Invited Review

The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis

Peter Feys, Ilse Lamers, Gordon Francis, Ralph Benedict, Glenn Phillips, Nicholas LaRocca, Lynn D Hudson, Richard Rudick and Multiple Sclerosis Outcome Assessments Consortium

Abstract: Impaired manual dexterity is a frequently reported disability in people with multiple sclerosis (MS) and is increasingly prevalent with worsening disease. While various tests and patient-reported outcome measures are available, the Nine-Hole Peg Test (NHPT) is considered as a gold standard measure of manual dexterity and most frequently used in MS research and clinical practice. The MS Outcome Assessments Consortium (MSOAC) includes representatives from advocacy organizations, Food and Drug Administration (FDA), European Medicines Agency (EMA), National Institute of Neurological Disorders and Stroke (NINDS), academic institutions, and industry partners along with persons living with MS. Among the MSOAC goals are acceptance and qualification by regulators of performance outcomes that are highly reliable and valid, practical, cost-effective, and meaningful to persons with MS. A critical step for these neuroperformance metrics is elucidation of clinically relevant benchmarks, well-defined degrees of disability, and gradients of change that are deemed clinically meaningful. This article addresses the NHPT, the proposed MSOAC measure for upper extremity function. We find that the NHPT is reliable within and between test sessions, discriminates between healthy subjects and MS patients with different levels of upper limb impairment, and shows high convergent validity with other manual dexterity as well as more comprehensive upper limb measures. Ecological validity is established by its relation to perceived upper limb use in daily life and perceived difficulty in performing activities of daily living. The NHPT is responsive to deterioration in longitudinal studies, and research suggests that a 20% change in test score is commonly used to define clinically meaningful worsening, a definition that needs further validation in all stages of the disease.

Keywords: Multiple sclerosis, upper extremity, performance outcomes, psychometric properties, Nine-Hole Peg Test, assessment

Correspondence to:
P Feys
Rehabilitation Research Center (REVAL), Biomedical Research Institute (BIOMED), Faculty of Medicine and Life Sciences, Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium.
peter.feys@uhasselt.be

Peter Feys
Ilse Lamers
Rehabilitation Research Center (REVAL), Biomedical Research Institute (BIOMED), Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

Gordon Francis
Neuroscience Clinical Development, San Francisco, CA, USA

Ralph Benedict
Department of Neurology, University at Buffalo, Buffalo, NY, USA

Glenn Phillips
Richard Rudick
Biogen, Cambridge, MA, USA

Nicholas LaRocca
National Multiple Sclerosis Society, New York, NY, USA

Lynn D Hudson
Critical Path Institute, Tucson, AZ, USA

Multiple Sclerosis Outcome Assessments Consortium
Multiple Sclerosis Outcome Assessments Consortium (MSOAC), Critical Path Institute, Tucson, AZ, USA

Overview and historical perspective

Like walking disability, visual problems, and cognitive deficits, upper limb dysfunction is a core deficit affecting multiple sclerosis (MS) patients. A combination of predominantly motor and sensory symptoms causes upper limb disability, which hampers the ability to perform activities of daily living (ADLs) and social activities, resulting in a decreased quality of life.[1] Upper limb disability in MS patients may present in the proximal or distal parts of the upper limb. Distal upper limb dysfunction is frequently referred to as impaired manual dexterity or hand dysfunction. According to Kister et al.,[2] impaired sensory function (85%), fatigue (81%), impaired hand function (60%), and mobility (50%) were the most frequently reported

symptoms in the first year of the disease. Recently, Bertoni et al.[3] reported that 75% of their study population (*n* = 110, median Expanded Disability Status Scale (EDSS) 6.5) had bilaterally (minimally) impaired manual dexterity as measured with the Nine-Hole Peg Test (NHPT).

An overview and description of upper limb outcome measures according to body function and structures as well as activity levels of the International Classification of Functioning (ICF) is provided by Lamers and Feys[4] Capacity measures assess the person's maximal ability in manual dexterity, gross motor function, or both, at a given moment in time, measured in a standardized environment.[5] Patient-reported outcome measures

address upper limb use and perceived difficulty of performing ADLs requiring one or both arms. A review on upper limb measures applied in MS rehabilitation context documented that the NHPT was by far the most frequent measure, utilized in 63% of published studies.[5] As such, the NHPT is widely considered a gold standard metric for manual dexterity. Besides the NHPT, other manual dexterity assessment tools such as the Purdue Pegboard test,[6] the Box and Block test,[7] and Coin Rotation Test[8] are less frequently used, and only limited studies have addressed their psychometric properties in MS patients.[5]

