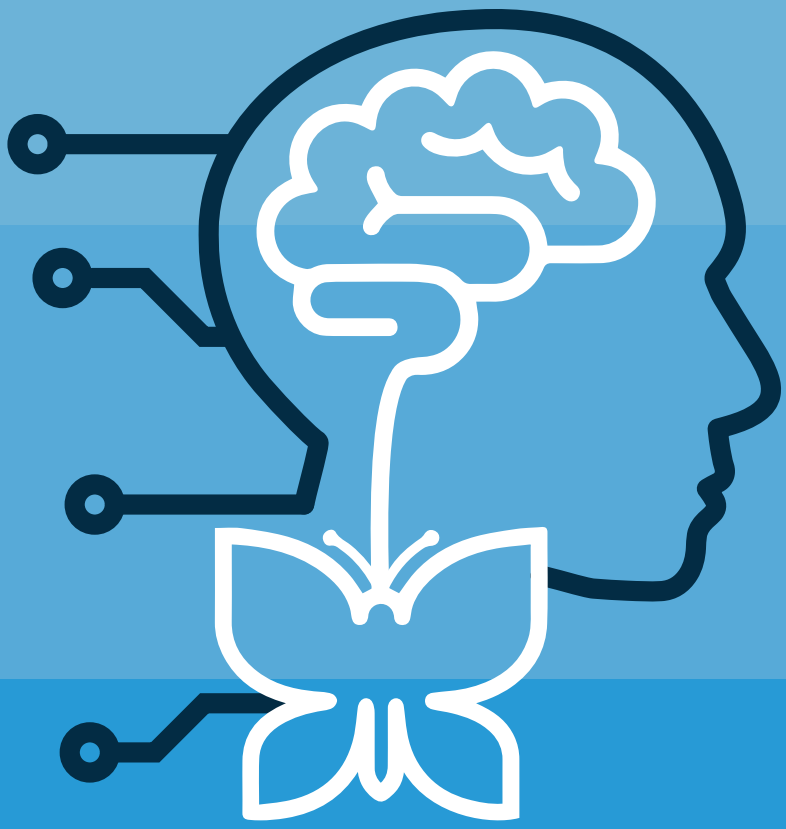
Disability in multiple sclerosis

van Munster, Caspar Erik Pieter

2021

document version
Publisher's PDF, also known as Version of record



Disability in multiple sclerosis

Improving clinical assessment

Caspar E.P. van Munster

Disability in multiple sclerosis

Improving clinical assessment

Caspar E.P. van Munster

Financial support for the printing of this thesis was kindly provided by Stichting MS Research, Roche Nederland B.V., Teva Nederland, Mylan B.V. and Bayer B.V.

ISBN: 978-94-6416-285-1

Elektronische ISBN: 978-94-6416-286-8

Cover and lay-out by: Publiss | www.publiss.nl

Printed by: Ridderprint | www.ridderprint.nl

VRIJE UNIVERSITEIT

Disability in multiple sclerosis

Improving clinical assessment

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 11 februari 2021 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Caspar Erik Pieter van Munster

geboren te Nijmegen

Promotor: Prof. dr. B.M.J. Uitdehaag

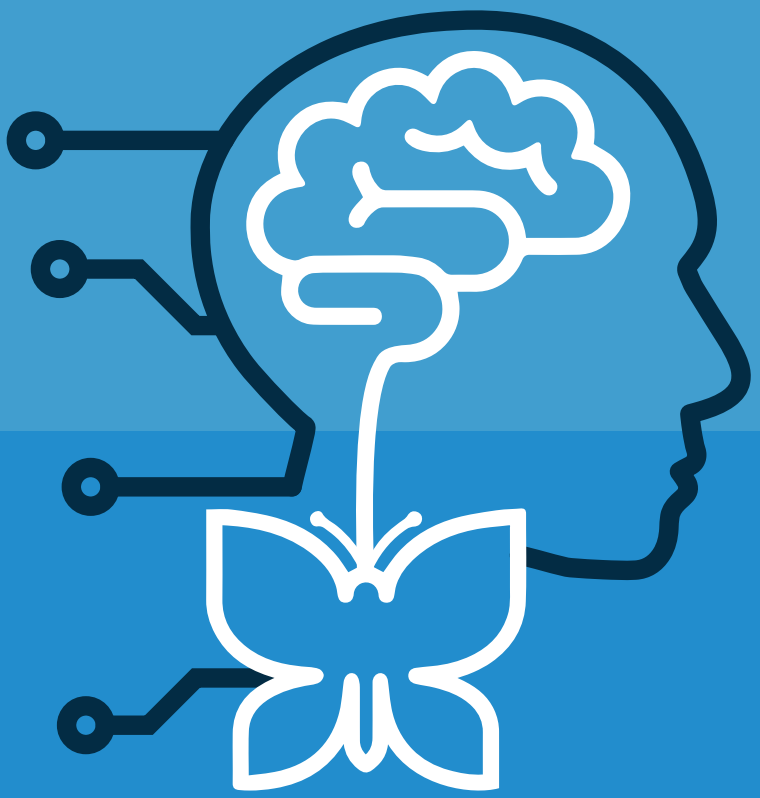
Co-promotor: Prof. dr. J. Killestein

Ter herinnering aan Pépé

CONTENTS

PART 1

Introduction



CHAPTER 1

Introducing multiple sclerosis, aims and outline of the thesis

Background of multiple sclerosis

Multiple Sclerosis (MS) is a devastating neurological disease that strikes young people in the prime of their life. Over 2 million people worldwide are affected.(1) It is estimated that approximately 17,000 people are affected in the Netherlands, and this figure may be an underestimate.(2) The disease most commonly impacts individuals between the ages of 20 and 40 year, and there is a female preponderance.(3) Despite its relative low prevalence, MS has a great impact on a social and economic level, since it strikes at young working age, treatment costs are high and rise with increasing disability.(4)

The exact aetiology is unknown, although it involves interactions between environmental, genetic and epigenetic factors, all of which contribute to the development of MS.(5) The pathophysiology of MS is characterized by the development of focal inflammatory lesions in the central nervous system (i.e. brain and spinal cord) that leads to neuronal demyelination and axonal damage.(6) A central hallmark of the disease is that the neuronal inflammation occurs with dissemination in space and time. This essentially means that inflammation develops in different locations of the central nervous system, and at different moments in time, respectively. In addition, neurodegenerative damage appears as a consequence of the accumulated inflammatory lesion burden, as well as an independent pathophysiological process.(7) Inflammatory and neurodegenerative pathology are found in the white and grey matter as well.(6)

Clinical course and diagnosis

The clinical course of MS is highly heterogeneous. In accordance with dissemination in space and time that characterizes pathophysiology, clinical symptomatology presents as episodes of signs in different neurological systems. Typically, symptoms develop within days to weeks, before reaching a certain peak of severity, and subsequently resolves during weeks with or without residual deficits. Such an episode of symptoms is referred to as a clinical relapse (referred to as "*schub"* in Dutch). The number and severity of the episode, extent of the recovery, and which neurological system is affected differs widely between and within individuals with MS.

The clinical course of MS can be categorized into different phenotypes (schematic representation in figure 1).(8) The first clinical relapse, which is suggestive of MS but does not (yet) fulfil the diagnostic criteria, is defined as a clinically isolated syndrome (CIS). If a second relapse occurs, involving another neurological system, the clinical course is regarded as relapsing-remitting MS (RRMS). In approximately 30 – 60% of patients, the clinical course eventually shifts to a more slowly progressive symptomatology. This phase is referred to as secondary-progressive MS (SPMS). During this phase with predominantly progressive symptoms, superimposed relapses may

occur. A minority of patients (approximately 15%) presents with progressive symptoms since the onset of the disease. This phenotype is referred to as primary-progressive MS (PPMS). Patients with this phenotype are generally older than RRMS patients, and this phenotype is more common in males.

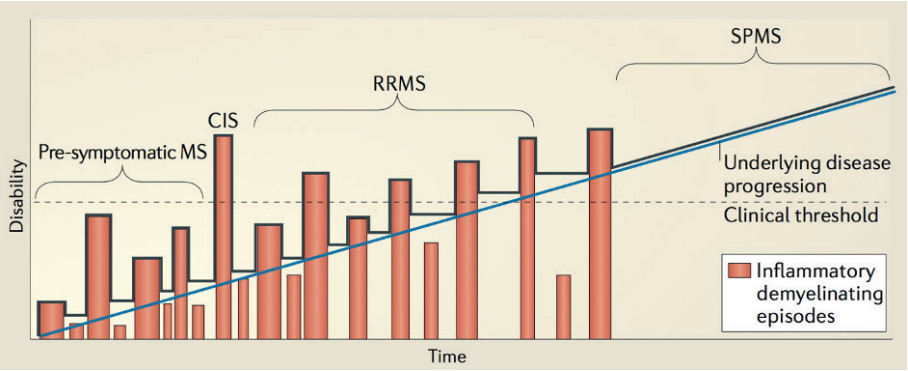


Figure 1, clinical phenotypes Multiple Sclerosis (reproduced from Stys et al. with permission from Springer Nature.(7)

The formal diagnosis of MS is based on a combination of clinical symptoms, coupled with findings from ancillary investigations, of which magnetic resonance imaging (MRI) and cerebrospinal fluid (CSF) are most relevant. Demyelinated lesions can be typically visualized using MRI. These lesions commonly occur in specific locations in the brain (i.e. periventricular, (juxta) cortical and infratentorial) and spinal cord. Supportive CSF findings include CSF-specific oligoclonal bands, an increased IgG index and no more than $50x10^6$ leucocytes per litre. Clinical symptoms, MRI and CSF findings are incorporated into the McDonald criteria that aids in the efforts to diagnose MS.(9) These criteria are summarized in figure 2, which also illustrates the importance of demonstrating dissemination in space and time in diagnosing MS.(10) The 2017 revision of the criteria allows an earlier diagnosis while preserving good specificity. Importantly, the McDonald criteria are principally *categorizing* criteria that primarily facilitates the phenotyping of the clinical course, rather than being true *diagnostic* criteria. This implies that the criteria must be used for the right patient in the right context; specifically a clinical syndrome suspect for MS. As such, other diseases with similar symptoms must therefore be ruled out sufficiently. A recent study illustrates that the clinical and radiological aspects of the criteria can be misunderstood and misapplied, which underscores the importance of educational efforts in the MS field.(11)

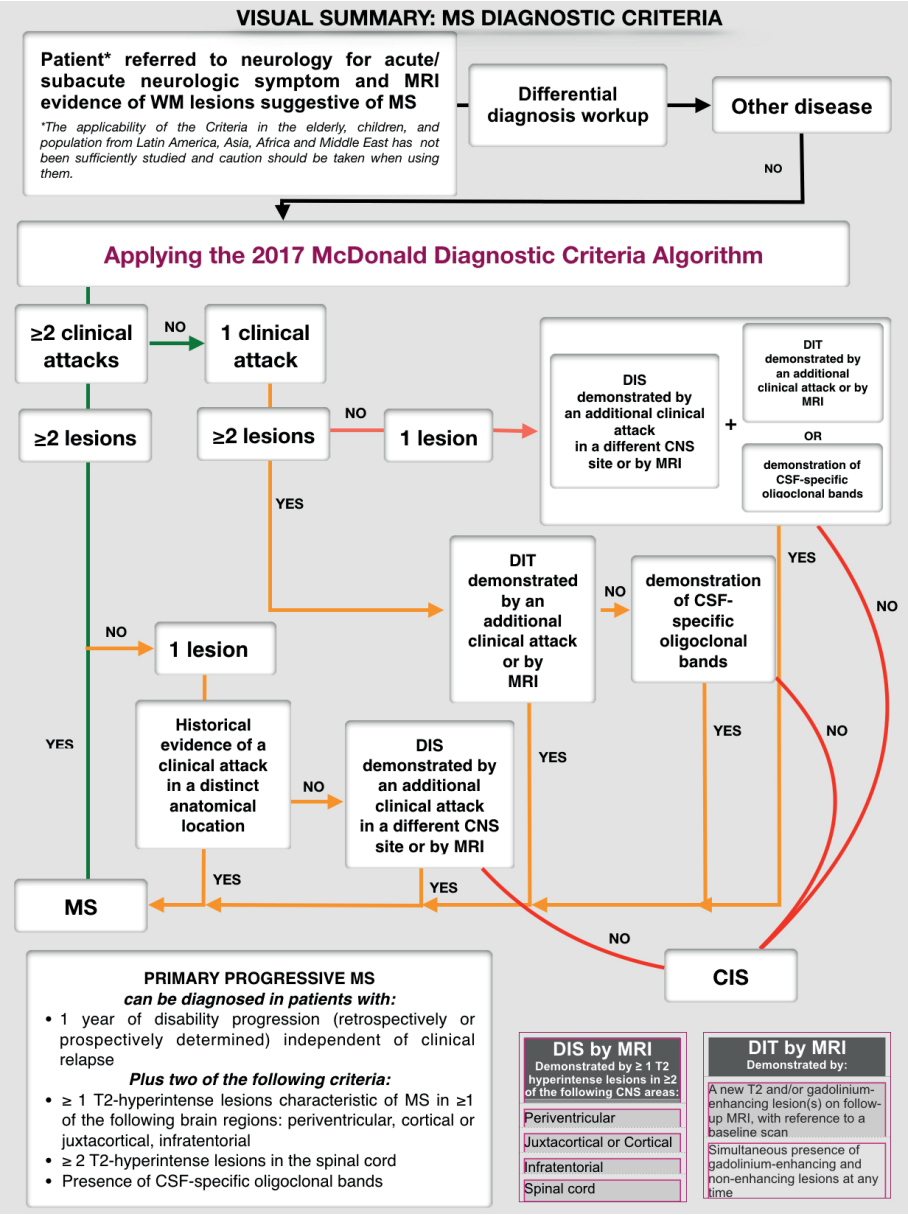


Figure 2, 2017 revised McDonald criteria (reproduced from de Angelis et al. with permission from BMJ Publishing Group Ltd.(10)

Clinical assessment in multiple sclerosis

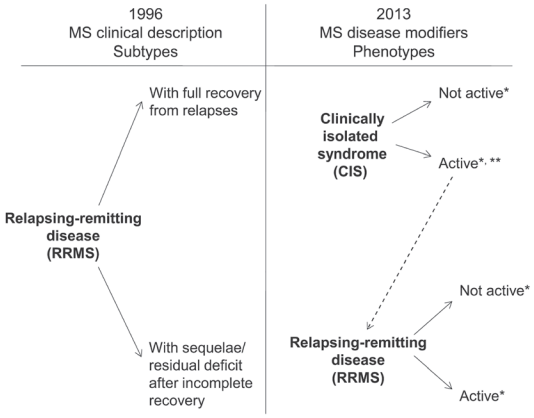
One of the biggest challenges facing MS care and research is the assessment of its clinical presentation. Before we point out some of these challenges and set the stage for this thesis, we will elaborate on the relevance of good clinical assessment.

Relevance

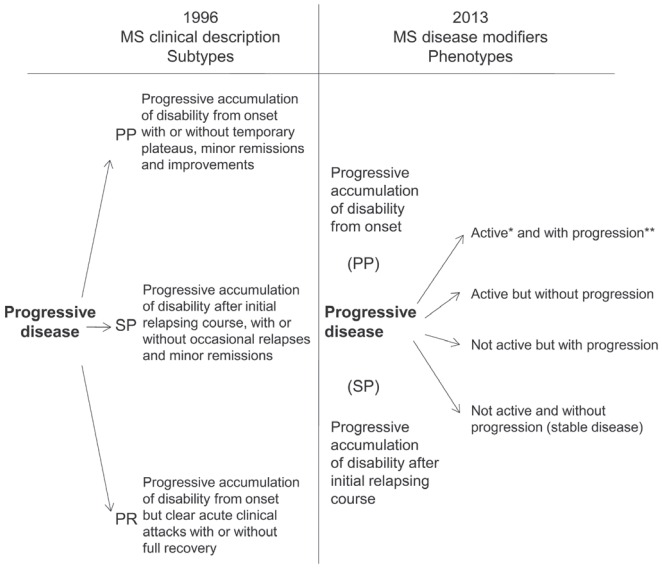
Good clinical assessment is relevant throughout all stages of patient care and research. In the diagnostic process, making the diagnosis of MS begins with good clinical assessment. This means that MS can only be diagnosed with a clinical syndrome suspect for MS. Ancillary investigations, such as MRI, are very valuable, but cannot be used in isolation to make the diagnosis. In fact, one could get into difficult situations when an MRI has been performed and typical MS lesion are found, while no symptoms are present (e.g. in the case of a "preventive" total body MRI) or symptoms are not specific for MS (such as headache or fatigue). A situation like this is referred to as a radiologically isolated syndrome (RIS). The exact therapeutic consequences of a RIS are largely unknown.(12)

The assessment of the clinical course of MS, irrespective of whether or not the patients is on disease modifying therapy (DMT), is based on two main pillars: disease activity and progression. Neuroinflammation and neurodegeneration respectively are the main underlying pathophysiological processes.(13) *Disease activity* is reflected clinically in the occurrence of relapses, and radiologically in the increase of lesion load or formation of gadolinium enhancing lesion. *Disease progression* is defined as a slow accrual of disability independent from clinical relapses. These pillars are used to categorize patients into different phenotypes, which are displayed in figure 3.(14) Adequate assessment of these pillars is crucial for effective treatment decision making.

