Original Research Paper

A digital version of the nine-hole peg test: Speed may be a more reliable measure of upper-limb disability than completion time in patients with multiple sclerosis

Xiaotong Jiang [iD], Marisa McGinley [iD], Joshua Johnston, Jay Alberts, Robert Bermel [iD], Daniel Ontaneda [iD], Robert T Naismith [iD], Robert Hyde*, Nick Levitt*, Johan van Beek*, Zhaonan Sun, Nolan Campbell* and Christian Barro* [iD]

Abstract

Background: A digital adaptation of the nine-hole peg test (9HPT) was developed with the potential to provide novel disability features for patients with multiple sclerosis (PwMS).

Objectives: The objectives were to evaluate the 9HPT features based on reliability, prognosis, and discrimination between treatment groups.

Methods: The MS partners Advancing Technology and Health Solutions (MS PATHS) cohort data were used to derive new features including completion time and speed. Association and reliability between features and clinical outcomes were tested by intraclass correlation coefficients (ICCs) with repeated measures. The added prognostic value of the features for a clinically meaningful decline was assessed by time-to-event analyses with likelihood ratio tests. The estimated effect size between treatment efficacy groups was acquired from linear mixed-effects models. Sample size was calculated for a hypothetical randomized clinical trial.

Results: For the 10,843 PwMS, speed and completion time were associated with MS disability. Compared with time, speed showed higher reliability (ICC = 0.78 vs 0.74), added benefits in predicting disability worsening ($p < 0.001$), better discrimination between high- and low-efficacy groups (effect size: 0.035 vs 0.015), and an 18% reduction in required sample size for a 1-year clinical trial.

Conclusion: Integrating horizontal hand distances traveled over the 9HPT pegboard can be a more reliable measure of hand function.

Correspondence to:
X Jiang
Biogen, 225 Binney St,
Cambridge, MA 02142,
USA.
phoebe.jiang@biogen.com

Xiaotong Jiang
Robert Hyde
Nick Levitt
Johan van Beek
Zhaonan Sun
Nolan Campbell
Christian Barro
Biogen, Cambridge, MA,
USA

Marisa McGinley
Robert Bermel
Daniel Ontaneda
Mellen Center, Cleveland
Clinic, Cleveland, OH, USA

Joshua Johnston
Jay Alberts
Biomedical Engineering,
Cleveland Clinic, Cleveland,
OH, USA

Robert T Naismith
Washington University
School of Medicine, St.
Louis, MO, USA

*Employee of Biogen at time
of this study.

Introduction

Multiple sclerosis (MS) is a highly debilitating disease with a heterogeneous clinical presentation.[1] Accruing in multiple functional domains, disability is the result of acute disease activity or chronic neurodegeneration.[2] Using disease-modifying therapies (DMTs) resulted in a dramatic reduction in acute disease activity measured as the number of relapses or focal lesions on magnetic resonance imaging.[3] However, irreversible disability worsening continues to occur, and functional recovery by remyelination remains elusive to therapeutic efforts.[4,5]

Current gold standard measurements of disability such as the Expanded Disability Status Scale (EDSS) have greatly benefited the MS community. The EDSS has known drawbacks, including inadequate reflection of cognition, fatigue, and mood deficits, which are most troublesome for patients with multiple sclerosis (PwMS). It is prone to human errors, heavily impacted by ambulatory functions, and lacks disability confirmation within the same functional system, potentially missing disability worsening events and potential treatment effects that focused measures such as the nine-hole peg test (9HPT) can capture.[3] These

novel therapeutical landscapes require rethinking our ways of measuring disease stability. The 9HPT, which is more objective and focused on a specific functional domain, may better capture meaningful disability changes. Previous attempts to slow disease progression improved manual dexterity in the 9HPT, but not disability as quantified by the EDSS.[6,7]

Completion time of the 9HPT is the gold standard of upper-limb function testing in PwMS, but does not reveal which movement aspects (e.g. peg movement strategy, speed, distance) caused delays in completion.[8] In this study, we developed methods to identify new biomarkers and validate a digital adaptation of the 9HPT, the manual dexterity test (MDT), within the software suite—the Multiple Sclerosis Performance Test (MSPT).[9,10] Effective digital biomarkers are crucial for detecting "silent" progression, such as disability worsening in the absence of acute attacks or subclinical disease activities like brain atrophy, and for identifying suboptimal treatment responses more quickly. While treatment escalation is often delayed due to adherence issues and therapeutic inertia, the data collected from digital monitoring have the potential for timely detection of disease progression and treatment response. In addition to capturing the standard time to complete the task, the MDT records location and timing of each peg movement, allowing the capture of metrics such as speed, peg movement strategy, and total distance required to complete the task as separate potential sources of variability in task performance.[9]

In this study, we aimed to (1) explore the relationship between intra-task features and MS disease progression; (2) determine if speed is a more reliable measure of MS-related decline in upper-limb function than the overall completion time; (3) quantify the potential benefits of using speed versus completion time when predicting disability progression or quantifying treatment effects.

Methods

Study design

This study leveraged clinical assessments collected as part of the MS partners Advancing Technology and Health Solutions (MS PATHS), a multicenter learning health system that enrolled and monitored patients between 2016 and 2022.[11] Patients were included if they were under the care of a physician at an MS center participating in MS PATHS and had a confirmed diagnosis of MS or clinically isolated syndrome as determined by clinicians (based on available clinical data and diagnostic criteria available at the time of diagnosis). The database lockdown date for the current analysis is 13 July 2020.