The NHPT was originally introduced by Kellor et al.[9] in 1971 as a measure of dexterity, in an official publication of the American Society for Occupational Therapy. The report provided approximate dimensions of the material and general procedures of administration. In 1985, Mathiowetz et al.[10] provided detailed test instructions and adult normative values according to hand, sex, and age. As well, reliability and correlation with the Purdue Pegboard was reported in healthy subjects. The NHPT was introduced to the MS community by Goodkin et al.[11] in 1988. In 1997, the National MS Society's Clinical Outcomes Assessment Task Force recommended the use of the NHPT as an upper limb outcome measure in MS.[12,13] In 1999, Cutter et al.[14] published the newly developed Multiple Sclerosis Function Composite (MSFC) including the NHPT in combination with the Paced Auditory Serial Addition test and Timed 25-Foot Walk as a potential clinical trial outcome measure in MS patients. Since then, the NHPT has been frequently included in MS research and clinical practice.

Review objective
The NHPT was selected for this review based on the widespread adoption and extensive data available for this capacity measure in MS research. The literature cited derives from a systematic review conducted through the MS Outcome Assessments Consortium (MSOAC). MSOAC's mission to develop a clinical outcome assessment tool for clinical trials to better capture MS-related disability was born out of a consensus paper by the International Advisory Committee on Clinical Trials in Multiple Sclerosis.[15] MSOAC includes representatives from the National MS Society as well as 6 other MS advocacy organizations, Food and Drug Administration (FDA), European Medicines Agency (EMA), National Institute of Neurological Disorders and Stroke (NINDS), 21 academic institutions, and 9 industry partners. The goals are acceptance and qualification by regulators of performance outcomes that reflect core MS impairments, and that

are highly reliable and valid, practical, cost-effective, and meaningful to MS patients. This review benefitted from a formal MSOAC-sponsored literature search, conducted in Embase, Medline, PsychInfo, and Cumulative Index of Nursing and Allied Health Literature, followed by an enrichment technique (key papers identified by MSOAC members added and informed search criteria) including work identified from prior reviews. Like the companion reviews of the Symbol Digit Modalities Test, Timed 25-Foot Walk, and Low Contrast Letter Acuity, we describe the NHPT in detail, and then cover its psychometric validity, concluding with an appraisal of the clinical meaningfulness of the measure.

Description of the test
The NHPT requires participants to repeatedly place and then remove nine pegs into nine holes, one at a time, as quickly as possible. Roughly, 53% of the variance in the NHPT score is explained by muscle strength, tactile sensitivity of the thumb, and presence of intention tremor.[16]

Over the years, the original instructions for the NHPT described by Kellor et al.[9] and Mathiowetz et al.[10] have been further refined with additional specifications. In 2001, the NHPT was included in a MSFC manual was released by the National MS Society's Clinical Outcomes Assessment Task Force. The manual, which can be downloaded from http://www.nationalmssociety.org, includes our recommended administration instructions, the placement of the NHPT apparatus compared to the tested hand, and frequently asked questions and answers documenting how to handle different situations a tester may face such as a peg ending up on the table or floor.

Different versions of the NHPT apparatus are available (Figure 1)[9,10] varying in the type of material, dimension of the surrounding surface and the shape of the container, but not the size of the holes or pegs, nor the distances between the holes. Importantly, as reported by Oxford et al.,[17] the performance times on the original wooden square version described by Mathiowetz et al.[10] and the plastic commercially available version (Smith & Nephew) as illustrated in Figure 1 are not significantly different. The latter plastic version seems the most frequently used and is recommended for maximal standardization.

The time needed to complete the NHPT in seconds is the most frequently reported metric in the literature. Recently, some studies have reported the number of pegs per second (pegs/s) instead of the seconds
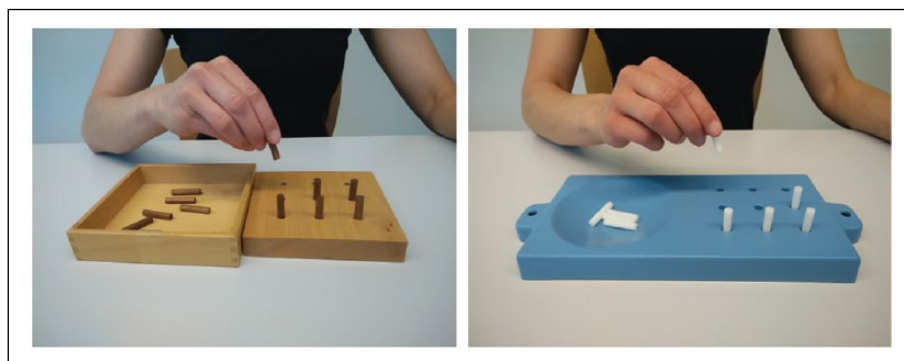
Figure 1. Original wooden square version (left) and the most commonly used commercially available plastic version (right) of the NHPT.