Accurate and early diagnosis is increasingly important because of major advantages in terms of treatment possibilities. The development of DMT in MS in the past three decades is depicted in figure 4.(15) Treatment goals in the early 1990's were restricted to moderate suppression of disease activity and possibly delay of disability progression. Thanks to increasing DMT efficacy, present treatment goals are much more ambitious. Using the novel high efficacy DMT, the complete absence of disease activity and progression is achieved in 23 – 48% of patients.(16) This outcome is referred to as 'no evidence of disease activity' (NEDA). Notably, in order to achieve this result, it is important that the appropriate treatment is either initiated early on in the disease course, or treatment is switched swiftly to another (more effective) DMT when disease activity persists. Early and accurate diagnosis has also become more important for the progressive forms of MS (SPMS and PPMS), since DMTs have recently become available for these phenotypes.



*Activity determined by clinical relapses and/or MRI activity (contrast-enhancing lesions; new or unequivocally enlarging T2 lesions assessed at least annually); if assessments are not available, activity is "indeterminate." **CIS, if subsequently clinically active and fulfilling current multiple sclerosis (MS) diagnostic criteria, becomes relapsing-remitting MS (RRMS).



*Activity determined by clinical relapses assessed at least annually and/or MRI activity (contrast-enhancing lesions; new and unequivocally enlarging T2 lesions). **Progression measured by clinical evaluation, assessed at least annually. If assess-ments are not available, activity and progression are "indeterminate." MS 5 multiple sclerosis; PP 5 primary progressive; PR 5 progressive relapsing; SP 5 secondary progressive.

Figure 3, clinical phenotypes according to disease activity and progression, compared with conventional phenotypical classification (reproduced from Lublin et al. with permission from Wolter Kluwer Health, Inc.(14)

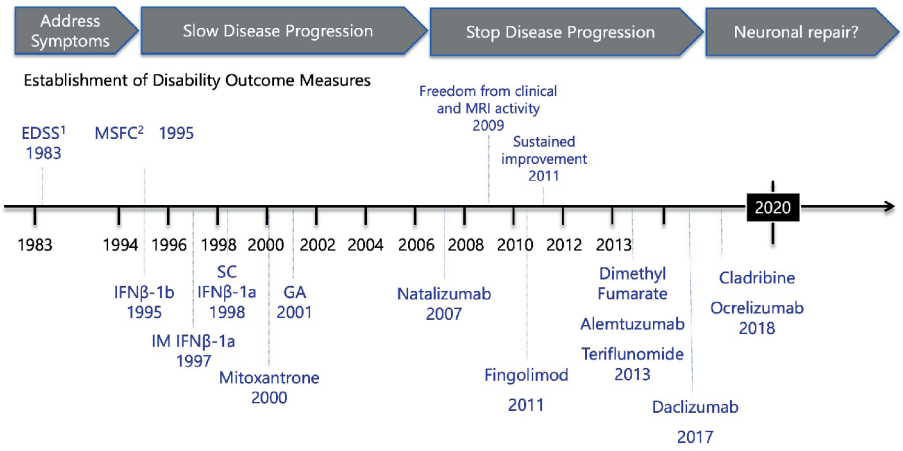


Figure 4, historical overview of disease modifying treatment (adapted from Thompson et al. with permission from Elsevier.(15)

Several other arguments are also relevant in emphasizing the importance of accurate clinical assessment. Firstly, clinical symptomatology does not necessarily correlate with radiological pathology (i.e. the clinic-radiological paradox).(17) This implies that not all disease activity is captured by MRI in isolation (and vice versa). Secondly, accrual of disability is currently the only meaningful measure to assess neurodegeneration in clinical practice. Although various other measures are used in research as surrogate markers for neurodegeneration (e.g. brain atrophy,(18) optical coherence tomography(19) and neurofilament light chain(20)), they are not yet sufficient to use in individual cases. Lastly, clinical assessment can be done quickly, regularly and cost-effectively (relative to performing an MRI for example).

Challenges

Several challenges exist in the clinical assessment of MS patients in daily practice and clinical trials. The most important challenge is the high variability of disease expression and course. Patients exhibit a wide symptomatologic variation in terms of relapse frequency and severity, as well as accrual of disability. Furthermore, relapses present with essentially all neurological symptoms in which there also is an age-dependent distribution.(21) The extent to which the clinical syndrome of a relapse contributes to disability is also variable. The heterogeneity of clinical presentation also hampers defining (uniform) outcome measures of disability for clinical trials.

Another challenge is the overall slow accumulation of disability. In clinical practice, this may cause difficulties to assess whether a patient is stable (on

DMT) or exhibits progressive symptoms. This is particularly important in order to establish an indication to start, or to evaluate the efficacy of treatment in the progressive phenotypes. In clinical trials, the slow accumulation dictates long-term follow-up to assess treatment effects.

Lastly, there are several confounding factors exist that influence the assessment of disability, which may not directly be related to disease activity. Examples include co-existing fatigue, mood disturbances, deconditioning, spasticity and side effects of medication. These factors will have to be accounted for to achieve an individualized treatment approach in clinical practice.

Aims and outline of this thesis

Good clinical assessment is essential in MS care and research, yet this remains challenging. As such, this thesis aims to improve the clinical assessment of MS patients. To create the framework for the studies presented here, an overview of clinical and paraclinical outcome measures is given in **chapter 2**.

Part II, describes the assessment of upper extremity functioning (UEF) and mobility. In **chapter 3**, UEF is assessed using several measures in subgroups of MS patients with different levels of ambulatory impairment. In **chapter 4**, various aspects of existing clinical measures for UEF and mobility are assessed, and their relative value is determined. **Chapter 5** sets out to determine a minimal clinically important difference of improvement of a patient-reported outcome measure for UEF. In **chapter 6**, the responsiveness of various measures for UEF and mobility are assessed in several subgroups of MS patients.

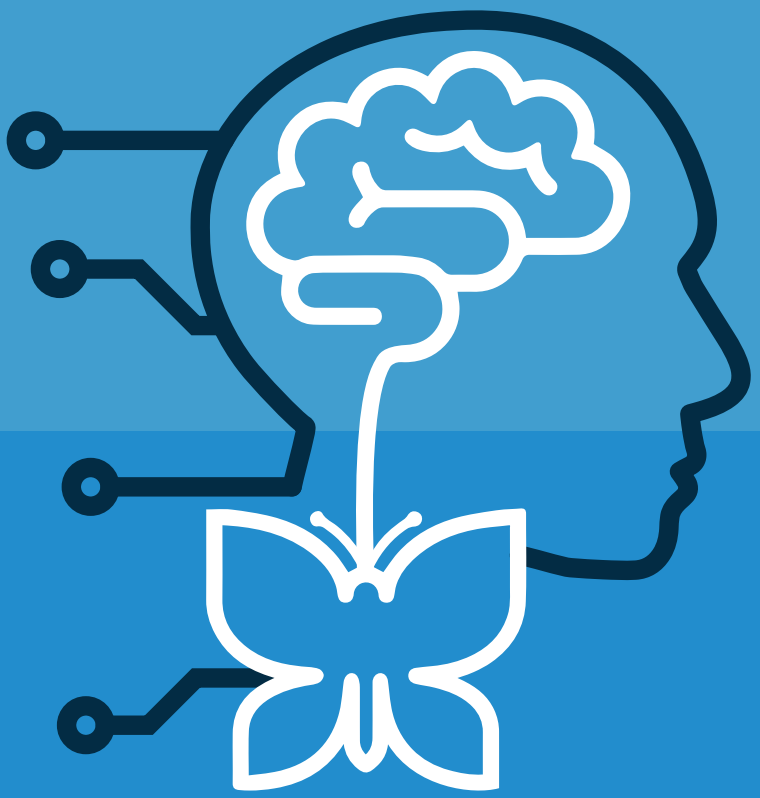
In **part III**, three studies are presented in which several aspects of video-assisted assessment of motor functioning are evaluated. **Chapter 7** presents the added value of reference videos in the assessment of motor functioning. In **chapter 8**, combinations of video-assisted assessment were compared with classical performance measures of UEF and mobility. In **chapter 9**, a proof of concept study is presented that describes the use of auto-encoders to ensure data privacy when sharing videos of patients performing movements.

Part IV – chapter 10, summarises the conclusions of this thesis and discusses their implications for clinical practice and future studies.

REFERENCES

1. Group GBDNDC. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Neurol. 2017;16(11):877-97.
2. https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/incidenties-en-prevalenties [
3. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. N Engl J Med. 2018;378(2):169-80.
4. Uitdehaag B, Kobelt G, Berg J, Capsa D, Dalen J, European Multiple Sclerosis P. New insights into the burden and costs of multiple sclerosis in Europe: Results for the Netherlands. Mult Scler. 2017;23(2_suppl):117-29.
5. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol. 2017;13(1):25-36.
6. Geurts JJ, Barkhof F. Grey matter pathology in multiple sclerosis. Lancet Neurol. 2008;7(9):841-51.
7. Stys PK, Zamponi GW, van Minnen J, Geurts JJ. Will the real multiple sclerosis please stand up? Nat Rev Neurosci. 2012;13(7):507-14.
8. Leemhuis EJ, Arnoldus EPJ, Visser LH. [Prognosis after diagnosis of multiple sclerosis]. Ned Tijdschr Geneeskd. 2017;161:D943.
9. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162-73.
10. De Angelis F, Brownlee WJ, Chard DT, Trip SA. New MS diagnostic criteria in practice. Pract Neurol. 2019;19(1):64-7.
11. Solomon AJ, Pettigrew R, Naismith RT, Chahin S, Krieger S, Weinshenker B. Challenges in multiple sclerosis diagnosis: Misunderstanding and misapplication of the McDonald criteria. Mult Scler. 2020:1352458520910496.
12. De Stefano N, Giorgio A, Tintore M, Pia Amato M, Kappos L, Palace J, et al. Radiologically isolated syndrome or subclinical multiple sclerosis: MAGNIMS consensus recommendations. Mult Scler. 2018;24(2):214-21.
13. Coetzee T, Thompson AJ. Unified understanding of MS course is required for drug development. Nat Rev Neurol. 2018;14(4):191-2.
14. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-86.
15. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet. 2018;391(10130):1622-36.
16. Giovannoni G. Disease-modifying treatments for early and advanced multiple sclerosis: a new treatment paradigm. Curr Opin Neurol. 2018;31(3):233-43.
17. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. Curr Opin Neurol. 2002;15(3):239-45.
18. van Munster CE, Jonkman LE, Weinstein HC, Uitdehaag BM, Geurts JJ. Gray matter damage in multiple sclerosis: Impact on clinical symptoms. Neuroscience. 2015;303:446-61.

19. Petzold A, Balcer LJ, Calabresi PA, Costello F, Frohman TC, Frohman EM, et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. Lancet Neurol. 2017;16(10):797-812.
20. Khalil M, Teunissen CE, Otto M, Piehl F, Sormani MP, Gattringer T, et al. Neurofilaments as biomarkers in neurological disorders. Nat Rev Neurol. 2018;14(10):577-89.
21. Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J, et al. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. Brain. 1989;112 (Pt 1):133-46.



CHAPTER 2

Outcome measures in clinical trials for multiple sclerosis

Caspar E.P. van Munster, Bernard M.J. Uitdehaag.

CNS drugs. 2017 Mar; 31(3): 217-236

ABSTRACT

Due to the heterogeneous nature of the disease it is a challenge to capture disease activity of multiple sclerosis (MS) in a reliable and valid way. Therefore, it can be difficult to assess the true efficacy of interventions in clinical trials. In phase 3 trials in MS, the traditionally used primary clinical outcome measures in MS trials are the *expanded disability status scale* and the relapse rate. Secondary outcome measures in these trials are the number or volume of T2-hyperintense lesions and Gadolinium enhancing T1-lesions on Magnetic Resonance Imaging (MRI) of the brain. These secondary outcome measures are often primary outcome measures in phase 2 trials in MS. Despite several limitations, the traditional clinical measures are still the mainstay for assessing treatment efficacy. Newer and potentially valuable outcome measures increasingly used or explored in MS trials are clinically the *MS functional composite* and patient-reported outcome measures, and on MRI brain atrophy and the formation of persisting black-holes. Several limitations of these measures have been addressed and further improvements will probably be proposed. Major improvements are the coverage of additional functional domains, such as cognitive functioning and assessment of the ability to carry out activities of daily living. The development of multidimensional measures is promising because these measures have the potential to cover the full extent of MS activity and progression. In this review, we provide an overview of the historical background and recent developments of outcome measures in MS trials. We discuss the advantages and limitations of various measures, including newer assessments such as optical coherence tomography, biomarkers in body fluids and the concept of *no evidence of disease activity*.

1 BACKGROUND

Multiple sclerosis (MS) has a female predominance and typically develops at young age with a peak incidence between 20 and 40 years.(1) Clinically, it is characterized by a large variability of symptoms arising from focal inflammation of the central nervous system which may occur at various points in time. Symptoms generally last for several days to weeks, but occasionally persist for many months, with subsequent full or partial recovery. These periods are being referred to as relapses. Radiologically, MS is characterized by typical white matter lesions that are best visualized with magnetic resonance imaging (MRI). The occurrence of clinical relapses or new white matter lesions on MRI are used to estimate *disease activity*. Demonstrating dissemination in time and place, clinical or radiological, is the core feature in the diagnostic criteria.(2)

The occurrence of relapses is the dominant clinical picture in the vast majority of patients during the earlier disease stages and is defined as relapsing-remitting MS (RRMS). If a patient only experienced a single episode with clinical symptoms, it is referred to as a clinically isolated syndrome (CIS). Relapses eventually subside and the disease course often evolves to a slow worsening of symptoms, leading to disability accrual (i.e. *disease progression*). When there is a disease progression independent from relapses, this is referred to as secondary-progressive MS (SPMS). Approximately 15% of patients have slowly progressive disease from onset without evident relapses and are categorized as primary-progressive MS (PPMS).The first effective immunomodulatory treatments were the injectables interferon-β and glatiramer acetate that were introduced in the 1990s.(3) After a decade the more potent natalizumab (in 2004) and the first oral drug fingolimod (in 2010) were introduced. More recently approved treatments include teriflunomide, dimethylfumurate, alemtuzumab and daclizumab. Ocrelizumab and cladribine are expected to be approved in the near future. In the phase 3 trials of these treatments, the outcome measures used to evaluate efficacy were relapse rate, disability worsening and MRI (formation of new T2-hyperintense lesions (T2HL) or Gadolinium enhancing T1-lesions (GdT1L)). These measures have been generally accepted as measures of (short-term) treatment effects.

Clearly, treatment options in MS are rapidly expending and are applied in patients with different clinical phenotypes. It is therefore important to have clear, comprehensive and universally accepted outcome measures. For this purpose, an outcome measure has to be valid, reliable and responsive. In practical terms this means it must measure what it intends to measure, it should be free of measurement errors and able to detect true change of performance (due to disease activity or progression).(4) Furthermore, it needs to capture clinically relevant changes and ideally has predictive value.

Unfortunately, standardized definitions of outcome measures in MS research are lacking for which there are several explanations. First, the clinical disease expression and course are highly variable which hampers defining a uniform concept of disability in MS.(5-7) There is wide variation between patients concerning relapse frequency (including seasonal variation(8)) and accrual of (relapse-related) disability. Also, patients may present with virtually all neurological symptoms that exhibit an age-dependent distribution (table 1).(7) Moreover, the extent in which symptoms contribute to overall disability is variable. This may be more dependent on the location of the lesion than on the size or activity. For example, a severe persisting hemiparesis may have a greater impact on disability than a mild sensory deficit, while both may result from pathologically comparable lesions. In fact, lesions may occur subclinically without causing disability worsening.(9) Another difficulty is that disability often accumulates slowly. Consequently, long term follow-up is needed to assess treatment effect, which makes trials time-consuming and expensive. Lastly, disability is influenced by confounding factors that may not be directly related to disease activity (e.g. fatigue, mood disturbances, deconditioning, spasticity and side effects of medication).(10)

Table 1, distribution of patients (%) by presenting clinical symptoms to age of onset.(7)

age at onset of MS (years)	optic neuritis	diplopia or vertigo	acute motor symptoms	insidious motor symptoms	balance or limb ataxia	sensory symptoms
<20	23	18	6	4	14	46
20-29	23	12	7	6	11	52
30-39	13	11	7	14	15	44
40-49	9	17	3	31	13	33
≥50	6	13	4	47	11	32

With all these difficulties in mind, we aim to provide a non-systematical comprehensive overview of clinical and paraclinical outcome measures that are used in clinical research of MS (summarized in table 2). We elaborate on traditional and newer measures such as brain atrophy, optical coherence tomography (OCT), biomarkers in body fluids and the concept of *no evidence of disease activity* (NEDA). We highlight the most important advantages, limitations and caveats of these measures.

Table 2, primary, secondary and exploratory outcome measures in phase 3 trials for MS.

Primary outcome measures
Clinical: - expanded disability status scale (EDSS): 3- or 6 months confirmed disability worsening or improvement - relapses: annualized relapse rate, time to second relapse (conversion to clinically definite MS)
Secondary outcome measures
Clinical: - MS functional composite (MSFC): timed 25-foot walk test, nine-hole peg test, paced auditory serial addition task or symbol digit modalities test Paraclinical: - T2-hyperintense lesions - Gadolinium enhancing T1-lesions - whole brain atrophy
Exploratory outcome measures
Clinical: - as candidate component of MSFC: low-contrast letter acuity test - patient-reported outcome measures: e.g. quality of life, depression and anxiety, fatigue, specific functional domains Paraclinical – imaging: - volumetric measures of specific structures (e.g. thalamus, upper cervical cord area) - persisting black-holes - functional MRI for analysis of functional connectivity - diffusion tensor imaging to examine brain tissue integrity - magnetization transfer ratio MRI as a marker for brain myelin content - optical coherence tomography Paraclinical – biomarkers: - biomarkers in body fluids: in CSF or blood Composite: - no evidence of disease activity (NEDA): typically covering (confirmed) EDSS progression, relapse rate and formation of MRI lesions; whole brain volume increasingly included (i.e. "NEDA-4") Electronical devices: - e.g. Assess MS system, Glove analyzer, accelerometers

2 CLINICAL OUTCOME MEASURES

Outcome measures can be generic or disease-specific, physician- or patient-based, direct or indirect, and may cover all or specific aspects of MS. Various clinical outcome measures are available, assessing different disease characteristics. Which characteristics are important largely depends on the aim of the study. Here, we first describe the traditional measures *expanded disability status scale* (EDSS) and relapses. Subsequently, the more recently developed *MS functional composite* (MSFC) will be discussed. Finally, we elaborate on patient-reported outcome measure (PROM)s as these patient-based measures are increasingly being used in MS trials.