Study visits and assessments

The MDT is an iPad-based version of the traditional 9HPT, part of the MSPT.[9,12] A finite-state machine algorithm was implemented to define the functionality of MSPT and preprocess the raw MDT data, where a state was defined as the set of pegs that were being held up at each moment (Supplementary Figure 1). The MDT assessments were deemed valid if comprising 18 pairs of peg movements (36 in all) of the nine pegs being picked up and then put down into a different region (first 9 from row to grid, then 9 from grid to row).

We investigated four main features for each MDT assessment, identified based on clinical relevance and statistical correlation with MS disability:

1. completion time in second (s)=the total time used to complete an MDT assessment;
2. total distance of hand traveled in centimeter (cm)=the sum of distance of all horizontal hand movements over the peg board, including hand movements without the peg in hand (after dropping down a peg and before picking up the next peg);
3. speed in cm/s=total distance of hand traveled divided by completion time;
4. peg movement strategy=the number of positions that differed between the adopted movement strategy and the most common movement strategy. (This is known as the Hamming distance in information theory to determine the similarity between two strings. These two strings must be of equal length. In our case, the peg movement strategy had been preprocessed to be a string of 36 peg positions for moving nine pegs between row and grid. We utilized the Hamming distance to quantify the difference between each peg strategy with the most common one.)

These features were derived from peg-timing raw data acquired during routine clinical testing for the dominant hand. For each visit, we also had access to the demographic and clinical parameters including Patient Determined Disease Step (PDDS), Processing Speed Test (PST), Quality of Life in Neurological Disorders (NQ), and treatment information.

Statistical methods

Overall analyzed population. Baseline patient characteristics were summarized as mean and standard deviation (SD) for continuous or count variables and proportion (%) for categorical variables. Baseline was defined as the first valid MDT assessment using the dominant hand.

Intra-patient correlation analysis. Intra-patient hand distance range, defined as the difference between maximum distance and minimum distance among MDT assessments of the same patient, was calculated in patients with ≥2 cleaned MDT assessments. Intra-patient longitudinal change associations between the four MDT features and MS-related disability (PDDS, PST, NQ) were assessed by repeated measures correlations among patients with ≥6 months of follow-up and ≥5 cleaned MDT assessments. The 95% confidence intervals (CI) provided a measure of variability.

Test–retest analysis. The reliability, intra-patient agreement, and responsiveness for completion time versus speed were compared among patients with ≥2 cleaned MDT assessments performed within 7 to 60 days from each other. Reliability was quantified by intraclass correlation coefficient (ICC). Agreement was visualized by the Bland Altman plots. We calculated the minimal detectable change with 95% CI (MDC95), defined as the smallest change in completion time or speed that can be detected with 95% CIs, to distinguish it from random measurement error (details in Supplementary Materials). Responsiveness was assessed in terms of the number of patients whose worst and best performance difference exceeded the MDC95 in the treatment initiation analysis below.

Prognostic analysis. The prognostic value of speed/time for a clinically meaningful decline was assessed on the event: a 3-month confirmed 20% worsening in completion time occurred >2 years post-baseline. This survival analysis included patients with ≥3 cleaned MDT assessments in years 0–2 and ≥2 cleaned MDT assessments after 2 years. Three Cox proportional hazards models were applied to predict the event: (1) a nested model using *completion time* as the predictor; (2) a nested model using *speed* as the predictor; and (3) a full model using both *speed and completion time* as predictors (details in Supplementary Materials). The Akaike and Bayesian information criterion (AIC and BIC) informed on model performance. Likelihood ratio tests compared full and nested models to evaluate the added prognostic value of either feature.

Treatment initiation analysis. This analysis included patients who started a new DMT with ≥3 cleaned MDT assessments, ≥180 days of follow-up, and known previous treatment group. To help us understand and capture the efficacy component of the treatment, we focused on patients who started a new treatment in this analysis. High-efficacy treatments offer better control over MS progression but carry higher risks, whereas low-efficacy treatments are safer with moderate effectiveness. To assess the ability to discriminate between treatment efficacy groups, linear mixed-effects models were applied to the annualized changes in speed or time relative to baseline (the outcome) and were weighted to balance baseline characteristics (details in Supplementary Materials). Effect sizes and variances between the high versus low DMT efficacy groups were estimated. The sample size needed for a hypothetical 1-year randomized clinical trial with 80% power and 5% type-1 error was calculated for time and speed.

To assess disability responsiveness, we calculated the overall disability response score (ODRS) based on changes relative to baseline in speed and time. This integrated metric was used to evaluate disability improvement and worsening for PwMS over time and was adjusted and estimated with linear mixed-effect models (details in Supplementary Materials).[13] A *t*-test compared the difference in adjusted ODRS at week 48 between high- and low-efficacy groups. Analyses were conducted in Python 3.9 and R 4.2.2 with 0.05 as the significance level.