needed to complete the test.[3,16,18] The pegs/s is calculated based on nine pegs placed compared to the time needed to complete the test. In case the person is not able the complete the test, the pegs/s is calculated using the number of pegs placed compared to the time limit of 300 seconds. The calculation of pegs/s has the advantage of avoiding floor effects in people with severe upper limb dysfunction. Furthermore, it allows for a more normal distribution of the data, having advantages for statistical analysis. Four trials are conducted (two trials for each hand).

The MSFC manual recommends that the two trials for each hand be averaged, converted to the reciprocals of the mean times for each hand, and then the two reciprocals are averaged. The *Z*-score is obtained by subtracting the mean of the reference population from the test result and then dividing by the standard deviation (SD) of the reference population. While the latter calculation procedure provides a single score for manual dexterity, and its meaning is difficult to understand because it does not indicate potential asymmetry that may occur between hands, or which hand is most affected. Another argument for reporting results separately for each hand is the fact that normative data revealed significantly different performance scores, with the dominant hand task completed more quickly than the non-dominant hand.[10] Inconsistency in the reporting NHPT scores makes it difficult to compare results across studies, and we suggest that the average time to complete the task be reported for both the dominant and non-dominant hands separately.

Psychometric validity

Reliability
The reliability of the NHPT has been investigated in five studies including 255 MS patients, and results are

displayed in Table 1. The inter-rater and test–retest reliability of the NHPT are consistently high (range, $r = 0.86$–0.98), independent of whether repeated testing was performed within one session or on different test days. Reliability coefficients are also high across a wide range of disability levels, spanning normal to grossly abnormal hand function.

Reliability may be influenced by learning effects with repeated administrations of any neuroperformance test. In the case of the NHPT, one study, Solari et al.,[23] found significant practice effects and recommended that the NHPT be administered five times (four times pre-baseline) to attenuate these practice effects. However, the study did not report on reliability between first and last measures. One may question whether such practice is needed in routine clinical care, given our impression that the NHPT is frequently used in not only research, but also MS clinical centers such that most patients are regularly exposed to the test. Increasing test time with four pre-baseline tests for each hand may be mentally and physically fatiguing and too time-consuming in routine clinical practice, so only indicated for naïve subjects.

Discriminative validity
Normative data according to hand dominance, sex, and age have been provided for the original wooden version and commercial plastic NHPT devices.[10,17] Based on these, one can differentiate persons with normal hand function versus minimal hand dysfunction. Studies have shown that the NHPT discriminates manual dexterity in MS patients from healthy controls at a highly significant level ($p < 0.05$).[24,25]

Based on cross-sectional work, a number of research groups have recommended cut scores to identify levels of impairment on the NHPT (Table 2). Drake et al.[24]

Table 1. Reliability of the NHPT.

Study	*n*	Sample characteristics	Inter-rater reliability	Intra-rater reliability	Test–retest reliability
Bovend'Eerdt et al.[19]	26[a]	7 persons with MS Mean age: 51 years Mean disease duration: 6.5 years Ratio NHPT (s/peg) = 0.55			$r = 0.98$
Erasmus et al.[20]	189	Mean age: 40 years Median EDSS: 5.5 Median NHPT better hand (s) = 35.75 Median NHPT worse hand (s) = 52.80			Better hand: $\rho = 0.92$ Worse hand: $\rho = 0.86$
Rasova et al.[21]	17	Mean age: 43.3 years Mean EDSS: 3.5 Mean disease duration: 10.1 years Mean ± SD NHPT (s) = 24.93 ± 4.66			ICC = 0.88 (0.69–0.96)
Rosti-Otajärvi et al.[22]	10	Mean age: 43.1 years Mean EDSS: 2.5 Mean disease duration: 12.2 years Mean NHPT reported in figure	ICC = 0.98 (0.96–0.99)	ICC = 0.98 (0.95–0.99)	
Solari et al.[23]	32	Mean age: 43.1 years Mean EDSS: 4.5 Mean disease duration: 11.7 years Mean ± SD NHPT (s) = 27.90 ± 8.30	ICC = 0.93 (0.84–0.96)	ICC = 0.96–0.98 (0.91–0.99)	

NHPT: Nine-Hole Peg Test; MS: multiple sclerosis; SD: standard deviation; EDSS: Expanded Disability Status Scale; ICC: intra-class correlation coefficient (95% confidence interval); ρ: Spearman correlation coefficient; *r*: Pearson correlation coefficient.
[a]Not only people with MS were included.