2.1 The expanded disability status scale

The EDSS intends to capture disability of MS patients based on neurological examination by describing symptoms and signs in eight functional systems (FS). Furthermore, it encompasses ambulatory function and the ability to carry out activities of daily living (ADL). An overall score can be given on an ordinal scale ranging from 0 (normal neurological examination) to 10 (death due to MS). Scores from 0 to 4.0 are determined by FS scores, which means that in this range the EDSS is essentially a measure of impairment. Scores from 4.0 and higher basically address disability. Ambulatory function and the use of walking aids heavily determine the range of 4.0 to 7.0, and scores between 7.0 and 9.5 are largely determined by the ability to carry out ADL. A schematic representation of the EDSS is given in figure 1.

In clinical trials of MS, the EDSS is the most widely used outcome measure to determine disability worsening and define relapse-related change in neurological function. Furthermore, it is used as an inclusion criterion and to characterize study populations. The value of the EDSS as a surrogate outcome measure for future disability is limited.(11-15)

2.1.1 Limitations and caveats

Despite general acceptance of the EDSS, there are many limitations and caveats (summarized in table 3).(16) First of all, EDSS holds high intra- and interrater variability.(10, 11, 17-19) This can be explained by the subjective nature of the neurological examination itself on which the EDSS is largely based, particularly in the lower EDSS range. Also, complex and ambiguous scoring rules for the FS probably explain some of the variability.

Non-linearity of the EDSS is another limitation (visualized in figure 1). The staying time in the middle scores is shortest and this results in a bimodal distribution with peaks at 1.0 to 3.0 and 6.0 to 7.0.(7, 20) It means that the

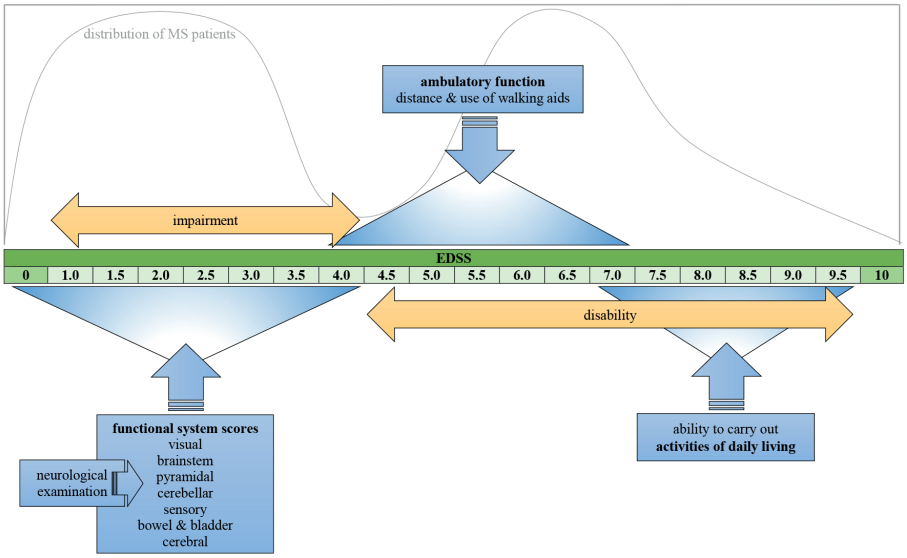


Figure 1, schematic representation of expanded disability status scale depicting the factors that determine overall score; the graph shows the distribution of patients over the expanded disability status scale (EDSS)(7)

rate of progression as assessed by the EDSS varies depending on baseline score. Furthermore, responsiveness of the EDSS is limited.(21, 16) Scores higher than 4.0 are less influenced by changes in FS scores. For example, development of a paresis in a patient with an EDSS of 6.0 will not result in a higher EDSS. Reversely, EDSS would have changed with a baseline EDSS of 4.0.

The non-linearity and limited responsiveness should both be accounted for when interpreting changes over time.(22) Nevertheless, EDSS change is often presented without accounting the baseline score. As a result, statistical significant change may erroneously be presented as clinically relevant and vice versa. An increasingly used clinically meaningful change is a change of 1.0 or more if EDSS at baseline was 0 to 5.5, and 0.5 or more for higher baseline EDSS scores. This is more driven by reproducibility data than by clinical relevance data.

Because the EDSS is an ordinal scale, non-parametric statistics should be used in statistical analysis. This implies that significant differences between groups can be calculated, but the magnitude of differences cannot. In line with this, results should not be presented with means and standard deviation, but with median values and interquartile ranges. Also, a caveat of numeric values is that they might give the false impression of being precise.

Table 3, limitations, caveats and improvements for clinical outcome measures. PASAT = paced auditory serial addition task.

Expanded disability status scale (EDSS)	
Limitations and caveats	*Improvements*
- High intra- and interobserver variability - Non-linearity (bimodal distribution) - Limited responsiveness - Necessity to use non-parametric statistics (ordinal scale) - Uneven distribution of relapsing-remitting and progressive patients -Several functional domains not assessed	- Accounting for baseline score when determining change (e.g. change ³1.0 with baseline score 0 to 5.5, and ³0.5 for higher baseline scores) - Determining disability worsening with confirmation of the EDSS progression after at least six months - Using standardized scripts for questioning patients (improving reliability and decreasing risk of unblinding) - Simplification of scoring rules (decreasing variability) - Streamlining by stripping components of the functional systems that are less informative - Modification to improve linearity and facilitate statistical analysis
Relapses	
Limitations and caveats	*Improvements*
- Strong subjectivity - Recovery of signs or symptoms before confirmation of relapse - Recall bias of patient and observer bias of examiner - Newly reported symptoms not always clearly depicted in change of the EDSS - Identification largely depended on patient reporting it - Higher relapse rate prior to inclusion: over-reporting to fulfill inclusion criteria, high relapse rate inclusion criterion leading to decrease of relapse rate because of regression to the mean, placebo effect, decrease of relapse due to natural course of MS	- Confirming a relapse by another examiner - Increasing number of visits to identify more relapses

Multiple sclerosis functional composite (MSFC)	
Limitations and caveats	*Improvements*
- moderate reliability, sensitivity and responsiveness of the PASAT - the PASAT often disliked by patients, requirement of mathematical ability and ceiling effect - Several important functional domains are not assessed - lack of a clear dimension of the overall score (resulting in difficult interpretability) - Z-scores are influenced by results of the reference population and obscure the meaning of crude scores	- Replacing the PASAT with the symbol digit modalities test - Adding the low-contrast letter acuity test (covering visual domain) - Adding other functional domains - Determining minimal clinically relevant changes of the Z-scores and confirming change after six months - Determining clinical relevance - keeping elements separated instead of combining them into a single score
Patient-reported outcome measures (PROM)	
Limitations and caveats	*Improvements*
- Unblinding nature - Potential expectance bias - Assessment of quality of life may be influenced by multiple factors - Possible response shift over time	- Weighing of individual questions appropriately - Using (computer) adaptive testing to reduce test length and improve tolerability

Another limitation is that clinical phenotypes are unevenly distributed across the EDSS. Because ambulatory dysfunction is one of the main characteristic in patients with progressive disease (SPMS and PPMS), these patients represent a larger proportion in the range of 4.0 to 7.5.(23, 24)

Lastly, several domains are not (sufficiently) assessed. Examples are cognitive function, mood, energy level and quality of life. Symptoms in these domains are frequently observed in MS patients and they may influence FS scores, ambulation and ADL function.

2.1.2 Suggested improvements

During an *international conference on disability outcomes in MS* (ICDOMS), that was held in 2011, several refinements for the EDSS were suggested to improve performance.(25) Firstly, standardized script for questioning patients (which is necessary for some FS scores) might improve reliability and decrease the risk of unblinding in clinical trials (an example of the Neurostatus form may be found on http://www.neurostatus.net/). Secondly, simplification of scoring

rules might reduce intra- and interrater variability. Thirdly, long term disability worsening should be assessed with confirmation of EDSS worsening at six rather than three months. The main reason for this is that relapses may improve beyond three months, and thus EDSS worsening may be temporary.(26) Fourthly, streamlining of the EDSS might be achieved by finding the components of FS that contribute most to confirmed worsening of disability and omitting the other less informative components. Lastly, modification of the EDSS to improve linearity of measurement will facilitate statistical analysis and clinical understanding.

Whatever its limitations, the EDSS probably continues to be the main disability measure in the near future because of the vast experience with it and the possibility to make historical comparisons. Until we have better alternatives, clinical assessment can be improved by using the EDSS in conjunction with other measures.

2.2 Relapses

The other traditional outcome measure is assessment of relapses. By consensus, a relapse has been defined as new or worsening neurological symptoms that are objectified on neurological examination in the absence of fever and last for more than 24 hours, and have been preceded by a period of clinical stability of at least 30 days, without another explanation than MS.(27, 28) The relation of number of relapses with disability worsening is not completely clear, although conclusions may be drawn from natural history studies. Various of these studies showed that relapses early in the course of MS were associated with long term disability and increased risk of conversion to SPMS, which probably relates to faster disability worsening.(29-32) However, superimposed relapses in the progressive phase did not lead to faster disability worsening.(33) Treatment effects on relapses is confined in the change of annualized relapse rate or time to second relapse (i.e. conversion to clinically definite MS).(34) Treatment effect on relapses gives a fair reflection of short-term efficacy.

2.2.1 Limitations and caveats

There are several caveats when using relapses as an outcome measure (summarized in table 3). First of all, identification of a relapse is subjective. Ensuring perfect blinding for treatment is therefore essential. To limit subjectivity, a second assessment can be performed to objectify the relapse. The problem with this approach is that symptoms or signs may already have recovered, and recall bias of the patient and observer bias from the examiner may influence the second assessment.(35)

Another caveat is that identification of a relapse largely dependents on a patient reporting new symptoms. When a patient only reports new symptoms on scheduled visits and not spontaneously, the established relapse rate will be lower than in reality. In fact, increasing the number of visits in a trial period may increase the relapse rate.(36)

An interesting phenomenon is that relapse rate is often remarkably high prior to inclusion into trials. Various explanations may be given for this. (37, 38) First of all, relapses in the preceding period of a trial are usually determined retrospectively and patients may over report the exact number to qualify for inclusion. Secondly, the inclusion criterion of relapse rate is often high, meaning that only patients with very active disease are included. As a consequence, it can be expected that relapse rate of these patients will decrease towards a disease average during the trial (i.e. *regression to the mean*). Thirdly, patients participating in a trial may do better merely because of a placebo effect or better comprehensive care during the trial. Lastly, during the natural course of MS the relapse rate will eventually decrease, independent of treatment.(39) These factors may obscure the interpretation of absolute relapse rate reduction in treatment trials.

2.3 The multiple sclerosis functional composite

Because of the limitations of the EDSS and assessment of relapses, the MSFC was developed to improve clinical assessment.(40, 41) It was introduced in the early 1990s, a time when the first effective treatments were introduced. In contrast with the EDSS, the MSFC covers three functional domains: ambulatory, hand and cognitive function (a schematic summary is given in figure 2). The results of the tests that assess these domains are depicted in an interval scale (seconds or number of correct responses) and can be converted to a Z-score that is based on values of a reference population.(42) An overall score can be calculated by averaging the Z-score of the subtests.

The MSFC has been extensively evaluated. The overall score of MSFC correlated strongly with EDSS (43) and subtest scores did moderately.(40) Also, change of MSFC correlated with EDSS change and relapse rate.(40, 44, 45) Furthermore, it was predictive of conversion from RRMS to SPMS.(44) Concerning the relation with MRI abnormalities, MSFC correlated with white matter lesion load and various atrophy measures.(46-48) Lastly, correlations with several PROMs,(49, 50, 43, 51) employment status(52) and driving performance(53) were found.

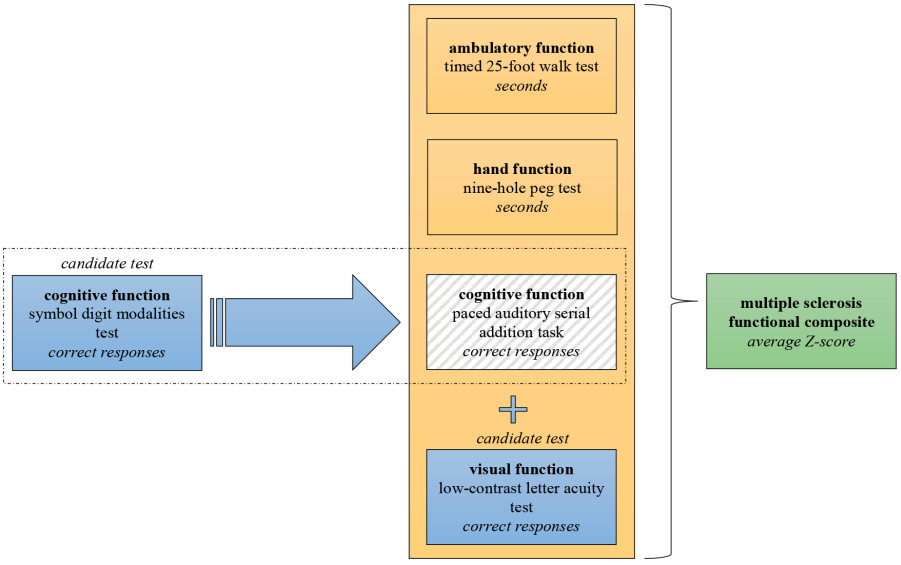


Figure 2, schematic representation of the multiple sclerosis functional composite (MSFC) with candidate components.

2.3.1 The original components

Ambulatory function is tested with the *timed 25-foot walk test* (T25W, explained in table 4). The T25W is a reliable test for patients with more severe gait impairment, because it primarily assesses walking speed. Assessing walking speed seems clinically relevant, because it relates to the capacity to perform outdoor activities important in daily life.(54) For patients with mild gait impairment the T25W may not be sensitive enough to detect abnormalities and because of that has a floor ceiling effect.(55) For these patients it may be more appropriate to assess walking *endurance* with longer walking distances, for example with a 6 minutes walking test.(56)

2.3.2 Candidate components

A candidate cognitive test that may replace the criticized PASAT is the *symbol digit modalities test* (SDMT, explained in table 4).(62, 63) It measures information processing speed. The advantages of the SDMT are that it is easily administrated, better tolerated by patients (probably because there is no time pressure),(64) and more robust and reliable than the PASAT. (65, 66) Moreover, the SDMT correlated more strongly with white matter abnormalities than PASAT.(67, 68) It also correlated with worsening of cognitive impairment(69, 70) and MRI abnormalities (atrophy measures in

Table 4, description of components of the multiple sclerosis functional composite (MSFC).

Original components	
Timed 25-foot walk test (T25W)	The patient is directed to one end of a clearly marked 25-foot course and is instructed to walk 25 feet as quickly as possible, but safely. The task is immediately administered again by having the patient walk back the same distance. Patients may use assistive devices when doing this task. In clinical trials, it is recommended that the treating neurologist select the appropriate assistive device for each patient.[42]
Nine-hole peg test (9HPT)	The patient is asked to take nine small pegs one-by-one from a small shallow container, place them into nine holes and then remove them and place them back into the container. Results are depicted in seconds to complete the task of both the dominant and non-dominant hand; two trials for each side.[42]
Paced auditory serial addition task (PASAT)	The PASAT is presented on audiocassette tape or compact disc to control the rate of stimulus presentation. Single digits are presented either every 3″ (or every 2″ for the optional 2″ PASAT) and the patient must add each new digit to the one immediately prior to it. The test score is the number of correct sums given (out of 60 possible) in each trial. To minimize familiarity with stimulus items in clinical trials and other serial studies, two alternate forms have been developed; the order of these should be counterbalanced across testing sessions. The PASAT is the last measure of the MSFC that is administered at each visit.[42]
Candidate components	
Symbol digit modalities test (SDMT)	Patients are presented with a key that includes nine numbers, each paired with a different symbol. Below this key is an array of these same symbols in pseudorandom order paired with empty spaces. Patients must then provide the correct numbers that accompany the symbols as indicated in the key.[64]
Low-contrast letter acuity test (LCLA)	Seven charts with different levels of contrast (0.6 – 100%) are presented to the patient. On each chart multiple rows are depicted with gray letters with decreasing size on a white background. The letter scores indicate the number of letters identified correctly. Each chart is scored separately.

particular).(71, 72) A limitation is that a patient has to have an intact visual system, which may be impaired in MS patients. Although there is a ceiling effect, it is less pronounced than for the PASAT. Taken all points together, the SDMT probably is a good replacement for the PASAT.