Results

MS PATHS overall analyzed population

Patient characteristics. After preprocessing, 90% of the original raw MDT assessments remained in the analysis; 5% assessments were canceled and 5% were removed due to errors during the task performance (e.g. fumbles, bumps), duplicate trials, and > or <36 peg movements. Fumbles refer to instances where the individual mishandled the peg, including dropping, losing grip, or failing to properly grasp or manipulate the peg as intended such that it fell back into its original positions. Bumps refer to instances where the individual inadvertently hits the pegboard, causing pegs to jump and temporarily lose connection with the pegboard. Both instances can result in missing or disturbed data, making it difficult to process. Given the low frequency of these events, we chose to exclude them. Overall, patient demographic and clinical characteristics were similar across analyses (Table 1).

Table 1. Baseline and follow-up characteristics of patients with MDT assessments.

|	Analysis population			
	Overall analyzed population	Intra-patient correlation analysis	Test–retest analysis	Prognostic analysis
N (unique patients)	10843	3239	812	1213
Baseline patient characteristics				
Age at symptom onset, y, mean (SD)	33 (11)	32 (11)	31 (11)	32 (11)
Missing, *n* (%)	688 (6)	25 (1)	74 (9)	5 (<1)
Age at baseline, years, mean (SD)	47 (12)	45 (11)	43 (12)	45 (11)
Missing, *n* (%)	613 (6)	22 (1)	69 (9)	10 (1)
Female, *n* (%)	7544 (70)	2373 (73)	561 (69)	894 (74)
Missing, *n* (%)	574 (5)	–	62 (8)	–
Years of education, *n* (%)				
0-12	3128 (29)	1053 (33)	224 (28)	366 (30)
13-16	4888 (45)	1469 (45)	365 (45)	561 (46)
17-20	2253 (21)	717 (22)	161 (20)	286 (24)
Missing (or not requested), *n* (%)	574 (5)	–	62 (8)	–
PDDS score, *n* (%)				
0	3873 (36)	1155 (36)	257 (32)	410 (34)
1	1883 (17)	621 (19)	142 (18)	245 (20)
2-3	2273 (21)	785 (24)	200 (25)	310 (26)
≥4	2143 (20)	644 (20)	141 (17)	237 (20)
Missing, *n* (%)	671 (6)	34 (1)	72 (9)	11 (1)
Had ≥1 relapse in the past 12 months, *n* (%)	4774 (44)	1484 (46)	401 (50)	585 (48)
Missing, *n* (%)	714 (7)	39 (1)	77 (10)	9 (1)
MS duration, years, mean (SD)	12 (9)	11 (8)	10 (8)	11 (8)
Missing, *n* (%)	877 (8)	80 (3)	94 (12)	29 (2)
Post-baseline patient follow-up statistics				
Number of MDT assessments per patient, median (Q1–Q3)	3 (2–5)	6 (5–8)	N/A	7 (6–9)
Number of follow-up days, median (Q1–Q3)	628 (372–835)	840 (718–974)	N/A	986 (916, 1099)

MDT: manual dexterity test; MS: multiple sclerosis; N/A: not applicable; PDDS: Patient-Determined Disease Steps; Q1: 25th percentile; Q3: 75th percentile; SD: standard deviation.

Intra-patient hand distance range. The distance traveled by the same patient varied substantially across longitudinal MDT assessments (Figure 1), with 63% >10 cm differences and 28% >20 cm differences.

Association of intra-task features with MS-related disability. Speed and completion time, but not hand distance or peg strategy, were consistently associated with MS disability, cognitive performance, and quality of life (Figure 2). In our sensitivity analysis, associations were consistent in baseline impaired population and population with 20% longitudinal worsening on upper-limb function.

Reliability responsiveness and agreement. In the test–retest subset ($n=812$), the estimated ICC was 0.74 (95% CI=0.71–0.77) for completion time and 0.78 (95% CI=0.75–0.80) for speed. The test–retest variability increased with increasing completion time but remained relatively consistent across the range of speed (Figure 3). The MDC95 (%MDC95) was 3.6 cm/s (25.4%) for speed and 9.2 seconds (23.3%)
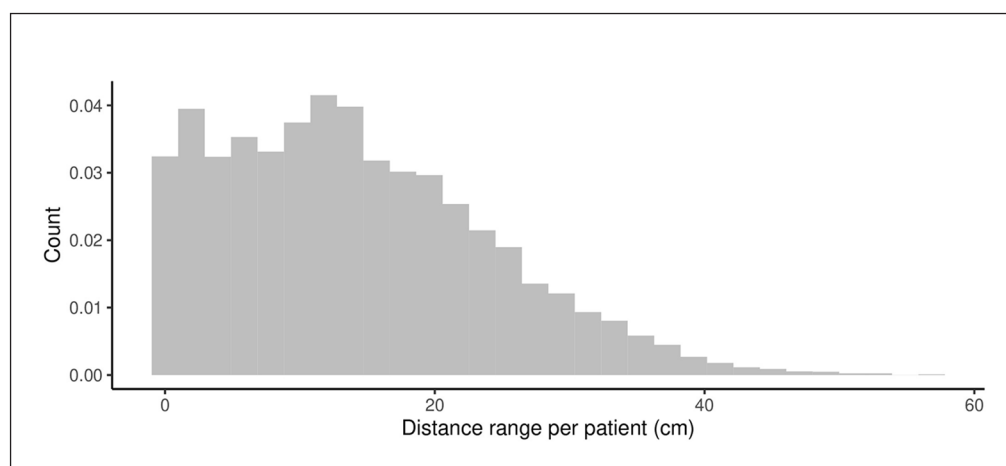
Figure 1. Intra-patient distance range (cm).
Range = maximum distance minus minimum distance in centimeter, among assessments of the same patient.

for completion time. This confirmed the cut-off for a meaningful change around 20%, already broadly used for completion time and now here also shown for speed.[13] Applied to the treatment initiation cohort, more PwMS exceeded the %MDC95 threshold based on speed ($n=255$, 20%) than completion time ($n=177$, 14%).