recommended 18 seconds as an appropriate threshold to discriminate normal versus abnormal hand function. Johansson et al.[26] considered the NHPT score in MS patients abnormal when the NHPT score was higher than the age- and sex-related normative value + 1 SD. Kierkegaard et al.[18] suggested a flat cut-off value of 0.5 pegs/s corresponding to 18 seconds performed with the right hand as an indicator of normal functioning which could be used to identify people with MS at risk for activity limitations and participation restrictions. This cut-off value was calculated on the basis of the statistical risk for dependency on the activity and participation level as indicated on the Frenchay Activities Index,[27] and the Katz[28] personal and instrumental ADL. One must note, however, that these outcome measures are also influenced by a person's walking ability, the ability to perform a transfer, and cognitive function. Bertoni et al.[3] applied a more conservative cut-off score of >1.95 SD of the norms, in order to avoid misclassifying hands as showing reduced manual dexterity. Lamers et al.[16] recommended cut-off

score for differentiating persons with mild versus marked to severe upper limb dysfunction being 0.27 pegs/s corresponding to 33.3 seconds. This cut-off value was arbitrary defined based on the median NHPT score of a large MS sample (*n* = 105). Almost all persons with MS with a score below 33.3 seconds performed maximally on the Action Research Arm Test (ARAT), which includes lifting and manipulating objects using different hand grips as well as gross movements. Persons with MS with NHPT scores above 33.3 seconds did not reach maximal scores on the ARAT indicating more severe global hand and upper limb dysfunction. Further validation of the arbitrary cut-off score by comparison to external anchors is needed (Table 2).

Criterion and ecological validity
Many studies investigating validity of upper limb function tests in MS have selected the NHPT as the

Table 2. Overview cut-off values.

Cut-off value	Goal
Age- and sex-related normative value + 1 SD[26]	To differentiate between normal versus abnormal hand function
>1.95 SD of the age- and sex-related normative values[3]	To differentiate between normal versus abnormal hand function
18 seconds[24]	To differentiate between normal versus abnormal hand function
0.5 pegs/s or corresponding 18 seconds (right hand)[18]	Indicator of normal functioning which could be used to identify MS patient at risk for activity limitation and participation restrictions
0.27 pegs/s or corresponding 33.3 seconds[16]	To differentiate between mild and marked to severe hand dysfunction
SD: standard deviation.	

gold standard or reference value. As shown in Table 3, correlation coefficients between the NHPT and other capacity measures for hand and total upper limb function such as the Purdue Pegboard test,[30] Box and Block test,[11,30] and the ARAT[16] exceeded 0.40 indicating moderate associations. Importantly, the highest correlation coefficients were found with the modified Jebsen Taylor Hand function test[19,29] (range, $r = 0.86$–0.88) and the TEMPA[29] (range, $r = -0.79$ to -0.90 and 0.81–0.90) both of which assess the ability to perform functional ADL-like tasks requiring fine and gross manual dexterity during unilateral and bimanual tasks. These findings support not only the criterion validity of the NHPT, but also its ecological validity, as the score is robustly correlated with the manual manipulation of everyday use objects ranging from coins, playing cards, bins, pitchers, and glasses, all requiring hand and pinch grips and other upper limb movements.

Studies investigating the relationship between NHPT and patient-reported outcomes reveal a range of correlation coefficients (range, $r = -0.37$ to -0.79).[16,25,30,31] The correlation was the lowest with the ABILHAND[30] ($r = -0.37$), which may be related to the fact that this patient-reported outcome includes only bimanual activities, while the NHPT is performed unilaterally. Correlations with the manual ability measure (MAM-36)[16] and motor activity log (MAL)[25] were higher (range, $r = -0.56$ to -0.79 and 0.66, respectively), probably because the MAM-36 includes unilateral items and the MAL includes questions about unilateral use of the upper limb. In fact, the NHPT explained 44% of the variance in the MAM-36, which evaluates the perceived ease or difficulty that MS patients may have experienced when performing 36 common ADL tasks.[16] The NHPT of the most impaired upper limb explained 55% of the variance in the MAL, reflecting the quantity and quality of upper limb use during ADL tasks.[25] These findings confirm the high

ecological validity of the NHPT from a patient's perspective.

The relationship between the NHPT and actual upper limb performance measured with accelerometers was investigated in one study.[25] This relationship may provide us information on whether the NHPT score reflects the actual upper limb performance of a MS patient in daily life. Lamers et al.[25] reported that the correlation coefficient between the two measures was high ($r = -0.66$) for the non-dominant hand but non-significant for the dominant hand ($r = 0.04$). NHPT was not a retained variable in the regression models predicting the variance in the accelerometers. Future research is needed to replicate these findings and to improve the understanding of how upper extremity accelerometers reflect daily function, and consensus on which extracted parameters are most relevant.

