on spontaneous stride velocity, stride length, and walking activity in a non-controlled environment. *Orphanet J Rare Dis*. 2021;16(1):318. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2013196063&from=export.