Prognostic value of speed and completion time. All Cox models suggested PDDS and prior relapse activity as significant predictors (Table 2). Both features by themselves were significant for predicting completion time to subsequent meaningful decline (both $p<0.001$). When both included, speed remained a significant predictor ($p=0.001$), while completion time did not ($p=0.168$). Model performance was superior when speed was added to time ($p<0.001$), but there was no difference in performance when time was added to speed ($p=0.174$). Sensitivity analyses were performed using the first assessment as baseline and 6-month confirmation as the event, and results remained consistent; however, the 6-month confirmation results showed wider CIs compared to the 3-month results in Table 2, likely due to the stricter event definition and consequently a smaller sample size.

In addition, speed detected more 20% confirmed improvement events than completion time ($n=150$, 12% vs 97, 8%, respectively). Patients who improved had a higher average completion time (29 seconds vs 24 seconds), likely explaining the worse performance of completion time in detecting improvement events. The results of 20% confirmed worsening events were similar between speed and time ($n=55$, 4% vs 73, 6%, respectively).

Treatment initiation analysis population
Patient characteristics. The high-efficacy DMT group included PwMS treated with ocrelizumab (71%), natalizumab (15%), alemtuzumab (9%), and rituximab (6%). The low-efficacy group included teriflunomide (41%), glatiramer acetate (38%), interferon beta-1a (18%), and peginterferon beta-1a (3%). Relative to the low-efficacy group, patients in the high-efficacy group on average had higher PDDS scores and were younger, more likely to be treated previously, and less likely to be female, have relapsing-remitting multiple sclerosis (RRMS) (vs progressive MS), or relapses in the past 12 months (Supplementary Table 1).

Ability to discriminate between treatment efficacy groups. With covariate balance adjusted (Supplementary Figure 2), the effect size in absolute value for annualized changes over time between the efficacy groups was higher for speed compared with completion time (Figure 4). This indicated that speed can potentially better differentiate known group treatment effects than time. Based on the effect sizes and assuming 1:1 arm assignment, speed required 18% fewer samples than time ($n=2906$ vs 3536).