When the MSFC was developed, no data on suitable tests to assess visual function were available. The past decade, various visual outcome measures for MS research have been studied.(73) Of these the *low-contrast letter acuity test* (LCLA, explained in table 4) may be a good candidate to add to the MSFC.(74) Results correlated with clinical phenotypes, MRI abnormalities and PROMs for visual impairment and quality of life (which supports clinical relevance).(75, 76) Moreover, some clinical trials showed treatment effect on the LCLA in the active group compared with placebo.(77)

2.3.3 Limitations and caveats

There are several limitation and caveats of the MSFC (summarized in table 3). A frequently postulated objection to the MSFC is that the overall score lacks a clear dimension, which hinders interpretability and therefore appears to be difficult for the interpreter to get familiar with the score. In other words, it is difficult to form a *mental picture* of it.(78) This difficulty may be addressed by keeping the elements of the MSFC score separated instead of combining them into a single score. Nonetheless, comparison of subtest results between studies will remain impossible due to the Z-scores that obscure the meaning of crude scores.

Another problem is that results of the reference group strongly influences the Z-scores of patients.(79) With that, assessing changes in time is problematic, because the overall score is influenced by variability between time-points of both the reference and patient group. Consequently, it is impossible to determine if change is a result of statistical variance or true progression of disability.(38)

A potential solution to some of the statistical caveats of Z-scores might be to determine the minimal clinically relevant change.(21, 80) This means that change should be confirmed on a subsequent time point, preferably at six months (because of possible disability improvement after a relapse). This approach has been tested in a clinical trial dataset.(45) Sensitivity of worsening was found to be similar between MSFC and EDSS, and it correlated with other clinical and MRI outcome measures. However, the downside of this approach is that it will hamper sensitivity to change, which is of particular importance in patients with severe disability.

Despite its disadvantages, the MSFC is an appealing alternative for the EDSS. It can be performed within 20 minutes, covers three domains, has good intra- and interrater reliability and it results in a score on a continuous scale. The MSFC has once been used as primary outcome in a treatment trial in SPMS.(49) While MSFC progression was slowed, treatment effects were not observed with the EDSS. When applying its components in a sensible way, it may be used as a primary endpoint in future clinical trials.

2.4 Patient-reported outcome measures

A PROM is defined as "*any report of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else*".(81) A PROM may provide valuable insight into the patient-perspective of a treatment or matter of interest. For example, treatment success for a patient might be more influenced by adverse events than a physician perceives or deducts from other outcome measures. Furthermore, it may detect clinically meaningful changes and leave out changes with no clinical relevance. A PROM can assess perceived efficacy, side-effect, depression and anxiety, fatigue, mobility, quality of life, ability to carry out ADL, sexual dysfunction and symptoms specific for MS. A list of PROMs that are being used in MS research is presented in table 5.(82-105)

Patient-reported outcome measures that assess the ability to carry out ADL may be of particular value. They are able to demonstrate clinical relevance of MS-specific outcome measures. For example, one study found a correlation between the EDSS and a 42 items ADL scale which was mostly driven by impairment of mobility.(106) Another advantage is that measuring ADL activity allows comparison between studies of MS as well as other diseases. Currently, no MS-specific ADL measures are available. Nevertheless, PROMs that were developed for stroke patients (Ranking scale(107, 108) and Bartel index(109)) were used in some MS trials.(110, 111)

There are several limitations of PROMs (summarized in table 3). Among these, are the unblinding nature and potential expectance bias. Also, questionnaires assessing quality of life are prone to be influenced by more than only disability. Other factors that are commonly seen in MS patient contribute as well (e.g. fatigue, depression, anxiety and physical comorbidities).(112) Also, the individual questions should be weighted appropriately. Summing up all the subscores assumes equal importance which is generally not the case. Lastly, PROMs are prone to response shift over time.(113) Response shift occurs when a patient answers an item differently than previously done due to change of internal standards, values or conceptualization of the purposed domain (e.g. quality of life).

Typically, PROMs are fixed in length and all patients have to fill in the complete questionnaire. The number of questions that have to be answered can be reduced with computer adaptive testing.(114) It leads the patient through an iterative process in which the answer to a question determines what question is presented next. For example, if a patient is fully dependent on a wheelchair, a question about climbing stairs is irrelevant. With these methods patients' tolerability for a questionnaire may be improved.

Table 5, patient-reported outcome measures that are used in MS research.

Measure
Quality of life
MS quality of life-54[103]
MS quality of life inventory[86]
European quality of life-5D[87]
Health utilities index mark 3[87]
World health organization quality of life brief form[100]
Sickness impact profile[83]
Life satisfaction questionnaire[96]
Hamburg quality of life questionnaire in MS[91]
Quality of life index[85]
Leeds MS quality of life scale[90]
Disability and impact profile[101]
The MS international quality of life questionnaire[102]
Functional assessment of MS[84]
Depression and anxiety
Beck depression inventory[82]
Patient health questionnaire-9[95]
Hospital anxiety and depression Scale[94]
Fatigue
Modified fatigue impact scale[89]
Fatigue impact scale for daily use[88]
Single functional domain
MS walking scale-12[93]
Arm function in MS questionnaire[98]
Visual function questionnaire-25[99]
Multiple domains
Short form-36[104]
MS impact scale-29[92]
Guy's neurological disability scale[97]
MS impact profile[105]

3 PARACLINICAL OUTCOME MEASURES

Numerous paraclinical outcome measures are available and could be used as adjunct to clinical measures to obtain information on treatment efficacy. Some are potentially valuable (e.g. cerebrospinal fluid (CSF), visual evoked potentials) while others are less suitable (e.g. brainstem auditory evoked potentials).(115) Here, we shortly discuss the value of white matter pathology as detected on MRI. Subsequently, we will elaborate on newer outcome

measures, such as brain atrophy, *persisting black-holes* (PBH), OCT and biomarkers in body fluids.

3.1 Magnetic resonance imaging

3.1.1 White matter pathology

Magnetic resonance imaging is sensitive to detect, characterize and quantify lesions in the white matter. It plays a fundamental role in the McDonald diagnostic criteria for MS to demonstrate dissemination in time and space in addition to clinical signs.(2) Radiological dissemination in space is defined as having at least one lesion in at least two for MS typical areas in the central nervous system. Dissemination in time is determined when at least one new lesion is demonstrated on a follow-up MRI, or if one asymptomatic Gadolinium-enhancing and one non-enhancing lesion is demonstrated on the initial MRI.

The *MAGNIMS workgroup* recently proposed a revision of these criteria allowing even earlier diagnosis with MRI.(116) The value of MRI as a diagnostic tool is principally the high sensitivity to detect (past) disease activity. Formation of new T2HL and GdT1L may occur subclinical and are thus more frequently seen than clinical relapses.(9, 117) The moderate correlation of T2HL load with relapse rate(118, 26) and disability(119, 120) is possibly related with this phenomenon. Nevertheless, white matter pathology has predictive value for the clinical disease course. For example, patients with a CIS and a high T2HL load at baseline had an increased risk to reach an EDSS of 3.0.(121) Also, the presence of two or more GdT1L in patients treated with interferon-β predicted EDSS worsening at 15 years.(122)

Because of the high sensitivity for detecting disease activity, MRI has been widely accepted as a secondary endpoint in clinical trials. Moreover, demonstrating efficacy on MRI lesions is crucial in the development of immunomodulatory treatments. Treatment effects on MRI could also act as a surrogate endpoint for clinical disease activity. A study supported this by showing that treatment effect on MRI activity explained more than 80% of the variance of treatment effect on relapse rate.(123) Other studies confirmed this by showing the related MRI effects on relapse rate and accumulation of disability worsening (up to 16 years).(124-126)

These classical MRI parameters largely depicts (past) neuro-inflammation in MS. However, the neurodegenerative aspect of MS is being increasingly studied with MRI. One reason for this is that with the current therapy we are now able to suppress neuro-inflammation effectively, but the ultimate goal of therapy is prevention of neuronal tissue loss or, in the long run, to stimulate neuronal repair. Another reason is that neuropathological and

MRI techniques have improved our insight in underlying neurodegenerative processes of MS.(127) Consequently, measures that reflect these processes are more frequently used as secondary outcome measures. The most widely used neurodegenerative MRI measures are atrophy and PBH.

3.1.2 Atrophy

Brain volume loss in MS patients occurs considerable faster than in healthy people: 0.5 to 1.0% versus 0.1 to 0.3% brain volume loss per year.(128, 129) Atrophy may be found throughout the disease course, even in the early phases.(130) Remarkably, the atrophy rate of gray matter structures accelerates in patients with SPMS to a 14-fold of that of healthy persons.(131) Virtually all gray matter structures are affected, although variation exists between clinical phenotypes.(132)

Brain volume can be visualized in various ways. The somewhat older measures assess loss of brain volume indirectly by measuring corpus callosum size,(133) bicaudate ratio(72) and ventricular volumes(133, 72). Also, whole brain volume can be measured directly with conventional MRI. (72, 128) Nowadays, segmentation of the brain into white and gray matter compartments or specific gray matter structures is possible and several automated methods reduced processing time.(134-136)

The relation of atrophy measures with clinical signs has been extensively investigated. Whole brain and gray matter atrophy correlated strongly with disability and cognitive impairment, both cross-sectional and longitudinal. (132) These correlations existed throughout the disease course and clinical phenotypes. Atrophy of gray matter structures may even be more closely related to clinical signs than white matter lesion or whole brain atrophy. (137) Atrophy of several structures correlated remarkably strong with certain clinical symptoms. For example, cerebellar gray matter atrophy correlated strongly with cerebellar symptoms and hand function,(138) upper cervical cord area with ambulatory dysfunction,(139) and hippocampal atrophy with memory deficits.(140) Thalamic volume showed a remarkably firm correlation with cognitive impairment.(141) Also, various atrophy measures showed predictive value for future disability and cognitive impairment.(142, 137, 143, 144)

Furthermore, spinal cord volumes can be assessed for which the upper cervical cord area is often used. Several studies showed a correlation with spinal cord volume loss and clinical disability. (145, 146, 144) It has also been correlated with long term disability.(147)

An extensive summary of clinical trials that used brain atrophy as a secondary endpoint may be found elsewhere.(148, 149) Noteworthy is a recent meta-

analysis that showed that 75% of the variance of treatment effect on disability was explained by whole brain atrophy and T2HL.(150) Another meta-analysis found evidence that whole brain atrophy in patients that received immunomodulatory treatment was lower than placebo group.(151)

Although volumetric measurements are appealing outcome measures, there are some caveats and limitations. Firstly, atrophy accumulates very slowly, what in generally means that longer follow-up is needed to detect significant changes. Clearly, this accounts particularly for treatment effects on smaller structures, such as thalamic volume. Secondly, the short-term effect of immunosuppression on brain tissue may cause decrease of brain volume due to resolution of inflammation. This volume loss is not a sign of neurodegeneration, because there is no loss of neuronal tissue. This is often referred to as *pseudo-atrophy*. Importantly, this effect may last up to one year after initiation of treatment.(152, 153) Thirdly, various physiological variations in the content of the intra- and extracellular compartments affect volumetric measurements.(154) Lastly, factors that are not MS-specific (such as dehydration, alcohol use, smoking, genetic variation, comorbidities and age) may influence brain volume.(154)

3.1.3 Persisting black-holes

Another MRI marker for neurodegeneration is formation of PBH. These lesions are often defined as non-enhancing T2HL with persisting signal intensity between that of the gray matter and the CSF on T1-weighted scans.(155) Approximately 30 to 40% of active T2HL will eventually evolve into PBH within six to 12 months.(156) The underlying neuropathology of PBH is severe and irreversible tissue damage.(156) Accumulation of PBH is associated with accrual of disability.(157, 158) Furthermore, the PBH load correlated with disability worsening over 10 years.(159) Some clinical trials found significant effects of treatment on the formation of PBH.(160-163)

Several more advanced MRI techniques are potentially valuable outcome measures, although they need further research to clarify the exact relevance. Examples are *functional MRI* for analysis of functional connectivity,(164) *diffusion tensor imaging* to examine brain tissue integrity(165) and *magnetization transfer ratio MRI* as a marker for brain myelin content.(166, 167)

3.2 Optical Coherence Tomography

The retina can be visualized non-invasively, safely and fast with OCT. This technique uses the reflection of near infra-red light on the retina. Different layers of the retina can be distinguished on high-resolution images. It has

been proven to be valuable in quantifying pathology in these layers, although the exact underlying pathophysiological processes of these findings are largely unclear.(168, 169)

Most findings of the research with OCT in MS point to neurodegenerative changes such as axonal loss and neuronal soma shrinkage.(170) Therefore, OCT is a good candidate outcome measure to assess treatment effect on neurodegeneration, which makes it an attractive tool in progressive MS trials. For this purpose, the *retinal nerve fiber layer* (RNFL) is of particular interest. The thickness of this layer may be decreased following optic neuritis,(171, 172) but also decreases more slowly in patients without prior optic neuritis. (171, 173) The latter may indicate ongoing neurodegeneration. Furthermore, RNFL thickness correlated with cerebral atrophy measures(174, 175) and with axonal loss in the anterior visual pathway.(176, 177)

Clinically, thinning of the RNFL correlated with worse performance on the LCLA (explained in table 4),(171, 178) and a reduced visual quality of life.(179) Correlations of RNFL thickness with EDSS were less consistent.(180, 181) In a recent large multicenter study of patients without prior optic neuritis, persons with a RNFL thickness in the lowest tertile at baseline had double the risk of disability worsening in two years compared with the other tertiles. (182) The risk further increased with a longer follow-up. The clinical relevance of other layers, such as macular volume(183) and retinal ganglion-cell/ inner plexiform layer thickness,(184, 185) is less clear.

The advantage of OCT over MRI that it is technically easier and widely accessible. When using a predefined scanning protocol it has a good reliability.(186) Nevertheless, further research is needed before OCT can be implemented as an outcome measure. This is particularly the case for longitudinal data of the various layers.

3.3 Biomarkers in body fluids

Both MRI and OCT allow detection of neuro-inflammation and -degeneration at various time-points, but have limited sensitivity to detect ongoing processes. Biomarkers in body fluids, such as CSF and blood, might be more useful for this purpose. Although it is beyond the scope of this review to discuss this topic thoroughly (recently reviewed elsewhere(187)), a few biomarkers are worth mentioning.

There are several potential valuable CSF biomarkers that might give a real-time reflection of ongoing neurodegeneration. A biomarker that reflects axonal injury is neurofilament. This protein is a major component of the axonal cytoskeleton and is released following neuronal damage.(188) Neurofilament levels in CSF are generally raised in MS patients, particularly during an acute

relapse.(189, 190) Furthermore, increased levels were associated with worse EDSS,(190) faster disability worsening in 15 years, (191) Gd-enhancing lesion load,(192) and atrophy (of the brain and spinal cord) in 15 years.(193) Neurofilament levels were also responsive to treatment with fingolimod(194) and natalizumab,(195) and therefore might be biomarkers for treatment effect.

Other proteins of the axonal cytoskeleton that can be measured in CSF are actin(196, 197) and tubulin.(198, 197) Proteins that indicate ongoing disease activity are sphingolipids (component of the myelin sheet),(199) glial fibrillary acidic protein (GFAP),(200) S100B(200) and Chitinase 3-like proteins.(201)

Compared to CSF, blood is generally less well studied for biomarkers, but clearly has the advantage that it is much easier to obtain. Similarly like in CSF, neurofilament in the blood might act as a biomarker for neurodegeneration. Neurofilament levels predicted recovery of spinal cord lesions,(202) and higher concentrations were associated with faster conversion to definite MS and more cerebral lesions.(203) Another biomarker that is used to determine bioactivity of interferon-β is myxovirus-resistance protein A (MxA). It also seems to be indicative for recent and future disease activity.(204, 205) Lastly, various small noncoding microRNAs are potentially valuable for predicting disease course and treatment response.(187)

The exact value of these biomarkers as outcome measures will have to be determined. If clinically meaningful, they will probably be used in combination with other measures. They may be particularly useful to assess treatment effects in trials with progressive MS, because identification of progression or neurodegenerative changes remains very challenging.

4 NO EVIDENCE OF DISEASE ACTIVITY

The concept of a *disease activity free status* as the ultimate treatment goal has been used in other medical conditions, including cancer and inflammatory diseases such as rheumatoid arthritis. It implies the absence of measurable disease activity. This concept has been translated to NEDA and is used in more recent MS trials as a secondary outcome measure.(206, 207) It is essentially a multidimensional measure that typically covers (confirmed) EDSS progression, relapse rate and formation of MRI lesions (T2HL or GdT1L). However, any parameter related with disease activity may be added.

A recent study in a cohort of RRMS patients found that NEDA at two years had a positive predictive value for absence of disability progression at seven years of 78%.(207) Furthermore, the predictive value of NEDA was greater than each of the individual components. Other studies also showed that

combinations of clinical and MRI parameters had better predictive value for disability progression than individual measures.(125, 208, 150, 209, 210) For example, a recent meta-analysis found that treatment effect on T2HL and brain volume combined explained 75% of the variance of disability progression in two years, and this was significantly higher than predictive values of the MRI measures individually.(150)

In clinical practice, NEDA-like models are used to identify responders and non-responders to treatment. Examples are the *Modified Rio Score*(211) and the *Canadian treatment optimization recommendation model*.(35) Such tools need to have good long-term predictive power for disability, before treatment decision can be based on it.

When using NEDA as an outcome measure to assess treatment efficacy it is important to consider the timing of assessment. The reason for this is that a treatment needs to have had enough time to become effective. This can be illustrated by the finding that 70% of patients had NEDA two years after initiating treatment with natalizumab with baseline assessment after one year, compared with 51% NEDA with a baseline at initiation of therapy.(212) For alemtuzumab timing is different, because the true treatment effect starts after the second infusion cycle, one year after the initial course.(213) This issue has implications when determining if NEDA can be a valid outcome measure for disability at the long run.