Finally, it is noteworthy that the NHPT showed high correlations (range = 0.71–0.81) with the activity and ADL-dependency measures Frenchay Activities index and the Katz ADL indexes[18] in spite of the fact that both are heavily weighted toward mobility.

Responsiveness and clinical relevance of the NHPT

As evident in Table 4, two studies calculated areas under the receiver operating characteristic curves (AUC) in comparison with anchors from clinician and patients perspectives. Overall, the NHPT can be considered as a responsive measure detecting progression with AUC ranging from 0.49 to 0.97. Remarkably, the reported AUC values of the NHPT are greater when relating to changes in activity and ADL measures[18] (Frenchay Activities Index[27] and Katz ADL index[28]) as external anchors compared to Global Rating Scales (GRSs)[33] for disability and the EDSS.[34,35]

Table 3. Correlation coefficients between the NHPT and other outcome measures on activity level.

Study	*N*	Sample characteristics	Measures	Correlation coefficients
Bovend'Eerdt et al.[19]	26[b]	7 persons with MS Mean age: 51 years Mean disease duration: 6.5 years Mean ± SD NHPT (s/peg) = 0.55 ± 0.33	Jebsen Taylor hand function test (modified)	$r = 0.86^*$–$0.88^{*[19]}$
Feys et al.[29]	43	Mean age: 46 years	Jebsen Taylor hand function test (modified)	$\rho = 0.83^a$–$0.95^{a[29]}$
		Mean EDSS: 7	TEMPA—functional rating	$\rho = -0.79^a$ to $-0.90^{a[29]}$
		Mean disease duration: 15 years Mean ± SD left NHPT (s) = 77 ± 52 mean ± SD right NHPT (s) = 69 ± 53	TEMPA—speed	$\rho = 0.81^a$–$0.90^{a[29]}$
Goodkin et al.[11]	68	Mean age: 47.16 years Mean EDSS: 4.73 Mean disease duration 11.32 Mean ± SD NHPT (s) = 32.15	Box and block test	$r = -0.70^{a[11]}$
Lamers et al.[25]	30	Mean age: 58.2 years	Motor activity log	$\rho = -0.56^*$ to $-0.79^{*[25]}$
		Median EDSS: 7.5	Accelerometry of the upper limbs	$\rho = 0.04$ to $-0.66^{*[25]}$
		Mean disease duration: 21.8 years Median Dom NHPT (s) = 45.99 Median NDom NHPT (s) = 104.91		
Lamers et al.[16]	105	Mean age: 53.7 years	Action Research Arm Test	$r = 0.66^{*[16]}$
		Median EDSS: 6.5	Manual Ability Measure-36	$r = 0.66^{*[16]}$
		Mean disease duration: 17.9 years Median NHPT (pegs/s) = 0.25		
Simone et al.[30]	150[b]	17 persons with MS Mean age: 57 years	Box and block test	$r = -0.41^{*[30]}$
			Purdue pegboard test	$r = -0.52^{*[30]}$
			ABILHAND	$r = -0.37^{*[30]}$
		Mean ± SD NHPT (s) = 25 ± 11		
Marrie and Goldman[31]	44	Mean age: 42.2 years Median EDSS: 3.5 Mean disease duration: 8.3 years Mean ± SD NHPT (*Z*-score) = 0.18 ± SD	Performance Scales: hand	$\rho = -0.59^{*[31]}$

NHPT: Nine-Hole Peg Test; EDSS: Expanded Disability Status Scale; MS: multiple sclerosis; SD: standard deviation.
Values are the ranges of Spearman correlation coefficients (ρ) or Pearson correlation coefficients (r) found between outcome measures reported in different articles.
[a]Significance not reported.
[b]Not only people with MS were included.
$^*p < 0.05$.

Some studies addressed the magnitude of clinically meaningful change that can be used for the interpretation of changes in the NHPT. By applying distribution-based statistical methods, De Groot et al.[32] reported a smallest real change (SRC) at an individual/group level of 5.32/0.53 seconds (anchor EDSS) and 2.82/0.55 seconds (anchor GRS).

Previous studies[36–40] have reported that 15%–20% change in the NHPT corresponds to predefined clinical meaningful changes in the EDSS, Guys Neurological Rating Scale (GNRS), MS-Impact Scale (MSIS), and global disability ratings. This change criterion, with a varying absolute change value of the NHPT corresponding to 15%–20% of the

Table 4. Responsiveness values of the NHPT.