Mean ODRS improved in both efficacy groups when ODRS was defined by speed, but not by completion time: The mean (95% CI) adjusted ODRS at week 48 in terms of speed was 0.08 (0.01, 0.15) for low-efficacy and 0.03 (−0.01, 0.08) for high-efficacy; both scores were positive indicating improvement. The mean (95% CI) ODRS at week 48 in terms of completion time as 0.01 (−0.06, 0.08) for low-efficacy group and −0.04 (−0.09, −0.003) for high-efficacy group; the negative value indicated worsening. We found no
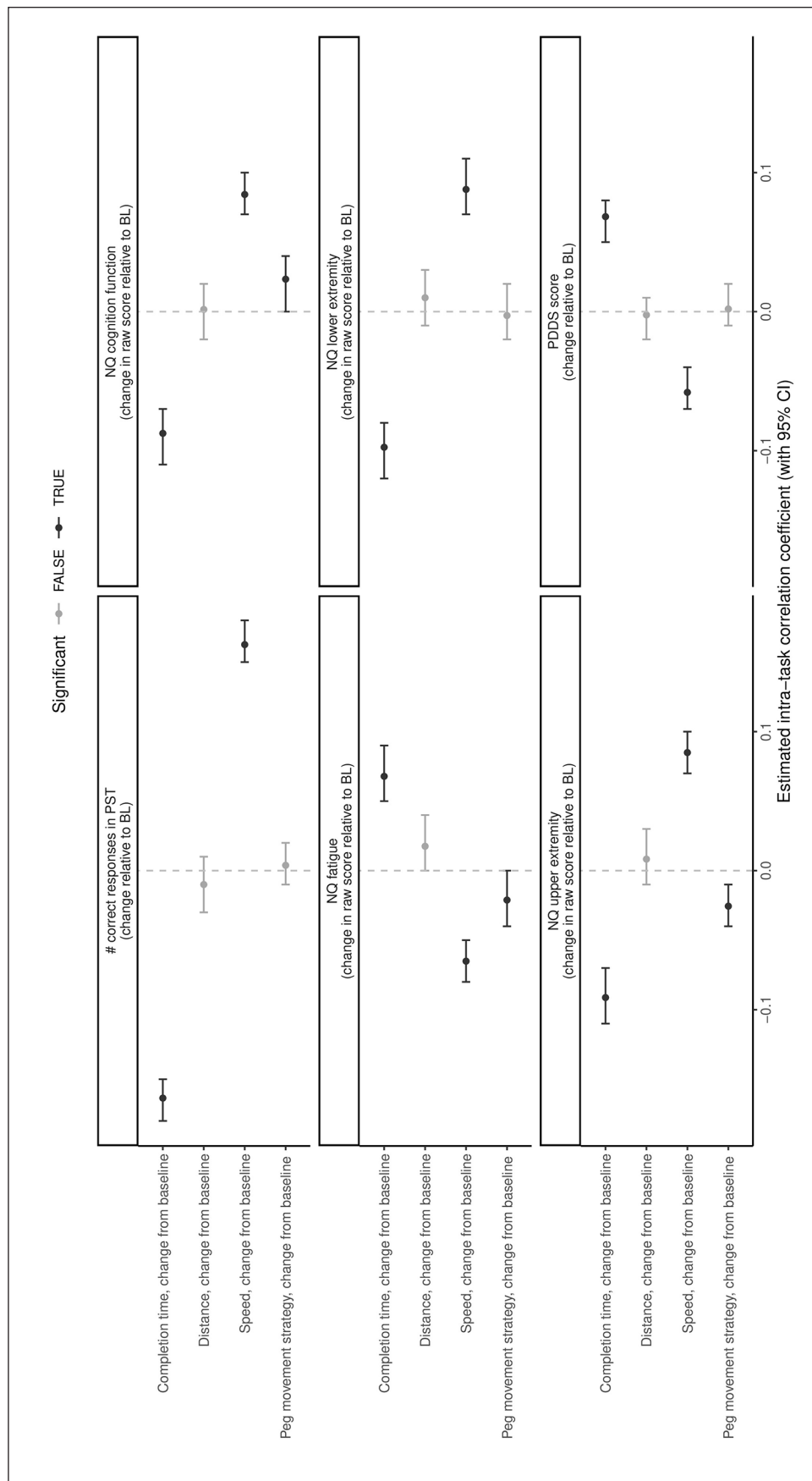
Figure 2. Intra-patient longitudinal correlations between main MDT features and MSPT assessments (BL: baseline, MDT: manual dexterity test, MSPT: Multiple Sclerosis Performance Test, NQ: NeuroQoL, PDDS: Patient-Determined Disease Steps).
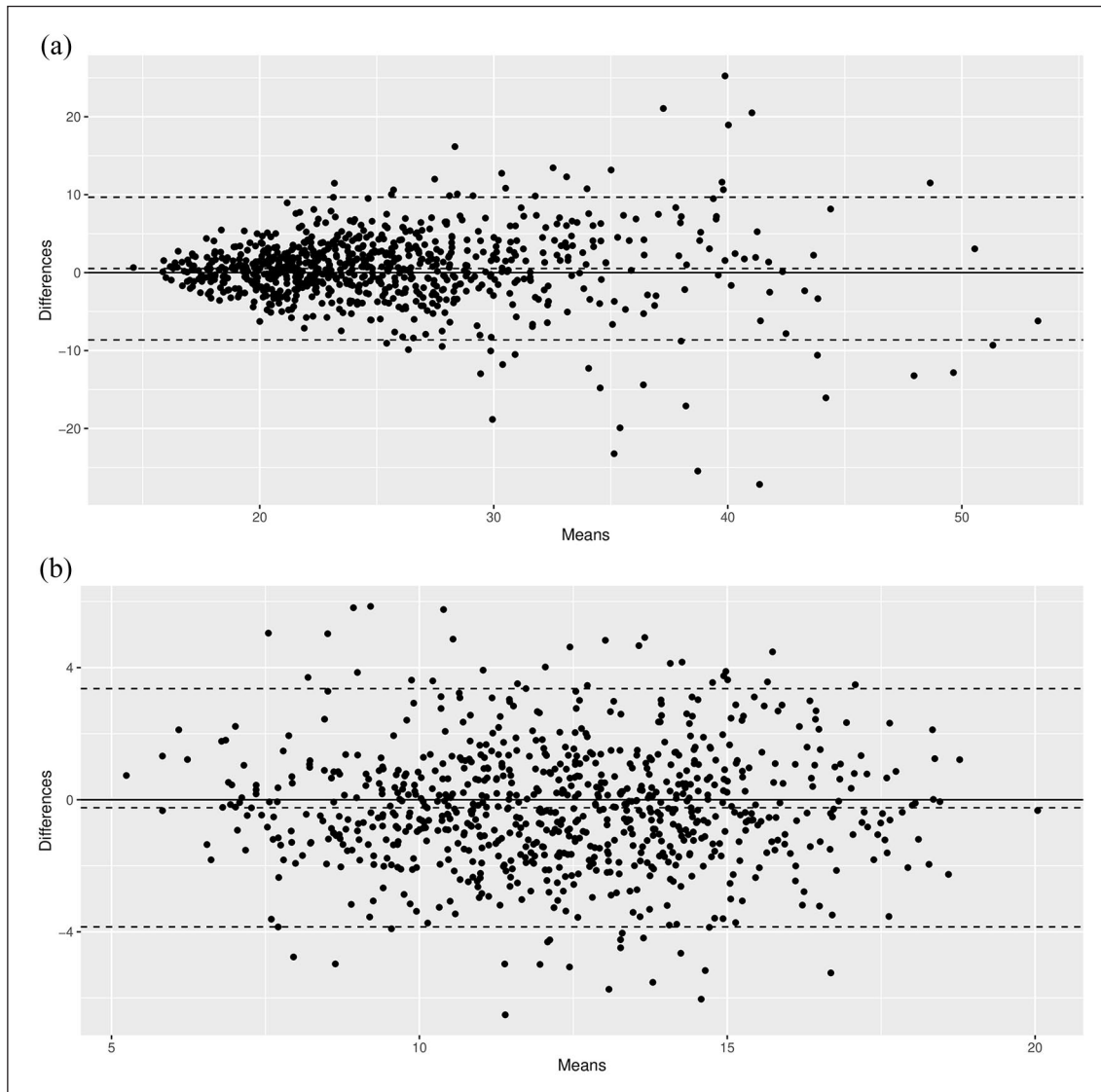
Figure 3. Bland–Altman plot for comparison of the tests and retests on completion time (a) and speed (b).