Although NEDA seems an appealing outcome measure in some ways, it is not yet clear which (functional) domains are important to include and when or how frequently these should be assessed. It should, for example, reflect what is important in daily life for patients. Therefore, including a PROM seems indispensable. Also, markers for neurodegeneration should be involved when tissue loss is considered to be the ultimate treatment goal. Therefore, brain volume is increasingly added to NEDA (referred to as NEDA-4).(214) However, adding more assessments likely reduces the number of patients fulfilling NEDA, and may raise the bar to a too high level resulting in the rejection of highly active, but not perfect, interventions.

Taken together, NEDA will continue to evolve while evidence accumulates about what are valuable outcome measures. Standardization of timing and functional subdomains are imperative for comparison between studies.

5 FUTURE PERSPECTIVES

The number and quality of outcome measures is increasing, and with that the assessment of treatment efficacy will improve over the coming years. Until new measures are validated and generally accepted, the traditional outcome measures EDSS and relapse rate will remain primary endpoints in clinical trials. However, it is very unlikely that these measures are sufficient to fully assess treatment efficacy. Eventually, measures that more explicitly capture multiple dimensions (e.g. MSFC and NEDA), will probably become the new standard. They are particularly useful to detect infrequent events (e.g. relapses) or small changes (e.g. brain atrophy and disability worsening) under treatment, which is increasingly important with highly effective therapy. The same accounts for treatment of progressive disease (SPMS and PPMS), in which small and gradual treatment effects can be expected. Moreover, multidimensional measures might decrease duration and size of clinical trials. The caveats of multidimensional measures that have to be taking into account are summarized in table 6.(25)

In addition to improvement of existing outcome measures, innovative techniques such as electronic devices and mobile device applications are potentially valuable. They allow, for instance, multiple or continuous assessment which might give a more adequate picture of a patients' ability or disability and the impact of the disease on daily living.

Table 6, limitations and caveats of multidimensional measure.

- Interpretation may not be straightforward, particularly if clinical relevance of (some) components are not immediately obvious
- An overall score lacks a clear dimension, which complicates the interpretability of the score
- Components should be normalized or weighted without obscuring the clinical meaning
- Components may shift in opposite directions (improvement vs harm) which might obscure interpretation of treatment efficacy
- Components should capture the expected (biological) effects of the intervention under investigation
- Increasing the number of components not necessarily increases sensitivity
- Redundant components might cause a large change of the composite score in patients that have symptoms in that domain, while the change may be smaller or absent in patients with symptoms in other domains
- Increasing sensitivity to change does not necessarily lead to higher sensitivity for treatment effects
- Dichotomization of the results (e.g. no evidence of disease activity) will inherently cause loss of information

Several electronic devices are under development to assess disability. An example of this is the *Assess MS system* that uses an infra-red camera to register movements of upper and lower limbs, trunk and ambulation for automatic quantification of these movements. Results from a pilot study in MS are promising and these preliminary results are currently being validated with a new high-resolution camera.(215) Another device that has been developed is the *Glove analyzer* system that is able to record data from finger movements to assess hand and arm function.(216) Also, accelerometers are potentially useful tools to measure mobility automatically.(217) Apart from other attractive aspect, electronic devices are free of intrarater variability.

Mobile device applications are increasingly being used in the medical field and are also potentially useful in assessing outcome in MS trials. Applications can be easily distributed and accessible for everyone with a smart phone. They can be used in several ways, for example, for assessing a PROM on regular basis up to several times per day. Also, applications may be connected online with investigators to get real-time access to or feedback from a patient status. This may decrease the number of visits needed or could help to decide whether or not a face to face contact with a patient is needed. In the past years, healthcare *hackathons* (i.e. an acronym of HACKers marATHONS) were organized to stimulate development and integration of medical devices and mobile device applications.(218, 219) However, many of these applications need rigorous scientific validation before they may be considered as outcome measures in clinical trials.

6 CONCLUSIONS

To conclude, assessing outcome in clinical trials in MS is not straightforward and therefore a challenging field. Although much has been achieved the past decades, "*old habits die hard*" and traditional measures probably remain the standard in the near future. When more advanced measures have proven their value, they need to earn general acceptance by health care providers and especially regulatory agencies. In the end, only multidimensional measures will allow full coverage of disease activity and progression of MS and are thus best suited to assess treatment efficacy in MS trials.

REFERENCES

1. Compston A. McAlpine's multiple sclerosis. 4th ed. Philadelphia: Churchill Livingstone Elsevier; 2005.
2. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292-302. doi:10.1002/ana.22366.
3. McGraw CA, Lublin FD. Interferon beta and glatiramer acetate therapy. Neurotherapeutics. 2013;10(1):2-18. doi:10.1007/s13311-012-0163-4.
4. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45. doi:10.1016/j.jclinepi.2010.02.006.
5. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372(9648):1502-17. doi:10.1016/S0140-6736(08)61620-7.
6. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology. 1996;46(4):907-11.
7. Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J et al. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. Brain. 1989;112 (Pt 1):133-46.
8. Jin Y, de Pedro-Cuesta J, Soderstrom M, Stawiarz L, Link H. Seasonal patterns in optic neuritis and multiple sclerosis: a meta-analysis. J Neurol Sci. 2000;181(1-2):56-64.
9. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. Curr Opin Neurol. 2002;15(3):239-45.
10. Amato MP, Fratiglioni L, Groppi C, Siracusa G, Amaducci L. Interrater reliability in assessing functional systems and disability on the Kurtzke scale in multiple sclerosis. Arch Neurol. 1988;45(7):746-8.
11. Ebers GC, Heigenhauser L, Daumer M, Lederer C, Noseworthy JH. Disability as an outcome in MS clinical trials. Neurology. 2008;71(9):624-31. doi:10.1212/01.wnl.0000313034.46883.16.
12. Freedman MS. Long-term follow-up of clinical trials of multiple sclerosis therapies. Neurology. 2011;76(1 Suppl 1):S26-34. doi:10.1212/WNL.0b013e318205051d.
13. Goodin DS, Traboulsee A, Knappertz V, Reder AT, Li D, Langdon D et al. Relationship between early clinical characteristics and long term disability outcomes: 16 year cohort study (follow-up) of the pivotal interferon beta-1b trial in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2012;83(3):282-7. doi:10.1136/jnnp-2011-301178.
14. Liu C, Blumhardt LD. Disability outcome measures in therapeutic trials of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts. J Neurol Neurosurg Psychiatry. 2000;68(4):450-7.
15. Rudick RA, Lee JC, Cutter GR, Miller DM, Bourdette D, Weinstock-Guttman B et al. Disability progression in a clinical trial of relapsing-remitting multiple sclerosis: eight-year follow-up. Arch Neurol. 2010;67(11):1329-35. doi:10.1001/archneurol.2010.150.
16. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain. 2000;123 (Pt 5):1027-40.

17. Francis DA, Bain P, Swan AV, Hughes RA. An assessment of disability rating scales used in multiple sclerosis. Arch Neurol. 1991;48(3):299-301.
18. Goodkin DE, Cookfair D, Wende K, Bourdette D, Pullicino P, Scherokman B et al. Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). Multiple Sclerosis Collaborative Research Group. Neurology. 1992;42(4):859-63.
19. Noseworthy JH, Vandervoort MK, Wong CJ, Ebers GC. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. Neurology. 1990;40(6):971-5.
20. Weinshenker BG, Rice GP, Noseworthy JH, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: a geographically based study. 4. Applications to planning and interpretation of clinical therapeutic trials. Brain. 1991;114 (Pt 2):1057-67.
21. Hobart J, Blight AR, Goodman A, Lynn F, Putzki N. Timed 25-foot walk: direct evidence that improving 20% or greater is clinically meaningful in MS. Neurology. 2013;80(16):1509-17. doi:10.1212/WNL.0b013e31828cf7f3.
22. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. Mult Scler. 1995;1(1):37-47.
23. Heesen C, Bohm J, Reich C, Kasper J, Goebel M, Gold SM. Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. Mult Scler. 2008;14(7):988-91. doi:10.1177/1352458508088916.
24. Sutliff MH. Contribution of impaired mobility to patient burden in multiple sclerosis. Curr Med Res Opin. 2010;26(1):109-19. doi:10.1185/03007990903433528.
25. Cohen JA, Reingold SC, Polman CH, Wolinsky JS, International Advisory Committee on Clinical Trials in Multiple S. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. Lancet Neurol. 2012;11(5):467-76. doi:10.1016/S1474-4422(12)70059-5.
26. Lublin FD, Baier M, Cutter G. Effect of relapses on development of residual deficit in multiple sclerosis. Neurology. 2003;61(11):1528-32.
27. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121-7.
28. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol. 1983;13(3):227-31. doi:10.1002/ana.410130302.
29. Debouverie M, Pittion-Vouyovitch S, Louis S, Guillemin F, Group L. Natural history of multiple sclerosis in a population-based cohort. Eur J Neurol. 2008;15(9):916-21. doi:10.1111/j.1468-1331.2008.02241.x.
30. Eriksson M, Andersen O, Runmarker B. Long-term follow up of patients with clinically isolated syndromes, relapsing-remitting and secondary progressive multiple sclerosis. Mult Scler. 2003;9(3):260-74.
31. Scalfari A, Neuhaus A, Degenhardt A, Rice GP, Muraro PA, Daumer M et al. The natural history of multiple sclerosis: a geographically based study 10: relapses and long-term disability. Brain. 2010;133(Pt 7):1914-29. doi:10.1093/brain/awq118.

32. Tremlett H, Yousefi M, Devonshire V, Rieckmann P, Zhao Y, Neurologists UBC. Impact of multiple sclerosis relapses on progression diminishes with time. Neurology. 2009;73(20):1616-23. doi:10.1212/WNL.0b013e3181c1e44f.
33. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. N Engl J Med. 2000;343(20):1430-8. doi:10.1056/NEJM200011163432001.
34. Sormani MP, Signori A, Siri P, De Stefano N. Time to first relapse as an endpoint in multiple sclerosis clinical trials. Mult Scler. 2013;19(4):466-74. doi:10.1177/1352458512457841.
35. Freedman MS, Selchen D, Arnold DL, Prat A, Banwell B, Yeung M et al. Treatment optimization in MS: Canadian MS Working Group updated recommendations. Can J Neurol Sci. 2013;40(3):307-23.
36. Thygesen P. (Evaluation of drug treatment of disseminated sclerosis). Ugeskr Laeger. 1965;127(45):1448-50.
37. Uitdehaag BM, Barkhof F, Coyle PK, Gardner JD, Jeffery DR, Mikol DD. The changing face of multiple sclerosis clinical trial populations. Curr Med Res Opin. 2011;27(8):1529-37. doi:10.1185/03007995.2011.591370.
38. Uitdehaag BM. Clinical outcome measures in multiple sclerosis. Handb Clin Neurol. 2014;122:393-404. doi:10.1016/B978-0-444-52001-2.00016-9.
39. Confavreux C, Aimard G, Devic M. Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients. Brain. 1980;103(2):281-300.
40. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain. 1999;122 (Pt 5):871-82.
41. Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J et al. Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. Ann Neurol. 1997;42(3):379-82. doi:10.1002/ana.410420318.
42. Fischer J, Jak AJ, Kniker JE, Rudick R. Administration and scoring manual, revised. 2001.
43. Miller DM, Rudick RA, Cutter G, Baier M, Fischer JS. Clinical significance of the multiple sclerosis functional composite: relationship to patient-reported quality of life. Arch Neurol. 2000;57(9):1319-24.
44. Rudick RA, Cutter G, Baier M, Fisher E, Dougherty D, Weinstock-Guttman B et al. Use of the Multiple Sclerosis Functional Composite to predict disability in relapsing MS. Neurology. 2001;56(10):1324-30.
45. Rudick RA, Polman CH, Cohen JA, Walton MK, Miller AE, Confavreux C et al. Assessing disability progression with the Multiple Sclerosis Functional Composite. Mult Scler. 2009;15(8):984-97. doi:10.1177/1352458509106212.
46. Fisher E, Rudick RA, Cutter G, Baier M, Miller D, Weinstock-Guttman B et al. Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. Mult Scler. 2000;6(6):373-7.
47. Hofstetter L, Naegelin Y, Filli L, Kuster P, Traud S, Smieskova R et al. Progression in disability and regional grey matter atrophy in relapsing-remitting multiple sclerosis. Mult Scler. 2014;20(2):202-13. doi:10.1177/1352458513493034.

48. Kalkers NF, Bergers L, de Groot V, Lazeron RH, van Walderveen MA, Uitdehaag BM et al. Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. Neurology. 2001;56(2):215-9.

49. Cohen JA, Cutter GR, Fischer JS, Goodman AD, Heidenreich FR, Kooijmans MF et al. Benefit of interferon beta-1a on MSFC progression in secondary progressive MS. Neurology. 2002;59(5):679-87.

50. Costelloe L, O'Rourke K, McGuigan C, Walsh C, Tubridy N, Hutchinson M. The longitudinal relationship between the patient-reported Multiple Sclerosis Impact Scale and the clinician-assessed Multiple Sclerosis Functional Composite. Mult Scler. 2008;14(2):255-8. doi:10.1177/1352458507081274.

51. Ozakbas S, Cagiran I, Ormeci B, Idiman E. Correlations between multiple sclerosis functional composite, expanded disability status scale and health-related quality of life during and after treatment of relapses in patients with multiple sclerosis. J Neurol Sci. 2004;218(1-2):3-7. doi:10.1016/j.jns.2003.09.015.

52. Honarmand K, Akbar N, Kou N, Feinstein A. Predicting employment status in multiple sclerosis patients: the utility of the MS functional composite. J Neurol. 2011;258(2):244-9. doi:10.1007/s00415-010-5736-8.

53. Shawaryn MA, Schultheis MT, Garay E, Deluca J. Assessing functional status: exploring the relationship between the multiple sclerosis functional composite and driving. Arch Phys Med Rehabil. 2002;83(8):1123-9.

54. Kempen JC, de Groot V, Knol DL, Polman CH, Lankhorst GJ, Beckerman H. Community walking can be assessed using a 10-metre timed walk test. Mult Scler. 2011;17(8):980-90. doi:10.1177/1352458511403641.

55. Goldman MD, Motl RW, Scagnelli J, Pula JH, Sosnoff JJ, Cadavid D. Clinically meaningful performance benchmarks in MS: timed 25-foot walk and the real world. Neurology. 2013;81(21):1856-63. doi:10.1212/01.wnl.0000436065.97642.d2.

56. Goldman MD, Marrie RA, Cohen JA. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. Mult Scler. 2008;14(3):383-90. doi:10.1177/1352458507082607.

57. Kragt JJ, van der Linden FA, Nielsen JM, Uitdehaag BM, Polman CH. Clinical impact of 20% worsening on Timed 25-foot Walk and 9-hole Peg Test in multiple sclerosis. Mult Scler. 2006;12(5):594-8.

58. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills. 1977;44(2):367-73. doi:10.2466/pms.1977.44.2.367.

59. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. Lancet Neurol. 2008;7(12):1139-51. doi:10.1016/S1474-4422(08)70259-X.

60. Nagels G, Geentjens L, Kos D, Vleugels L, D'Hooghe M B, Van Asch P et al. Paced visual serial addition test in multiple sclerosis. Clin Neurol Neurosurg. 2005;107(3):218-22. doi:10.1016/j.clineuro.2004.11.016.

61. Cohen JA, Cutter GR, Fischer JS, Goodman AD, Heidenreich FR, Jak AJ et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch Neurol. 2001;58(6):961-7.

62. Benedict RH, Duquin JA, Jurgensen S, Rudick RA, Feitcher J, Munschauer FE et al. Repeated assessment of neuropsychological deficits in multiple sclerosis using the Symbol Digit Modalities Test and the MS Neuropsychological Screening Questionnaire. Mult Scler. 2008;14(7):940-6. doi:10.1177/1352458508090923.

63. Ontaneda D, LaRocca N, Coetzee T, Rudick R, Force NMT. Revisiting the multiple sclerosis functional composite: proceedings from the National Multiple Sclerosis Society (NMSS) Task Force on Clinical Disability Measures. Mult Scler. 2012;18(8):1074-80. doi:10.1177/1352458512451512.

64. Walker LA, Cheng A, Berard J, Berrigan LI, Rees LM, Freedman MS. Tests of information processing speed: what do people with multiple sclerosis think about them? Int J MS Care. 2012;14(2):92-9. doi:10.7224/1537-2073-14.2.92.

65. Benedict RH. Effects of using same- versus alternate-form memory tests during short-interval repeated assessments in multiple sclerosis. J Int Neuropsychol Soc. 2005;11(6):727-36. doi:10.1017/S1355617705050782.

66. Sonder JM, Burggraaff J, Knol DL, Polman CH, Uitdehaag BM. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. Mult Scler. 2014;20(4):481-8. doi:10.1177/1352458513501570.

67. Papadopoulou A, Muller-Lenke N, Naegelin Y, Kalt G, Bendfeldt K, Kuster P et al. Contribution of cortical and white matter lesions to cognitive impairment in multiple sclerosis. Mult Scler. 2013;19(10):1290-6. doi:10.1177/1352458513475490.

68. Yu HJ, Christodoulou C, Bhise V, Greenblatt D, Patel Y, Serafin D et al. Multiple white matter tract abnormalities underlie cognitive impairment in RRMS. Neuroimage. 2012;59(4):3713-22. doi:10.1016/j.neuroimage.2011.10.053.