Study	N	Sample characteristics	Distribution-based method	Anchor-based method	
				Clinician's perspective	Patient perspective
De Groot et al.[32]	156	Mean age: 37.6 years Mean EDSS: 2.5 Mean disease duration 0.26 years Mean ± SD NHPT = 21.1 ± 4.0	*Anchors: EDSS, GRS* SRC_{ind} = 5.32 EDSS SRC_{gr} = 0.53 SRC_{ind} = 2.82 GRS SRC_{gr} = 0.55	*Anchor: EDSS* AUC = 0.67 (0.58–0.76) –	*Anchor: GRS* AUC = 0.59 (0.49–0.69) –
Kierkegaard et al.[18]	164	Mean age: 51 years EDSS: between 1 and 9 Mean NHPT score was not reported		*Anchors: Katz P-ADL and I-ADL* AUC = 0.94 (0.91–0.97) Cut-off value = 0.35 pegs/s AUC = 0.87 (0.81–0.92) Cut-off value = 0.52 pegs/s	*Anchor: FAI* AUC = 0.85 (0.79–0.91) Cut-off value = 0.48 pegs/s

NHPT: Nine-Hole Peg Test; SD: standard deviation; EDSS: Expanded Disability Status Scale; GRS: Global Rating Scale; Katz P-ADL: Katz Personal ADL; Katz I-ADL: Katz Instrumented ADL; FAI: Frenchay Activity Index; AUC: area under the receiver operating characteristic curve (values: 95% confidence interval); MIC: minimally important change; SRC_{ind}: smallest real change at individual level; SRC_{gr}: smallest real change at group level.

baseline value, appears robust in multiple longitudinal studies differentiating MS patients that improved or progressed versus patients remaining stable. One may question to what extent the NHPT score contributes to detection of overall progression in comparison to other related parameters, such as gait function. In this framework, the study of Kragt et al.[38] showed that the NHPT explained more variance than the T25FW on domains of sexual dysfunction, upper limb disability, fatigue, and mood measured by the GNRS. Similar results were found in the International MS Secondary Progressive Avonex Controlled Trial (IMPACT) trial reported by Cohen et al.[41] They found that the MSFC worsening was reduced by 40.4% after 2 years of interferon β-1a therapy compared to placebo in people with secondary progressive MS with the NHPT being the only MSFC metric showing a statistically significant benefit.

The above-mentioned studies on responsiveness and change were all based on longitudinal data over 2 years or more, and focused predominantly on detection of progression. Only a few studies reported on the sensitivity of the NHPT to detect response to treatment over a shorter period. Pascual et al.[42] showed that the NHPT was the most sensitive MSFC test to the effects of high-dose oral methylprednisolone. Savin et al.[43] reported that the NHPT and ABILHAND showed very similar sensitivity to improvements after Fampyra-PR, both in terms of the number of responders (33%) and the magnitude of response (reported as 12%–16%, slightly lower that the 15%–20% difference thought to be clinically meaningful).

These studies show that the NHPT is responsive, relates to real-life anchors, and can be used to form responder definitions based on a clinically meaningful change. Further work on the latter point is needed, with application of similar responder definitions of other upper extremity function metrics, as was done before in the area of ambulation.[44]

Conclusion and future directions
The NHPT is recommended as a gold standard test for measuring manual dexterity in MS patients, and it can be used as reference value to investigate validity of other, newly developed upper limb outcome measures. This review illustrates the clinical utility of the NHPT and its excellent psychometric properties regarding reliability, discriminant, concurrent, and ecological validity. The NHPT detects progression over time, and it is sensitive to treatment. As such, it is recommended for inclusion in clinical trials. The NHPT is correlated with other disability measures and patient-reported outcomes focusing on upper limb function.

From a practical perspective, the NHPT is easy to administer and acceptable to patients. To further facilitate electronic recording and potentially gain more insight concerning the quality of hand performance during the test instead of only a time score, current developments in technology-supported upper limb assessment tools may in the future supplement this clinical measure when the focus is on hand dysfunction. To date, promising results have been published

for the MS performance test (MSPT)[45] and the virtual peg insertion test (VPIT).[46,47] The MSPT test consists of a computer-assisted assessment of tasks analogous to the T25FW, NHPT, Symbol Digit Modalities Test (SDMT), and a measure of visual acuity, using a tablet and some technical additions such as audio-visual instructions and a recording device. The time scores on the instrumented NHPT are highly reproducible in MS patients ($r > 0.9$) and discriminate between MS patients and controls as well as the current technician-based version. The VPIT is performed with a small haptic robot and mimics the real dimensions and resistance of the NHPT device. It provides kinematic and dynamic parameters related to the movement trajectories, the interaction force with the virtual environment, and grasping force.[46,47] A high correlation ($r = 0.66$) was observed between the NHPT and the VPIT. Further validation work is needed to understand to which extent automated, computer-assisted test can better reflect upper limb function in future.