statistically significant difference between treatment groups in speed or time: the difference of high- versus low-efficacy in mean ODRS at 48 weeks was −0.04 (*p*=0.48) for speed and −0.06 (*p*=0.28) for time. There was also no significant separation in time to confirmed 20% worsening or improvement between the efficacy groups (Supplementary Figure 3).

Discussion

Recent MS studies increasingly incorporate digital tools such as tablets, smartphones, and wearable sensors (smart watches and visual reality) to provide detailed insights into the impact of MS on hand functions.[14–18] A recent study confirmed the reliability of a

mobile program called the Digital Self-Assessment for MS, showing a high correlation with EDSS and nuances that Multiple Sclerosis Functional Composite (MSFC) could not capture.[14] Another study on the Floodlight app found that finger velocity in pinching tests had a fair correlation with 9HPT, capturing unique limb movement characteristics not represented in standard clinical assessments.[15,19] Digital tools enhance traditional methods by capturing multiple metrics simultaneously, such as location, reaction time, and performance variability, revealing subtle changes overlooked by conventional methods and offering a more comprehensive view of hand function, further deepening our understanding of disease progression and treatment efficacy.

Table 2. Survival modeling[a] results of using standardized changes[b] in speed versus completion time to predict time to 3-month confirmed 20% worsening in terms of completion time.[c]

Variable	Completion time only (Model 1)			Completion time and speed (Model 2)			Speed only (Model 3)		
	Estimates	95% CI	p	Estimates	95% CI	p	Estimates	95% CI	p
Total time, standardized change[b]	2.16	1.89–2.47	<0.001	1.27	0.90–1.79	0.168			
Female	0.83	0.45–1.51	0.539	0.90	0.49–1.66	0.736	0.93	0.51–1.72	0.826
Age at symptom onset (years)	1.01	0.99–1.04	0.270	1.01	0.98–1.04	0.436	1.01	0.98–1.03	0.516
Baseline PDDS score 1[d]	1.02	0.33–3.16	0.968	1.05	0.34–3.27	0.934	1.01	0.33–3.16	0.980
Baseline PDDS score 2 or 3[d]	2.25	0.95–5.30	0.065	2.50	1.06–5.90	0.036	2.68	1.15–6.24	0.023
Baseline PDDS score ≥4[d]	2.16	0.91–5.13	0.080	2.41	1.02–5.71	0.046	2.51	1.06–5.95	0.036
Relapsed previously (yes vs no)	1.47	0.81–2.67	0.208	1.46	0.81–2.65	0.210	1.51	0.84–2.71	0.172
13–16 years of education[e]	1.22	0.65–2.30	0.535	1.12	0.60–2.10	0.721	1.08	0.58–2.01	0.817
17–20 years of education[e]	0.89	0.39–2.04	0.791	0.88	0.39–2.02	0.767	0.90	0.39–2.06	0.806
Speed, standardized change[b]				0.41	0.24–0.68	0.001	0.30	0.24–0.37	<0.001

Model	Description	AIC	BIC	LRT
Model 1 (nested)	Completion time only	615	633	Model 1 vs 2: $p < 0.001$
Model 2 (full)	Completion and speed	606	625	Model 2 vs 3: $p = 0.174$
Model 3 (nested)	Speed only	606	623	

AIC: Akaike information criterion; BIC: Bayesian information criterion; LRT: likelihood ratio test; MDS: manual dexterity test; PDDS: patient determined disease steps; SD: standard deviation.
[a]All models adjusted for the following baseline characteristics: age at symptom onset, sex, previously relapsed (yes/no), PDDS score, and number of years of education.
[b]Standardized changes are changes relative to baseline and standardized by the SD of the corresponding measure (time or speed).
[c]The number of observations is 7733, which corresponds to 1213 unique patients.
[d]Reference group: baseline PDDS score=0.
[e]Reference group: 0–12 years of education.

In this study, we investigated the advancements of a digital adaptation of the 9HPT via a comprehensive series of analyses. We found a large spread in peg strategy and distance traveled between longitudinal assessments for the same patient, but these longitudinal changes in distance traveled were not associated with objective or subjective MS disease changes. This variability in distance traveled is a potential confounder that can be minimized by utilizing the speed to execute the test, which demonstrated more consistent and favorable measurement properties. Speed was also a more sensitive predictor of disability than completion time, effectively distinguishing treatment effects between high- and low-efficacy groups. This allowed speed to be more responsive to improvements in hand function, increasing efficiency and reducing the required sample size as the primary endpoint in randomized clinical trials.