69. Benedict RH, Cookfair D, Gavett R, Gunther M, Munschauer F, Garg N et al. Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J Int Neuropsychol Soc. 2006;12(4):549-58.

70. Portaccio E, Goretti B, Zipoli V, Iudice A, Pina DD, Malentacchi GM et al. Reliability, practice effects, and change indices for Rao's Brief Repeatable Battery. Mult Scler. 2010;16(5):611-7. doi:10.1177/1352458510362818.

71. Benedict RH, Bruce JM, Dwyer MG, Abdelrahman N, Hussein S, Weinstock-Guttman B et al. Neocortical atrophy, third ventricular width, and cognitive dysfunction in multiple sclerosis. Arch Neurol. 2006;63(9):1301-6. doi:10.1001/archneur.63.9.1301.

72. Benedict RH, Weinstock-Guttman B, Fishman I, Sharma J, Tjoa CW, Bakshi R. Prediction of neuropsychological impairment in multiple sclerosis: comparison of conventional magnetic resonance imaging measures of atrophy and lesion burden. Arch Neurol. 2004;61(2):226-30. doi:10.1001/archneur.61.2.226.

73. Balcer LJ, Miller DH, Reingold SC, Cohen JA. Vision and vision-related outcome measures in multiple sclerosis. Brain. 2015;138(Pt 1):11-27. doi:10.1093/brain/awu335.

74. Balcer LJ, Baier ML, Cohen JA, Kooijmans MF, Sandrock AW, Nano-Schiavi ML et al. Contrast letter acuity as a visual component for the Multiple Sclerosis Functional Composite. Neurology. 2003;61(10):1367-73.

75. Wu GF, Schwartz ED, Lei T, Souza A, Mishra S, Jacobs DA et al. Relation of vision to global and regional brain MRI in multiple sclerosis. Neurology. 2007;69(23):2128-35. doi:10.1212/01.wnl.0000278387.15090.5a.

76. Balcer LJ, Frohman EM. Evaluating loss of visual function in multiple sclerosis as measured by low-contrast letter acuity. Neurology. 2010;74 Suppl 3:S16-23. doi:10.1212/WNL.0b013e3181dbb664.

77. Balcer LJ, Galetta SL, Calabresi PA, Confavreux C, Giovannoni G, Havrdova E et al. Natalizumab reduces visual loss in patients with relapsing multiple sclerosis. Neurology. 2007;68(16):1299-304. doi:10.1212/01.wnl.0000259521.14704.a8.

78. Uitdehaag BM, Ader HJ, Roosma TJ, de Groot V, Kalkers NF, Polman CH. Multiple sclerosis functional composite: impact of reference population and interpretation of changes. Mult Scler. 2002;8(5):366-71.

79. Fox RJ, Lee JC, Rudick RA. Optimal reference population for the multiple sclerosis functional composite. Mult Scler. 2007;13(7):909-14. doi:10.1177/1352458507076950.

80. Schwid SR, Goodman AD, McDermott MP, Bever CF, Cook SD. Quantitative functional measures in MS: what is a reliable change? Neurology. 2002;58(8):1294-6.

81. Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. http://wwwfdagov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282pdf. December 2009.

82. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry. 1961;4:561-71.

83. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Med Care. 1981;19(8):787-805.

84. Cella DF, Dineen K, Arnason B, Reder A, Webster KA, karabatsos G et al. Validation of the functional assessment of multiple sclerosis quality of life instrument. Neurology. 1996;47(1):129-39.

85. Ferrans CE, Powers MJ. Quality of life index: development and psychometric properties. ANS Adv Nurs Sci. 1985;8(1):15-24.

86. Fischer JS, LaRocca NG, Miller DM, Ritvo PG, Andrews H, Paty D. Recent developments in the assessment of quality of life in multiple sclerosis (MS). Mult Scler. 1999;5(4):251-9.

87. Fisk JD, Brown MG, Sketris IS, Metz LM, Murray TJ, Stadnyk KJ. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. J Neurol Neurosurg Psychiatry. 2005;76(1):58-63. doi:10.1136/jnnp.2003.017897.

88. Fisk JD, Doble SE. Construction and validation of a fatigue impact scale for daily administration (D-FIS). Qual Life Res. 2002;11(3):263-72.

89. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Can J Neurol Sci. 1994;21(1):9-14.

90. Ford HL, Gerry E, Tennant A, Whalley D, Haigh R, Johnson MH. Developing a disease-specific quality of life measure for people with multiple sclerosis. Clin Rehabil. 2001;15(3):247-58.

91. Gold SM, Heesen C, Schulz H, Guder U, Monch A, Gbadamosi J et al. Disease specific quality of life instruments in multiple sclerosis: validation of the Hamburg Quality of Life Questionnaire in Multiple Sclerosis (HAQUAMS). Mult Scler. 2001;7(2):119-30.

92. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain. 2001;124(Pt 5):962-73.

93. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). Neurology. 2003;60(1):31-6.

94. Honarmand K, Feinstein A. Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. Mult Scler. 2009;15(12):1518-24. doi:10.1177/1352458509347150.

95. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-13.

96. Melin R, Fugl-Meyer KS, Fugl-Meyer AR. Life satisfaction in 18- to 64-year-old Swedes: in relation to education, employment situation, health and physical activity. J Rehabil Med. 2003;35(2):84-90.

97. Mokkink LB, Knol DL, Uitdehaag BM. Factor structure of Guy's Neurological Disability Scale in a sample of Dutch patients with multiple sclerosis. Mult Scler. 2011;17(12):1498-503. doi:10.1177/1352458511413098.

98. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BM. The Arm Function in Multiple Sclerosis Questionnaire (AMSQ): development and validation of a new tool using IRT methods. Disabil Rehabil. 2015:1-7. doi:10.3109/09638288.2015.1027005.

99. Mowry EM, Loguidice MJ, Daniels AB, Jacobs DA, Markowitz CE, Galetta SL et al. Vision related quality of life in multiple sclerosis: correlation with new measures of low and high contrast letter acuity. J Neurol Neurosurg Psychiatry. 2009;80(7):767-72. doi:10.1136/jnnp.2008.165449.

100. Ozakbas S, Akdede BB, Kosehasanogullari G, Aksan O, Idiman E. Difference between generic and multiple sclerosis-specific quality of life instruments regarding the assessment of treatment efficacy. J Neurol Sci. 2007;256(1-2):30-4. doi:10.1016/j.jns.2007.01.080.

101. Pfennings L, Cohen L, Ader H, Polman C, Lankhorst G, Smits R et al. Exploring differences between subgroups of multiple sclerosis patients in health-related quality of life. J Neurol. 1999;246(7):587-91.

102. Simeoni M, Auquier P, Fernandez O, Flachenecker P, Stecchi S, Constantinescu C et al. Validation of the Multiple Sclerosis International Quality of Life questionnaire. Mult Scler. 2008;14(2):219-30. doi:10.1177/1352458507080733.

103. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Qual Life Res. 1995;4(3):187-206.

104. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.

105. Wynia K, Middel B, van Dijk JP, de Ruiter H, de Keyser J, Reijneveld SA. The Multiple Sclerosis impact Profile (MSIP). Development and testing psychometric properties of an ICF-based health measure. Disabil Rehabil. 2008;30(4):261-74. doi:10.1080/09638280701256868.

106. Cohen RA, Kessler HR, Fischer M. The Extended Disability Status Scale (EDSS) as a predictor of impairments of functional activities of daily living in multiple sclerosis. J Neurol Sci. 1993;115(2):132-5.

107. Farrell B, Godwin J, Richards S, Warlow C. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. J Neurol Neurosurg Psychiatry. 1991;54(12):1044-54.

108. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. Scott Med J. 1957;2(5):200-15.

109. Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. Md State Med J. 1965;14:61-5.

110. Airlie J, Baker GA, Smith SJ, Young CA. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. Clin Rehabil. 2001;15(3):259-65.

111. Smith SJ, Young CA. The role of affect on the perception of disability in multiple sclerosis. Clin Rehabil. 2000;14(1):50-4.

112. Berrigan LI, Fisk JD, Patten SB, Tremlett H, Wolfson C, Warren S et al. Health-related quality of life in multiple sclerosis: Direct and indirect effects of comorbidity. Neurology. 2016. doi:10.1212/WNL.0000000000002564.

113. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med. 1999;48(11):1507-15.

114. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther. 2014;36(5):648-62. doi:10.1016/j.clinthera.2014.04.006.

115. van Munster ETL. Klinische en paraklinische parameters in het verloop van multiple sclerose. Thesis. 1991.

116. Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kappos L et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol. 2016;15(3):292-303. doi:10.1016/S1474-4422(15)00393-2.

117. McDonald WI, Miller DH, Thompson AJ. Are magnetic resonance findings predictive of clinical outcome in therapeutic trials in multiple sclerosis? The dilemma of interferon-beta. Ann Neurol. 1994;36(1):14-8. doi:10.1002/ana.410360106.

118. Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain. 2008;131(Pt 3):808-17. doi:10.1093/brain/awm329.

119. Brex PA, Ciccarelli O, O'Riordan JI, Sailer M, Thompson AJ, Miller DH. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. N Engl J Med. 2002;346(3):158-64. doi:10.1056/NEJMoa011341.

120. O'Riordan JI, Gawne Cain M, Coles A, Wang L, Compston DA, Tofts P et al. T1 hypointense lesion load in secondary progressive multiple sclerosis: a comparison of pre versus post contrast loads and of manual versus semi automated threshold techniques for lesion segmentation. Mult Scler. 1998;4(5):408-12.

121. Tintore M, Rovira A, Rio J, Otero-Romero S, Arrambide G, Tur C et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. Brain. 2015;138(Pt 7):1863-74. doi:10.1093/brain/awv105.

122. Bermel RA, You X, Foulds P, Hyde R, Simon JH, Fisher E et al. Predictors of long-term outcome in multiple sclerosis patients treated with interferon beta. Ann Neurol. 2013;73(1):95-103. doi:10.1002/ana.23758.

123. Sormani MP, Bonzano L, Roccatagliata L, Cutter GR, Mancardi GL, Bruzzi P. Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. Ann Neurol. 2009;65(3):268-75. doi:10.1002/ana.21606.

124. Dobson R, Rudick RA, Turner B, Schmierer K, Giovannoni G. Assessing treatment response to interferon-beta: is there a role for MRI? Neurology. 2014;82(3):248-54. doi:10.1212/WNL.0000000000000036.

125. Sormani MP, Bonzano L, Roccatagliata L, Mancardi GL, Uccelli A, Bruzzi P. Surrogate endpoints for EDSS worsening in multiple sclerosis. A meta-analytic approach. Neurology. 2010;75(4):302-9. doi:10.1212/WNL.0b013e3181ea15aa.

126. Sormani MP, Stubinski B, Cornelisse P, Rocak S, Li D, De Stefano N. Magnetic resonance active lesions as individual-level surrogate for relapses in multiple sclerosis. Mult Scler. 2011;17(5):541-9. doi:10.1177/1352458510391837.

127. Geurts JJ, Barkhof F. Grey matter pathology in multiple sclerosis. Lancet Neurol. 2008;7(9):841-51. doi:10.1016/S1474-4422(08)70191-1.

128. Kalkers NF, Ameziane N, Bot JC, Minneboo A, Polman CH, Barkhof F. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. Arch Neurol. 2002;59(10):1572-6.

129. Miller DH, Barkhof F, Frank JA, Parker GJ, Thompson AJ. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. Brain. 2002;125(Pt 8):1676-95.

130. Henry RG, Shieh M, Okuda DT, Evangelista A, Gorno-Tempini ML, Pelletier D. Regional grey matter atrophy in clinically isolated syndromes at presentation. J Neurol Neurosurg Psychiatry. 2008;79(11):1236-44. doi:10.1136/jnnp.2007.134825.

131. Fisher E, Lee JC, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: a longitudinal study. Ann Neurol. 2008;64(3):255-65. doi:10.1002/ana.21436.

132. van Munster CE, Jonkman LE, Weinstein HC, Uitdehaag BM, Geurts JJ. Gray matter damage in multiple sclerosis: Impact on clinical symptoms. Neuroscience. 2015;303:446-61. doi:10.1016/j.neuroscience.2015.07.006.

133. Simon JH, Jacobs LD, Campion MK, Rudick RA, Cookfair DL, Herndon RM et al. A longitudinal study of brain atrophy in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Neurology. 1999;53(1):139-48.

134. Ashburner J, Friston KJ. Voxel-based morphometry--the methods. Neuroimage. 2000;11(6 Pt 1):805-21. doi:10.1006/nimg.2000.0582.

135. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000;97(20):11050-5. doi:10.1073/pnas.200033797.

136. Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage. 2002;17(1):479-89.

137. Fisniku LK, Chard DT, Jackson JS, Anderson VM, Altmann DR, Miszkiel KA et al. Gray matter atrophy is related to long-term disability in multiple sclerosis. Ann Neurol. 2008;64(3):247-54. doi:10.1002/ana.21423.

138. Anderson VM, Fisniku LK, Altmann DR, Thompson AJ, Miller DH. MRI measures show significant cerebellar gray matter volume loss in multiple sclerosis and are associated with cerebellar dysfunction. Mult Scler. 2009;15(7):811-7. doi:10.1177/1352458508101934.

139. Lycklama G, Thompson A, Filippi M, Miller D, Polman C, Fazekas F et al. Spinal-cord MRI in multiple sclerosis. Lancet Neurol. 2003;2(9):555-62.

140. Sicotte NL, Kern KC, Giesser BS, Arshanapalli A, Schultz A, Montag M et al. Regional hippocampal atrophy in multiple sclerosis. Brain. 2008;131(Pt 4):1134-41. doi:10.1093/brain/awn030.

141. Houtchens MK, Benedict RH, Killiany R, Sharma J, Jaisani Z, Singh B et al. Thalamic atrophy and cognition in multiple sclerosis. Neurology. 2007;69(12):1213-23. doi:10.1212/01.wnl.0000276992.17011.b5.

142. Calabrese M, Filippi M, Rovaris M, Bernardi V, Atzori M, Mattisi I et al. Evidence for relative cortical sparing in benign multiple sclerosis: a longitudinal magnetic resonance imaging study. Mult Scler. 2009;15(1):36-41. doi:10.1177/1352458508096686.

143. Rocca MA, Mesaros S, Pagani E, Sormani MP, Comi G, Filippi M. Thalamic damage and long-term progression of disability in multiple sclerosis. Radiology. 2010;257(2):463-9. doi:10.1148/radiol.10100326.

144. Valsasina P, Rocca MA, Horsfield MA, Absinta M, Messina R, Caputo D et al. Regional cervical cord atrophy and disability in multiple sclerosis: a voxel-based analysis. Radiology. 2013;266(3):853-61. doi:10.1148/radiol.12120813.

145. Lukas C, Sombekke MH, Bellenberg B, Hahn HK, Popescu V, Bendfeldt K et al. Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. Radiology. 2013;269(2):542-52. doi:10.1148/radiol.13122566.

146. Rocca MA, Valsasina P, Damjanovic D, Horsfield MA, Mesaros S, Stosic-Opincal T et al. Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. J Neurol Neurosurg Psychiatry. 2013;84(1):35-41. doi:10.1136/jnnp-2012-303821.

147. Kearney H, Rocca MA, Valsasina P, Balk L, Sastre-Garriga J, Reinhardt J et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. Mult Scler. 2014;20(1):72-80. doi:10.1177/1352458513492245.

148. De Stefano N, Airas L, Grigoriadis N, Mattle HP, O'Riordan J, Oreja-Guevara C et al. Clinical relevance of brain volume measures in multiple sclerosis. CNS Drugs. 2014;28(2):147-56. doi:10.1007/s40263-014-0140-z.

149. Ziemssen T, Derfuss T, de Stefano N, Giovannoni G, Palavra F, Tomic D et al. Optimizing treatment success in multiple sclerosis. J Neurol. 2016;263(6):1053-65. doi:10.1007/s00415-015-7986-y.

150. Sormani MP, Arnold DL, De Stefano N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. Ann Neurol. 2014;75(1):43-9. doi:10.1002/ana.24018.

151. Tsivgoulis G, Katsanos AH, Grigoriadis N, Hadjigeorgiou GM, Heliopoulos I, Kilidireas C et al. The effect of disease modifying therapies on brain atrophy in patients with relapsing-remitting multiple sclerosis: a systematic review and meta-analysis. PLoS One. 2015;10(3):e0116511. doi:10.1371/journal.pone.0116511.

152 Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. Lancet Neurol. 2006;5(2):158-70. doi:10.1016/S1474-4422(06)70349-0.

153. Zivadinov R, Reder AT, Filippi M, Minagar A, Stuve O, Lassmann H et al. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. Neurology. 2008;71(2):136-44. doi:10.1212/01.wnl.0000316810.01120.05.

154. Giorgio A, Battaglini M, Smith SM, De Stefano N. Brain atrophy assessment in multiple sclerosis: importance and limitations. Neuroimaging Clin N Am. 2008;18(4):675-86, xi. doi:10.1016/j.nic.2008.06.007.

155. van Waesberghe JH, van Walderveen MA, Castelijns JA, Scheltens P, Lycklama a Nijeholt GJ, Polman CH et al. Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin-echo and magnetization transfer MR. AJNR Am J Neuroradiol. 1998;19(4):675-83.

156. van Walderveen MA, Kamphorst W, Scheltens P, van Waesberghe JH, Ravid R, Valk J et al. Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. Neurology. 1998;50(5):1282-8.

157. van Walderveen MA, Barkhof F, Hommes OR, Polman CH, Tobi H, Frequin ST et al. Correlating MRI and clinical disease activity in multiple sclerosis: relevance of hypointense lesions on short-TR/short-TE (T1-weighted) spin-echo images. Neurology. 1995;45(9):1684-90.