In whatever format, the NHPT is the optimal metric for measuring the impact of MS on upper extremity function. The majority of research suggests that a 20% change in test score is commonly used to define clinically meaningful worsening; evidence is needed to support this definition in other versions of the NHPT and at all stages of the disease.

Acknowledgements

Declaration of Conflicting Interests

Funding

References

1. Yozbatiran N, Baskurt F, Baskurt Z, et al. Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. *J Neurol Sci* 2006; 246: 117–122.

2. Kister I, Bacon TE, Chamot E, et al. Natural history of multiple sclerosis symptoms. *Int J MS Care* 2013; 15: 146–158.

3. Bertoni R, Lamers I, Chen CC, et al. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with multiple sclerosis. *Mult Scler* 2015; 21: 1566–1574.

4. Lamers I and Feys P. Assessing upper limb function in multiple sclerosis. *Mult Scler* 2014; 20: 775–784.

5. Lamers I, Kelchtermans S, Baert I, et al. Upper limb assessment in multiple sclerosis: A systematic review of outcome measures and their psychometric properties. *Arch Phys Med Rehabil* 2014; 95: 1184–1200.

6. Tiffin J and Asher EJ. The Purdue pegboard; Norms and studies of reliability and validity. *J Appl Psychol* 1948; 32: 234–247.

7. Mathiowetz V, Volland G, Kashman N, et al. Adult norms for the Box and Block Test of manual dexterity. *Am J Occup Ther* 1985; 39: 386–391.

8. Mendoza JE, Apostolos GT, Humphreys JD, et al. Coin rotation task (CRT): A new test of motor dexterity. *Arch Clin Neuropsychol* 2009; 24: 287–292.

9. Kellor M, Frost J, Silberberg N, et al. Hand strength and dexterity. *Am J Occup Ther* 1971; 25: 77–83.

10. Mathiowetz V, Weber K, Kashman N, et al. Adult norms for the Nine Hole Peg Test of finger dexterity. *OTJR: Occup Particip Health* 1985; 5: 24–38.

11. Goodkin DE, Hertsgaard D and Seminary J. Upper extremity function in multiple sclerosis: Improving assessment sensitivity with box-and-block and Nine-Hole Peg Tests. *Arch Phys Med Rehabil* 1988; 69: 850–854.

12. Rudick R, Antel J, Confavreux C, et al. Recommendations from the national multiple sclerosis society clinical outcomes assessment task force. *Ann Neurol* 1997; 42: 379–382.

13. Rudick R, Antel J, Confavreux C, et al. Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996; 40: 469–479.

14. Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 1999; 122(Pt 5): 871–882.

15. Cohen JA, Reingold SC, Polman CH, et al.; International Advisory Committee on Clinical Trials in Multiple Sclerosis. Disability outcome measures in multiple sclerosis clinical trials: Current status and future prospects. *Lancet Neurol* 2012; 11: 467–476.

16. Lamers I, Cattaneo D, Chen CC, et al. Associations of upper limb disability measures on different levels of the International Classification of Functioning, Disability and Health in people with multiple sclerosis. *Phys Ther* 2015; 95: 65–75.

17. Oxford GK, Vogel KA, Le V, et al. Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. *Am J Occup Ther* 2003; 57: 570–573.

18. Kierkegaard M, Einarsson U, Gottberg K, et al. The relationship between walking, manual dexterity, cognition and activity/participation in persons with multiple sclerosis. *Mult Scler* 2012; 18: 639–646.

19. Bovend'Eerdt T, Dawes H, Johansen-Berg H, et al. Evaluation of the modified Jebsen Test of Hand Function and the University of Maryland Arm Questionnaire for Stroke. *Clin Rehabil* 2004; 18: 195–202.

20. Erasmus LP, Sarno S, Albrecht H, et al. Measurement of ataxic symptoms with a graphic tablet: Standard values in controls and validity in multiple sclerosis patients. *J Neurosci Methods* 2001; 108: 25–37.

21. Rasova K, Martinkova P, Cattaneo D, et al. Physical therapy in multiple sclerosis differs across Europe: Information regarding an ongoing study. *J Int Med Res* 2014; 42: 1185–1187.

22. Rosti-Otajärvi E, Hamalainen P, Koivisto K, et al. The reliability of the MSFC and its components. *Acta Neurol Scand* 2008; 117: 421–427.