We also explored the potential advantage of using speed over time in distinguishing between treatment groups. While the advantage was evident in annualized changes, using speed did not withstand the 20% threshold for a clinically meaningful change. There was no clinically meaningful difference between the low- and high-efficacy groups as defined by a worsening or improvement above a 20% threshold for *either* completion time or speed. The difference in detecting disability improvement can be explained by two observations: (1) improvement events were more common in patients with longer baseline completion times and (2) the reliability of completion time
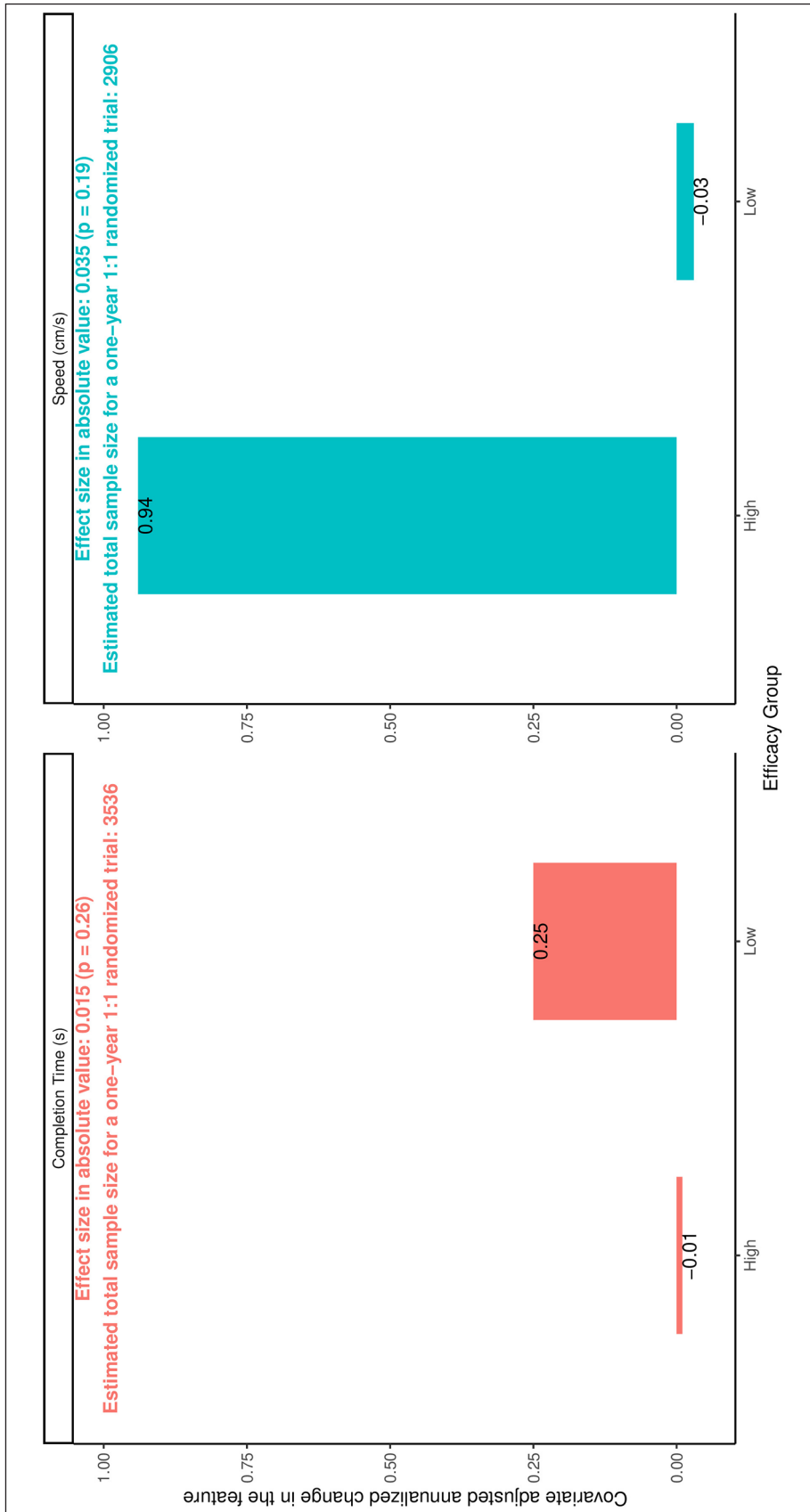
Figure 4. Covariate-adjusted annualized changes over time in completion time and speed for high- and low-efficacy groups and the corresponding effect size in absolute value. Speed and time are inversely related by definition. As speed increases, the time required to cover a given distance decreases, and vice versa. This inverse relationship explains why results involving these variables showed opposite signs or directions and why the effect size was calculated in absolute values.

decreased with increasing baseline completion time, which was not observed for speed (Figure 3).

Several studies have included speed in functional assessments for PwMS. For the evaluation of hand function, one study using a Virtual Peg Insertion Test found that speed, together with smoothness and grip force control, was most affected in PwMS.[20] Another study measured finer opposition movements with sensor-engineered gloves, showing a significantly lower movement rate in both spontaneous and maximal velocity conditions. Other speeds, such as walking speed from the Timed 25-Foot Walk Test and information processing speed, have also been shown to predict disability and are recommended for continued use in clinical practice and trials.[21,22] Implementing such digital tools in clinical practice involves setting up the necessary software and hardware, training staff, educating patients, and integrating the data with existing data platforms. Regular monitoring and continuous feedback are essential for effective use and process improvement.