158. Truyen L, van Waesberghe JH, van Walderveen MA, van Oosten BW, Polman CH, Hommes OR et al. Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. Neurology. 1996;47(6):1469-76.

159. Giorgio A, Stromillo ML, Bartolozzi ML, Rossi F, Battaglini M, De Leucio A et al. Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis. Mult Scler. 2014;20(2):214-9. doi:10.1177/1352458513494490.

160. Filippi M, Rovaris M, Rocca MA, Sormani MP, Wolinsky JS, Comi G et al. Glatiramer acetate reduces the proportion of new MS lesions evolving into "black holes". Neurology. 2001;57(4):731-3.

161. Filippi M, Rocca MA, Pagani E, De Stefano N, Jeffery D, Kappos L et al. Placebo-controlled trial of oral laquinimod in multiple sclerosis: MRI evidence of an effect on brain tissue damage. J Neurol Neurosurg Psychiatry. 2014;85(8):851-8. doi:10.1136/jnnp-2013-306132.

162. Dalton CM, Miszkiel KA, Barker GJ, MacManus DG, Pepple TI, Panzara M et al. Effect of natalizumab on conversion of gadolinium enhancing lesions to T1 hypointense lesions in relapsing multiple sclerosis. J Neurol. 2004;251(4):407-13. doi:10.1007/s00415-004-0332-4.

163. Barkhof F, van Waesberghe JH, Filippi M, Yousry T, Miller DH, Hahn D et al. T(1) hypointense lesions in secondary progressive multiple sclerosis: effect of interferon beta-1b treatment. Brain. 2001;124(Pt 7):1396-402.

164. Schoonheim MM, Meijer KA, Geurts JJ. Network collapse and cognitive impairment in multiple sclerosis. Front Neurol. 2015;6:82. doi:10.3389/fneur.2015.00082.

165. Rovaris M, Gass A, Bammer R, Hickman SJ, Ciccarelli O, Miller DH et al. Diffusion MRI in multiple sclerosis. Neurology. 2005;65(10):1526-32. doi:10.1212/01.wnl.0000184471.83948.e0.

166. Tortorella C, Viti B, Bozzali M, Sormani MP, Rizzo G, Gilardi MF et al. A magnetization transfer histogram study of normal-appearing brain tissue in MS. Neurology. 2000;54(1):186-93.

167. Schmierer K, Tozer DJ, Scaravilli F, Altmann DR, Barker GJ, Tofts PS et al. Quantitative magnetization transfer imaging in postmortem multiple sclerosis brain. J Magn Reson Imaging. 2007;26(1):41-51. doi:10.1002/jmri.20984.

168. Gabilondo I, Martinez-Lapiscina EH, Martinez-Heras E, Fraga-Pumar E, Llufriu S, Ortiz S et al. Trans-synaptic axonal degeneration in the visual pathway in multiple sclerosis. Ann Neurol. 2014;75(1):98-107. doi:10.1002/ana.24030.

169. Saidha S, Syc SB, Ibrahim MA, Eckstein C, Warner CV, Farrell SK et al. Primary retinal pathology in multiple sclerosis as detected by optical coherence tomography. Brain. 2011;134(Pt 2):518-33. doi:10.1093/brain/awq346.

170. Green AJ, McQuaid S, Hauser SL, Allen IV, Lyness R. Ocular pathology in multiple sclerosis: retinal atrophy and inflammation irrespective of disease duration. Brain. 2010;133(Pt 6):1591-601. doi:10.1093/brain/awq080.

171. Talman LS, Bisker ER, Sackel DJ, Long DA, Jr., Galetta KM, Ratchford JN et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. Ann Neurol. 2010;67(6):749-60. doi:10.1002/ana.22005.

172. Petzold A, de Boer JF, Schippling S, Vermersch P, Kardon R, Green A et al. Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. Lancet Neurol. 2010;9(9):921-32. doi:10.1016/S1474-4422(10)70168-X.

173. Herrero R, Garcia-Martin E, Almarcegui C, Ara JR, Rodriguez-Mena D, Martin J et al. Progressive degeneration of the retinal nerve fiber layer in patients with multiple sclerosis. Invest Ophthalmol Vis Sci. 2012;53(13):8344-9. doi:10.1167/iovs.12-10362.

174. Siger M, Dziegielewski K, Jasek L, Bieniek M, Nicpan A, Nawrocki J et al. Optical coherence tomography in multiple sclerosis: thickness of the retinal nerve fiber layer as a potential measure of axonal loss and brain atrophy. J Neurol. 2008;255(10):1555-60. doi:10.1007/s00415-008-0985-5.

175. Saidha S, Sotirchos ES, Oh J, Syc SB, Seigo MA, Shiee N et al. Relationships between retinal axonal and neuronal measures and global central nervous system pathology in multiple sclerosis. JAMA Neurol. 2013;70(1):34-43. doi:10.1001/jamaneurol.2013.573.

176. Trip SA, Schlottmann PG, Jones SJ, Altmann DR, Garway-Heath DF, Thompson AJ et al. Retinal nerve fiber layer axonal loss and visual dysfunction in optic neuritis. Ann Neurol. 2005;58(3):383-91. doi:10.1002/ana.20575.

177. Parisi V, Manni G, Spadaro M, Colacino G, Restuccia R, Marchi S et al. Correlation between morphological and functional retinal impairment in multiple sclerosis patients. Invest Ophthalmol Vis Sci. 1999;40(11):2520-7.

178. Henderson AP, Altmann DR, Trip AS, Kallis C, Jones SJ, Schlottmann PG et al. A serial study of retinal changes following optic neuritis with sample size estimates for acute neuroprotection trials. Brain. 2010;133(9):2592-602. doi:10.1093/brain/awq146.

179. Sakai RE, Feller DJ, Galetta KM, Galetta SL, Balcer LJ. Vision in multiple sclerosis: the story, structure-function correlations, and models for neuroprotection. J Neuroophthalmol. 2011;31(4):362-73. doi:10.1097/WNO.0b013e318238937f.

180. Naismith RT, Tutlam NT, Xu J, Shepherd JB, Klawiter EC, Song SK et al. Optical coherence tomography is less sensitive than visual evoked potentials in optic neuritis. Neurology. 2009;73(1):46-52. doi:10.1212/WNL.0b013e3181aaea32.

181. Jeanjean L, Castelnovo G, Carlander B, Villain M, Mura F, Dupeyron G et al. (Retinal atrophy using optical coherence tomography (OCT) in 15 patients with multiple sclerosis and comparison with healthy subjects). Rev Neurol (Paris). 2008;164(11):927-34. doi:10.1016/j.neurol.2008.03.008.

182. Martinez-Lapiscina EH, Arnow S, Wilson JA, Saidha S, Preiningerova JL, Oberwahrenbrock T et al. Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study. Lancet Neurol. 2016;15(6):574-84. doi:10.1016/S1474-4422(16)00068-5.

183. Burkholder BM, Osborne B, Loguidice MJ, Bisker E, Frohman TC, Conger A et al. Macular volume determined by optical coherence tomography as a measure of neuronal loss in multiple sclerosis. Arch Neurol. 2009;66(11):1366-72. doi:10.1001/archneurol.2009.230.

184. Walter SD, Ishikawa H, Galetta KM, Sakai RE, Feller DJ, Henderson SB et al. Ganglion cell loss in relation to visual disability in multiple sclerosis. Ophthalmology. 2012;119(6):1250-7. doi:10.1016/j.ophtha.2011.11.032.

185. Oberwahrenbrock T, Ringelstein M, Jentschke S, Deuschle K, Klumbies K, Bellmann-Strobl J et al. Retinal ganglion cell and inner plexiform layer thinning in clinically isolated syndrome. Mult Scler. 2013;19(14):1887-95. doi:10.1177/1352458513489757.

186. Pierro L, Gagliardi M, Iuliano L, Ambrosi A, Bandello F. Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. Invest Ophthalmol Vis Sci. 2012;53(9):5912-20. doi:10.1167/iovs.11-8644.

187. Teunissen CE, Malekzadeh A, Leurs C, Bridel C, Killestein J. Body fluid biomarkers for multiple sclerosis--the long road to clinical application. Nat Rev Neurol. 2015;11(10):585-96. doi:10.1038/nrneurol.2015.173.

188. Petzold A, Tisdall MM, Girbes AR, Martinian L, Thom M, Kitchen N et al. In vivo monitoring of neuronal loss in traumatic brain injury: a microdialysis study. Brain. 2011;134(Pt 2):464-83. doi:10.1093/brain/awq360.

189. Malmestrom C, Haghighi S, Rosengren L, Andersen O, Lycke J. Neurofilament light protein and glial fibrillary acidic protein as biological markers in MS. Neurology. 2003;61(12):1720-5.

190. Kuhle J, Plattner K, Bestwick JP, Lindberg RL, Ramagopalan SV, Norgren N et al. A comparative study of CSF neurofilament light and heavy chain protein in MS. Mult Scler. 2013;19(12):1597-603. doi:10.1177/1352458513482374.

191. Petzold A. The prognostic value of CSF neurofilaments in multiple sclerosis at 15-year follow-up. J Neurol Neurosurg Psychiatry. 2015;86(12):1388-90. doi:10.1136/jnnp-2014-309827.

192. Burman J, Zetterberg H, Fransson M, Loskog AS, Raininko R, Fagius J. Assessing tissue damage in multiple sclerosis: a biomarker approach. Acta Neurol Scand. 2014;130(2):81-9. doi:10.1111/ane.12239.

193. Petzold A, Steenwijk MD, Eikelenboom JM, Wattjes MP, Uitdehaag BM. Elevated CSF neurofilament proteins predict brain atrophy: A 15-year follow-up study. Mult Scler. 2016;22(9):1154-62. doi:10.1177/1352458516645206.

194. Kuhle J, Disanto G, Lorscheider J, Stites T, Chen Y, Dahlke F et al. Fingolimod and CSF neurofilament light chain levels in relapsing-remitting multiple sclerosis. Neurology. 2015;84(16):1639-43. doi:10.1212/WNL.0000000000001491.

195. Gunnarsson M, Malmestrom C, Axelsson M, Sundstrom P, Dahle C, Vrethem M et al. Axonal damage in relapsing multiple sclerosis is markedly reduced by natalizumab. Ann Neurol. 2011;69(1):83-9. doi:10.1002/ana.22247.

196. Doussau F, Augustine GJ. The actin cytoskeleton and neurotransmitter release: an overview. Biochimie. 2000;82(4):353-63.

197. Semra YK, Seidi OA, Sharief MK. Heightened intrathecal release of axonal cytoskeletal proteins in multiple sclerosis is associated with progressive disease and clinical disability. J Neuroimmunol. 2002;122(1-2):132-9.

198. Downing KH. Structural basis for the interaction of tubulin with proteins and drugs that affect microtubule dynamics. Annu Rev Cell Dev Biol. 2000;16:89-111. doi:10.1146/annurev.cellbio.16.1.89.

199. Checa A, Khademi M, Sar DG, Haeggstrom JZ, Lundberg JO, Piehl F et al. Hexosylceramides as intrathecal markers of worsening disability in multiple sclerosis. Mult Scler. 2015;21(10):1271-9. doi:10.1177/1352458514561908.

200. Petzold A, Eikelenboom MJ, Gveric D, Keir G, Chapman M, Lazeron RH et al. Markers for different glial cell responses in multiple sclerosis: clinical and pathological correlations. Brain. 2002;125(Pt 7):1462-73.

201. Hinsinger G, Galeotti N, Nabholz N, Urbach S, Rigau V, Demattei C et al. Chitinase 3-like proteins as diagnostic and prognostic biomarkers of multiple sclerosis. Mult Scler. 2015;21(10):1251-61. doi:10.1177/1352458514561906.

202. Kuhle J, Gaiottino J, Leppert D, Petzold A, Bestwick JP, Malaspina A et al. Serum neurofilament light chain is a biomarker of human spinal cord injury severity and outcome. J Neurol Neurosurg Psychiatry. 2015;86(3):273-9. doi:10.1136/jnnp-2013-307454.

203. Disanto G, Adiutori R, Dobson R, Martinelli V, Dalla Costa G, Runia T et al. Serum neurofilament light chain levels are increased in patients with a clinically isolated syndrome. J Neurol Neurosurg Psychiatry. 2016;87(2):126-9. doi:10.1136/jnnp-2014-309690.

204. van der Voort LF, Vennegoor A, Visser A, Knol DL, Uitdehaag BM, Barkhof F et al. Spontaneous MxA mRNA level predicts relapses in patients with recently diagnosed MS. Neurology. 2010;75(14):1228-33. doi:10.1212/WNL.0b013e3181f6c556.

205. van der Voort LF, Visser A, Knol DL, Oudejans CB, Polman CH, Killestein J. Lack of interferon-beta bioactivity is associated with the occurrence of relapses in multiple sclerosis. Eur J Neurol. 2009;16(9):1049-52. doi:10.1111/j.1468-1331.2009.02649.x.

206. Lublin FD. Disease activity free status in MS. Mult Scler Relat Disord. 2012;1(1):6-7. doi:10.1016/j.msard.2011.08.001.

207. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort. JAMA Neurol. 2015;72(2):152-8. doi:10.1001/jamaneurol.2014.3537.

208. Sormani MP, De Stefano N. Defining and scoring response to IFN-beta in multiple sclerosis. Nat Rev Neurol. 2013;9(9):504-12. doi:10.1038/nrneurol.2013.146.

209. Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2013;84(10):1082-91. doi:10.1136/jnnp-2012-304094.

210. Filippi M, Preziosa P, Copetti M, Riccitelli G, Horsfield MA, Martinelli V et al. Gray matter damage predicts the accumulation of disability 13 years later in MS. Neurology. 2013;81(20):1759-67. doi:10.1212/01.wnl.0000435551.90824.d0.

211. Sormani MP, Rio J, Tintore M, Signori A, Li D, Cornelisse P et al. Scoring treatment response in patients with relapsing multiple sclerosis. Mult Scler. 2013;19(5):605-12. doi:10.1177/1352458512460605.

212. Havrdova E, Galetta S, Hutchinson M, Stefoski D, Bates D, Polman CH et al. Effect of natalizumab on clinical and radiological disease activity in multiple sclerosis: a retrospective analysis of the Natalizumab Safety and Efficacy in Relapsing-Remitting Multiple Sclerosis (AFFIRM) study. Lancet Neurol. 2009;8(3):254-60. doi:10.1016/S1474-4422(09)70021-3.

213. Menge T, Stuve O, Kieseier BC, Hartung HP. Alemtuzumab: the advantages and challenges of a novel therapy in MS. Neurology. 2014;83(1):87-97. doi:10.1212/WNL.0000000000000540.

214. Kappos L, De Stefano N, Freedman MS, Cree BA, Radue EW, Sprenger T et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. Mult Scler. 2016;22(10):1297-305. doi:10.1177/1352458515616701.

215. D`Souza M, Burggraaff J, Steinheimer S, Kontschieder P, Morrison C, Dorn J et al. Accuracy of depth-sensing recordings in classifying Expanded Disability Status Scale subscores of motor dysfunction in patients with multiple sclerosis. American Academy of Neurology (AAN) poster presentation. 2016.

216. Bonzano L, Sormani MP, Tacchino A, Abate L, Mancardi GL, Uccelli A. Validation of a new quantitative and objective tool for the assessment of hand motor disability in multiple sclerosis. Mult Scler.17 (suppl 10):S43.

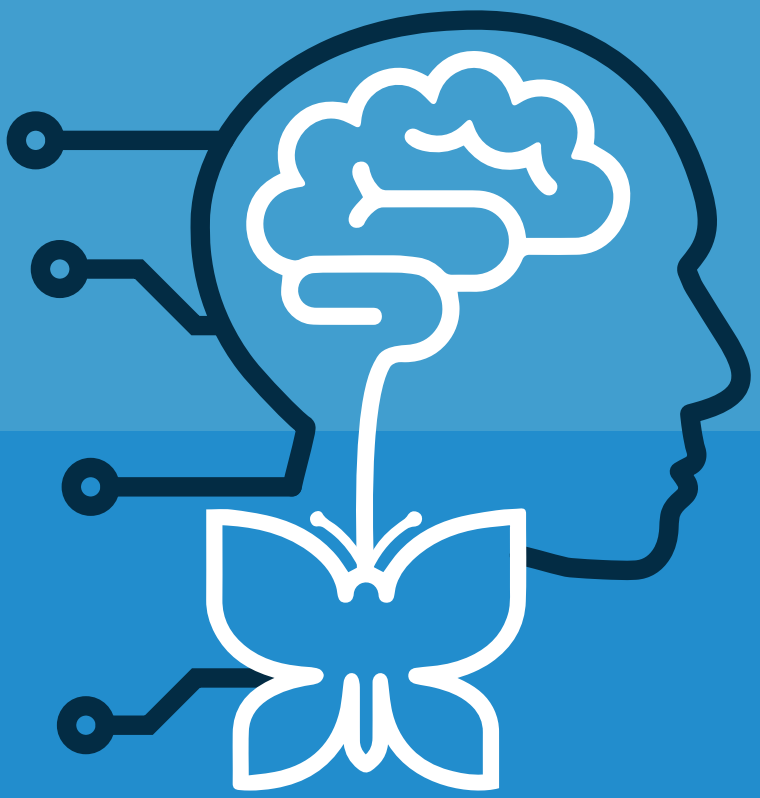
217. Weikert M, Suh Y, Lane A, Sandroff B, Dlugonski D, Fernhall B et al. Accelerometry is associated with walking mobility, not physical activity, in persons with multiple sclerosis. Med Eng Phys. 2012;34(5):590-7. doi:10.1016/j.medengphy.2011.09.005.