23. Solari A, Radice D, Manneschi L, et al. The multiple sclerosis functional composite: Different practice effects in the three test components. *J Neurol Sci* 2005; 228: 71–74.

24. Drake AS, Weinstock-Guttman B, Morrow SA, et al. Psychometrics and normative data for the multiple sclerosis functional composite: Replacing the PASAT with the Symbol Digit Modalities Test. *Mult Scler* 2010; 16: 228–237.

25. Lamers I, Kerkhofs L, Raats J, et al. Perceived and actual arm performance in multiple sclerosis: Relationship with clinical tests according to hand dominance. *Mult Scler* 2013; 19: 1341–1348.

26. Johansson S, Ytterberg C, Claesson IM, et al. High concurrent presence of disability in multiple sclerosis. Associations with perceived health. *J Neurol* 2007; 254: 767–773.

27. Holbrook M and Skilbeck CE. An activities index for use with stroke patients. *Age Ageing* 1983; 12: 166–170.

28. Katz S. Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc* 1983; 31: 721–727.

29. Feys P, Duportail M, Kos D, et al. Validity of the TEMPA for the measurement of upper limb function in multiple sclerosis. *Clin Rehabil* 2002; 16: 166–173.

30. Simone A, Rota V, Tesio L, et al. Generic ABILHAND questionnaire can measure manual ability across a variety of motor impairments. *Int J Rehabil Res* 2011; 34: 131–140.

31. Marrie RA and Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. *Mult Scler* 2007; 13: 1176–1182.

32. De Groot V, Beckerman H, Uitdehaag BM, et al. The usefulness of evaluative outcome measures in patients with multiple sclerosis. *Brain* 2006; 129: 2648–2659.

33. Jaeschke R, Singer J and Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.

34. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.

35. De Groot V, Beckerman H, Uitdehaag BM, et al. Physical and cognitive functioning after 3 years can be predicted using information from the diagnostic process in recently diagnosed multiple sclerosis. *Arch Phys Med Rehabil* 2009; 90: 1478–1488.

36. Bosma LV, Kragt JJ, Brieva L, et al. Progression on the multiple sclerosis functional composite in multiple sclerosis: What is the optimal cut-off for the three components? *Mult Scler* 2010; 16: 862–867.

37. Van Winsen LM, Kragt JJ, Hoogervorst EL, et al. Outcome measurement in multiple sclerosis: Detection of clinically relevant improvement. *Mult Scler* 2010; 16: 604–610.

38. Kragt JJ, van der Linden FA, Nielsen JM, et al. Clinical impact of 20% worsening on Timed 25-Foot Walk and 9-Hole Peg Test in multiple sclerosis. *Mult Scler* 2006; 12: 594–598.

39. Schwid SR, Goodman AD, McDermott MP, et al. Quantitative functional measures in MS: What is a reliable change? *Neurology* 2002; 58: 1294–1296.

40. Hoogervorst EL, Kalkers NF, Cutter GR, et al. The patient's perception of a (reliable) change in the multiple sclerosis functional composite. *Mult Scler* 2004; 10: 55–60.

41. Cohen JA, Cutter GR, Fischer JS, et al. Benefit of interferon beta-1a on MSFC progression in secondary progressive MS. *Neurology* 2002; 59: 679–687.

42. Pascual AM, Bosca I, Coret F, et al. Evaluation of response of multiple sclerosis (MS) relapse to oral high-dose methylprednisolone: Usefulness of MS functional composite and Expanded Disability Status Scale. *Eur J Neurol* 2008; 15: 284–288.

43. Savin Z, Lejbkowicz I, Glass-Marmor L, et al. Effect of fampridine-PR (prolonged released 4-aminopyridine) on the manual functions of patients with multiple sclerosis. *J Neurol Sci* 2016; 360: 102–109.

44. Baert I, Freeman J, Smedal T, et al. Responsiveness and clinically meaningful improvement, according to disability level, of five walking measures after rehabilitation in multiple sclerosis: A European multicenter study. *Neurorehabil Neural Repair* 2014; 28: 621–631.

45. 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.

46. Lambercy O, Fluet MC, Lamers I, et al. Assessment of upper limb motor function in patients with multiple sclerosis using the Virtual Peg Insertion Test: A pilot study. *IEEE Int Conf Rehabil Robot* 2013; 2013: 6650494.

47. Tobler-Ammann BC, de Bruin ED, Fluet MC, et al. Concurrent validity and test-retest reliability of the Virtual Peg Insertion Test to quantify upper limb function in patients with chronic stroke. *J Neuroeng Rehabil* 2016; 13: 8.