Typical shortcomings of observational data apply here. The data were noisy, contained self-reported disease information (which may be less reliable than clinical measurements), and had missing data due to arbitrary lost-to-follow-ups. Practice effects might persist despite our efforts to mitigate them by setting the second assessment as a reference. Hand distance was calculated based on direct physical distance between peg positions two-dimensionally. Capturing the three-dimensional (3D) movement of the hand could improve the use of speed over completion time as an outcome measure in clinical trials by better accounting for the hand distance traveled. Finally, all comparisons between speed and time were made within the 9HPT task, which prevents extrapolating our conclusions to cover the most used manual board. Future work should directly compare speed measurements from the digital 9HPT with completion time from the manual 9HPT and include studies with longer follow-ups.

In summary, the digital 9HPT introduced a novel outcome measure of execution speed with improved measurement properties compared to the traditional completion time. By accounting for differences in traveled distance, we achieved a more accurate assessment of hand function.

ORCID iDs
Xiaotong Jiang https://orcid.org/0000-0003-3698-4526
Marisa McGinley https://orcid.org/0000-0002-7463-6787
Robert Bermel https://orcid.org/0000-0003-2334-6883
Daniel Ontaneda https://orcid.org/0000-0002-2838-9148
Robert T Naismith https://orcid.org/0000-0003-0520-4283
Christian Barro https://orcid.org/0000-0002-7795-7383

Supplemental Material
Supplemental material for this article is available online.

References

1. Compston A and Coles A. Multiple sclerosis. *Lancet* 2008; 372(9648): 1502–1517.

2. Lublin FD, Häring DA, Ganjgahi H, et al. How patients with multiple sclerosis acquire disability. *Brain* 2022; 145(9): 3147–3161.

3. Lorscheider J, Buzzard K, Jokubaitis V, et al. Defining secondary progressive multiple sclerosis. *Brain* 2016; 139(9): 2395–2405.

4. 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(9): 1132–1140.

5. Cadavid D, Mellion M, Hupperts R, et al. Safety and efficacy of opicinumab in patients with relapsing multiple sclerosis (SYNERGY): A randomised, placebo-controlled, phase 2 trial. *Lancet Neurol* 2019; 18(9): 845–856.

6. Kapoor R, Ho PR, Campbell N, et al. Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): A phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension. *Lancet Neurol* 2018; 17(5): 405–415.

7. Cohen J, Cutter G, Fischer J, et al. Benefit of interferon β-1a on MSFC progression in secondary progressive MS. *Neurology* 2002; 59(5): 679–687.

8. 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(5): 711–720.

9. Rudick RA, Miller D, Bethoux F, et al. The Multiple Sclerosis Performance Test (MSPT): An iPad-based disability assessment tool. *J Vis Exp* 2014; 88: e51318.

10. Rao SM, Galioto R, Sokolowski M, et al. Multiple sclerosis performance test: Validation of self-administered neuroperformance modules. *Eur J Neurol* 2020; 27(5): 878–886.

11. McGinnis JM, Stuckhardt L, Saunders R, et al. *Best care at lower cost: The path to continuously learning health care in America*. Washington, DC: National Academies Press, 2013.

12. Rhodes JK, Schindler D, Rao SM, et al. Multiple sclerosis performance test: Technical development and usability. *Adv Ther* 2019; 36(7): 1741–1755.

13. Chang I, Kappos L, Giovannoni G, et al. Overall Disability Response Score: An integrated endpoint to assess disability improvement and worsening over time in patients with multiple sclerosis. *Mult Scler* 2022; 28(14): 2263–2273.

14. De Angelis M, Lavorgna L, Carotenuto A, et al. Digital technology in clinical trials for multiple sclerosis: Systematic review. *J Clin Med* 2021; 10(11): 2328.

15. Baker M, van Beek J and Gossens C. Digital health: Smartphone-based monitoring of multiple sclerosis using Floodlight. *Nature Portfolio*, https://www.nature.com/articles/d42473-019-00412-0

16. Van Hecke W, Costers L, Descamps A, et al. A novel digital care management platform to monitor clinical and subclinical disease activity in multiple sclerosis. *Brain Sci* 2021; 11(9): 1171.

17. Hsu W-Y, Rowles W, Anguera JA, et al. Assessing cognitive function in multiple sclerosis with digital tools: Observational study. *J Med Internet Res* 2021; 23(12): e25748.

18. Oh J, Capezzuto L, Kriara L, et al. Use of smartphone-based remote assessments of multiple sclerosis in Floodlight Open, a global, prospective, open-access study. *Sci Rep* 2024; 14(1): 122.

19. Graves JS, Elantkowski M, Zhang Y-P, et al. Assessment of upper extremity function in multiple sclerosis: Feasibility of a digital pinching test. *JMIR Form Res* 2023; 7: e46521.

20. Kanzler CM, Sylvester R, Gassert R, et al. Goal-directed upper limb movement patterns and hand grip forces in multiple sclerosis. *Mult Scler J Exp Transl Clin* 2022; 8(3): 20552173221116272.

21. Kalinowski A, Cutter G, Bozinov N, et al. The timed 25-foot walk in a large cohort of multiple sclerosis patients. *Mult Scler* 2022; 28(2): 289–299.

22. Ezegbe C, Zarghami A, van der Mei I, et al. Instruments measuring change in cognitive function in multiple sclerosis: A systematic review. *Brain Behav* 2023; 13(6): e3009.