218. Busby B, Lesko M, August, January Hackathon p, Federer L. Closing gaps between open software and public data in a hackathon setting: User-centered software prototyping. F1000Res. 2016;5:672. doi:10.12688/f1000research.8382.2.

219. Walker A, Ko N. Bringing Medicine to the Digital Age via Hackathons and Beyond. J Med Syst. 2016;40(4):98. doi:10.1007/s10916-016-0461-1.

PART 2

Clinical assessment of upper extremity function and mobility



CHAPTER 3

Measuring various aspects of upper extremity function independent from ambulation in multiple sclerosis enhances disability assessment

Caspar E.P. van Munster, Jessica Burggraaf, Saskia Steinheimer, Christian P. Kamm, Marcus d'Souza, Jonas Dorn, Jacques Boisvert, Manuela Diederich, Kristina Kravalis, Frank Dahlke, Ludwig Kappos, Bernard M.J. Uitdehaag.

Submitted for publication

ABSTRACT

Background
Upper extremity function (UEF) is often compromised in multiple sclerosis (MS), although its importance is commonly under-recognized relative to ambulation. We explored the concurrent presence of UEF and ambulatory impairment in MS, by examining various aspects of UEF across different levels of ambulation.

Methods
247 patients were included with clinically definite MS or clinically isolated syndrome according to the revised 2010 McDonald's criteria. The Nine-Hole-Peg test (9HPT) and Expanded Disability Status Scale were used to stratify patients into clinically different UEF and ambulation subgroups. Multiple aspects of UEF were examined with various clinical tests and the patient-reported outcome measure 'Arm Function in MS Questionnaire' (AMSQ).

Results
Patients with worse ambulatory impairment displayed worse performance on the 9HPT, clinical UEF tests and the AMSQ. Although most patients had no obvious ambulatory restrictions, more than 80% exhibited some level of UEF impairment. Most patients had mild UEF impairment (n=174), accounting for the largest proportion in all ambulation groups (52 – 78%).

Conclusion
UEF and ambulation showed distinct patterns of impairment in MS, affecting multiple aspects of UEF. Better assessment of multiple facets of disability may be helpful in treatment decisions, and support the development of rehabilitations strategies specially targeted towards UEF impairment.

1 INTRODUCTION

Walking impairment is a common consequence of multiple sclerosis (MS),(1) which is routinely assessed as an indicator of disability progression and to monitor the efficacy of treatment.(2) Although upper extremity function (UEF) is often compromised as well, its importance is under-recognized relative to ambulation.(3) UEF impairment can impact the ability to use walking aids and is important to maintain the capacity to perform activities of daily living (ADL). Additionally, the magnitude of upper extremity (UE) dysfunction was shown to negatively impact quality of life,(4) and was a significant predictor of direct disease related costs in MS.(5) Therefore, identifying UEF impairment and characterizing its magnitude and impact is of importance for MS management.

The traditional endpoint to rate disability in MS therapeutic trials is the Expanded Disability Status Scale (EDSS), which is heavily weighted on walking ability, especially in the higher scale range.(6) The use of UEF-specific measures in clinical trials has increased since the introduction of the MS Functional Composite (MSFC), a composite of quantitative measures of UEF (Nine Hole Peg test (9HPT)), walking speed (Timed 25-foot Walk test (T25WT)) and cognitive function.(7) Although this improved the assessment of UEF, the instrument does not fully capture the broader aspects of function necessary to define the level of severity on more complex tasks, such as ADL. Furthermore, the incorporation of specific patient-reported outcome (PRO) measures would be valuable for enhancing information on the functional impact of UEF impairment, which is increasingly being recommended as integral component in clinical trials.(8)

Recent developments in the assessment and management of UEF restrictions, and in the spectrum of interventions available, brings renewed hope. A large phase 3 trial (ASCEND) explored the effect of natalizumab on disease progression in secondary progressive MS (SPMS) patients. Although treatment did not have an effect on ambulation, it reduced progression of UEF impairment as measured with the 9HPT.(9) Similar data were presented recently from the ORATORIO trial, demonstrating that ocrelizumab reduced the risk of UEF impairment progression in primary progressive MS patients. (10) The positive effect on UEF of several therapies thus indicates that patients with a more advanced stage of MS may benefit from disease modification. Furthermore, UE rehabilitation studies revealed that different types of training programmes, such as multidisciplinary and robot-based rehabilitation, specially targeted toward the upper limbs can improve UE capacity and performance in MS.(11)

Ideally, clinicians can identify subgroups of patients who are likely to benefit from disease modifying therapies and select patients that may benefit from

rehabilitation strategies, according to the level of function on UEF as well as ambulation. Therefore, the aim of this study was to explore the concurrent presence of UEF and ambulatory impairment in patients with MS, by examining various aspects of UEF across different levels of ambulation.

2 METHODS

The data reported here are part of a larger study to develop the Assess MS system, which is a multinational project performed by large European MS centres.(12-16) Assess MS is being designed to automatically quantify motor dysfunction in MS, with the goal of providing a consistent and fine-grained measure of motor ability. Movements of MS patients were recorded non-invasively with a 3D depth sensing and colour camera (Microsoft Kinect®) and analysed using machine learning algorithms. Written informed consent was obtained from all subjects prior to study participation and the study was approved by the respective ethics committees.

Subjects

In total 247 patients were included (165 females; mean age 47.3±13.0 years; median disease duration 13 (range 0–57) years; median EDSS 3.5 (range 0.0–7.0)) with clinically definite MS (181 relapsing-remitting, 42 SPMS, 14 primary progressive MS) or clinically isolated syndrome (n = 10), according to the revised 2010 McDonald's criteria.(17) Further inclusion criteria required patients to have a Neurostatus-EDSS between 0.0 to 7.0,(18) aged above 18, without additional diseases that contribute to disability. Exclusion criteria were patients that were unable to follow procedures or read the consent due to psychological disorders, dementia or understand either the local language or English.

Ambulation and upper extremity function measures

All patients received a standardized EDSS assessment according to the Neurostatus definitions on the day of recording.(18) Furthermore, the 9HPT and T25WT were performed, as implemented in the MSFC.(7) For the T25WT, the performance of two trials was averaged for each patient. For the 9HPT, the averaged value of two trials were taken for the dominant and non-dominant hand, which was determined through questioning the patient. For this study three UEF movements from Assess MS were chosen: the 'finger-to-nose test' (FNT) to evaluate the level of ataxia (tremor/ dysmetria), the 'pronator drift test' (PDT) to evaluate the level of pyramidal dysfunction, and the 'drinking-from-cup' test (CUP) to evaluate the level of motor dysfunction affecting ADL.(19) For CUP, patients were asked to take a sip of water from a standardized ¾ filled plastic cup on a table in front of them at arm's length.

The Arm Function in Multiple Sclerosis Questionnaire (AMSQ)(20) and the Multiple Sclerosis Impact Scale 29 (MSIS-29)(21) were acquired to assess activity limitations due to impaired UEF in MS, and to examine the perceived physical and psychological impact of MS from the patients' perspective, respectively.

Video rating of movements

A comprehensive description of the methodology used for video rating can be found elsewhere.(22, 23) In short, all colour videos of the movements of patients were rated by two neurologists with broad experience in MS. Predetermined rating scales were used for FNT and PDT based on the Neurostatus-EDSS functional system scoring definitions, which are rated on a five- ('0=none' to '4=severe limb ataxia') and three-point Likert scale ('0=none' to '2=evident pronation and downward drift'), respectively.(18) For CUP a five-point Likert scale was created ranging from 0 (i.e. 'normal performance') through 4 (i.e. 'impossible to perform). Using an algorithm that takes into account individual rater bias, the videos were subsequently assigned a consensus score.(24) This consensus score was used in the statistical analysis. Videos of insufficient quality or if not performed according to the protocol, were primarily excluded from the analysis.

Classification of ambulatory and upper extremity function impairment groups

The EDSS score was used to stratify patients into three clinically different ambulation groups based on the Neurostatus-EDSS scoring definitions: 'fully ambulatory' (able to walk ≥500 meters; EDSS 0 – 3.5), 'mild ambulatory impairment' (unassisted walking distance of ≥100-300 meters, but <500 meters; EDSS 4.0 – 5.5) and 'severe ambulatory impairment' (assistance required when walking, able to take a few steps; EDSS 6.0 – 7.0). The 9HPT of the dominant hand (despite disability due to MS) was used to stratify patients into clinically different UEF groups: 'normal UEF' (<18 seconds), 'mild UEF impairment' (≥18 seconds and <33.3 seconds) and 'severe UEF impairment' (≥33.3 seconds). The lower benchmark was chosen based on previous work of Kierkegaard et al.(25) who provided evidence that 18 seconds on the 9HPT differentiated MS patients with no impaired UEF from patients with minimally impaired UEF, who were at risk for activity limitations and participation restrictions.(25) The upper benchmark was derived from the proposed cut-off value of Lamers et al.(26) who used a median split of the 9HPT score of a large MS sample (n=105) , which differentiated between 'mild' and 'marked to severe' UE dysfunction based on various measures of UEF (including assessment of strength, tremor, spasticity, pain) and participation level. Patients who were unable to perform the 9HPT due to MS-related impairment were categorized in the 'severely' impaired groups. This

classification resulted in nine clinically different patient groups, according to the level of ambulatory and UEF impairment.

To identify different aspects of UEF at various levels of ambulation, the following variables were compared between the aforementioned patient groups: FNT, PDT, CUP, AMSQ and MSIS; along with demographical and disease specific characteristics (age, sex, disease type, disease duration, T25WT). Disease type (i.e. relapsing / progressive), PDT (i.e. 0 / ≥1), FNT and CUP (i.e. 0 / ≥1 / ≥2) were further categorized for ease of interpretation.

To explore the relationship between the 9HPT, FNT, PDT and CUP, comparison Venn diagrams with multiple overlapping closed curves were created, each representing a set of patients performing abnormal [score of (≥1)] on either of the tests. Patients were included in the final analyses only if they completed all 4 tests.

Statistical analysis

Statistical analysis was performed in SPSS version 24 (Chicago, IL). A p-value of less than 0.05 was considered statistically significant. The normality of each variable was assessed using histograms and normality plots. For variables with a normal distribution mean values with standard deviations (SD) were calculated, and median values with interquartile ranges (IQR) for non-normal distributions. Variables following a normal distribution were compared between groups using analysis of variance models, as appropriate, whereas categorical variables were compared using the Mann–Whitney and Chi-square tests. The distribution of patients across the EDSS- and 9HPT-defined ambulation and UEF groups was compared using a Chi-square test.

3 RESULTS

Ambulation groups

The median EDSS of the total cohort was 3.5 (range 0 – 7.0). Based on the Neurostatus-EDSS, 147 patients were categorized as 'fully ambulatory', 46 patients as 'mild ambulatory impairment', and 54 patients as 'severe ambulatory impairment' (table 1). All groups were similar in terms of sex ($p=0.26$). A direct comparison between groups revealed that patients with worse levels of ambulatory impairment displayed: a higher age, more progressive MS phenotypes, longer disease duration, and higher MSIS and T25FW scores, with significant differences between all groups (overall $p<0.01$). However, age, disease duration and MSIS did not differ between the 'mild' and 'severe ambulatory impairment' groups ($p=0.17$, 0.21 and 0.27, respectively).

Table 1, demographic, clinical and upper extremity function measures of the MS patient ambulation groups

	Fully ambulatory[a] n = 147	Mild ambulatory impairment[a] n = 46	Severe ambulatory impairment[a] n = 54	Overall	Fully vs Mild	Mild vs Severe	Fully vs Severe
Demographic data & MS characteristics							
Age (in years)[b]	43.2 (11.1)	50.9 (15.1)	53.5 (10.9)	**<0.01**	**<0.01**	0.17	**<0.01**
Sex (female / male)	104 / 43	29 / 17	32 / 22	0.26	-	-	-
Disease type (Relapsing / Progressive)	142 / 5	34 / 12	15 / 39	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Disease duration (in years)[c]	10 (4-16)	18 (9-27)	20 (12-31)	**<0.01**	**<0.01**	0.21	**<0.01**
T25WT (in seconds)[c]	4.6 (4.0-5.2)	5.7 (4.9-5.2)	11.8 (8.5-15.9)	**<0.01**	**<0.01**	**<0.01**	**<0.01**
MSIS[c]	42 (33-57)	73 (56-84)	74 (62-95)	**<0.01**	**<0.01**	0.27	**<0.01**
Upper extremity function measures							
9HPT dominant hand (in seconds)[c]	19.9 (18.3-22.4)	25.0 (21.7-30.2)	33.0 (25.8-41.4)	**<0.01**	**<0.01**	**0.01**	**<0.01**
FNT dominant hand (0 / 1 / ≥2)	82 / 36 / 11	9 / 20 / 13	5 / 17 / 24	**<0.01**	**<0.01**	0.10	**<0.01**
FNT non-dominant hand (0 / 1 / ≥2)	78 / 41 / 13	6 / 21 / 14	2 / 22 / 26	**<0.01**	**<0.01**	0.09	**<0.01**
PDT (0 / ≥1)	124 / 15	27 / 14	21 / 29	**<0.01**	**<0.01**	**0.04**	**<0.01**
CUP dominant hand (0 / 1 / ≥2)	100 / 30 / 5	10 / 23 / 8	4 / 17 / 26	**<0.01**	**<0.01**	**<0.01**	**<0.01**
CUP non-dominant hand (0 / 1 / ≥2)	101 / 30 / 5	15 / 22 / 7	1 / 24 / 24	**<0.01**	**<0.01**	**<0.01**	**<0.01**
AMSQ[c]	33 (31-40)	50 (38-74)	81 (54-89)	**<0.01**	**<0.01**	**<0.01**	**<0.01**

Abbreviations: Relapsing = clinically isolated syndrome and relapsing-remitting multiple sclerosis; Progressive = secondary and primary progressive multiple sclerosis; T25WT= timed 25-foot walk test; 9HPT= nine-hole peg test; AMSQ= Arm Function in Multiple Sclerosis Questionnaire; CUP= drinking from cup test; FNT= finger to nose test; PDT= pronator drift test; MSIS= Multiple Sclerosis Impact Scale. [a] Fully ambulatory = EDSS 0 – 3.5, mild ambulatory impairment = EDSS 4.0 – 5.5, severe ambulatory impairment = EDSS 6.0 – 7.0); [b] Data are mean with standard deviation for normally distributed variables; [c] Because of non-normal distribution, median and interquartile range are provided; P-values in bold represent significant values (after Bonferroni correction).

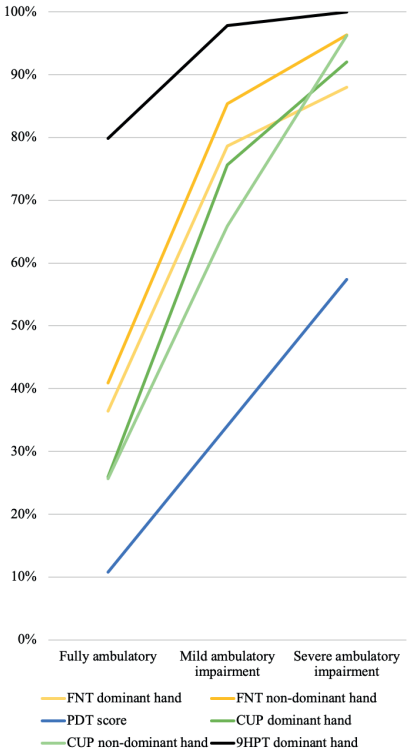


Figure 1, distribution of patients with abnormal upper extremity test performance among the clinically different ambulatory groups. Abbreviations: 9HPT= nine-hole peg test; FNT = finger-to-nose test; PDT = pronator drift test; CUP = drinking-from-a-cup. Shown are the percentages of patients that perform abnormal (score ≥1) on the 9HPT, FNT, PDT and CUP test among the clinically different ambulatory groups.

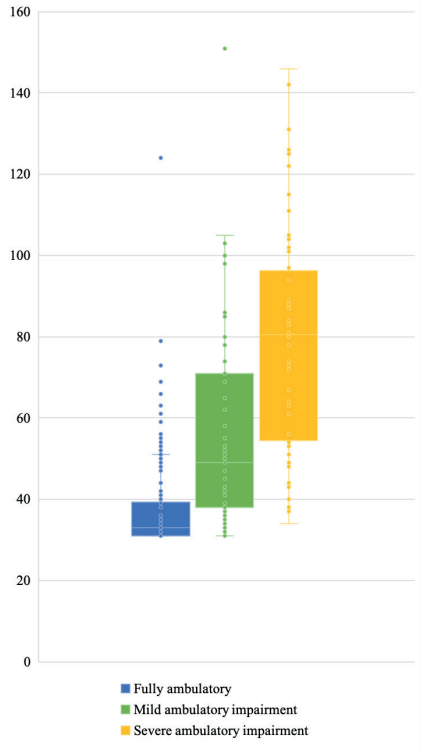


Figure 2, distribution of Arm Function in Multiple Sclerosis Questionnaire (AMSQ) results among the clinically different ambulatory groups. Shown are the median values with (interquartile) range.

A similar trend was observed for all UEF measures, with a worse level of UEF performance along with a worse level of ambulatory impairment (all group differences overall $p<0.01$). This is illustrated by figure 1 and 2, which display the distribution of patients that performed abnormally on the 9HPT, FNT, PDT and CUP (percentages), and AMSQ (median with IQR) among the ambulation groups. This clearly illustrates that most patients were restricted on the 9HPT in each group, followed by FNT, CUP and PDT.

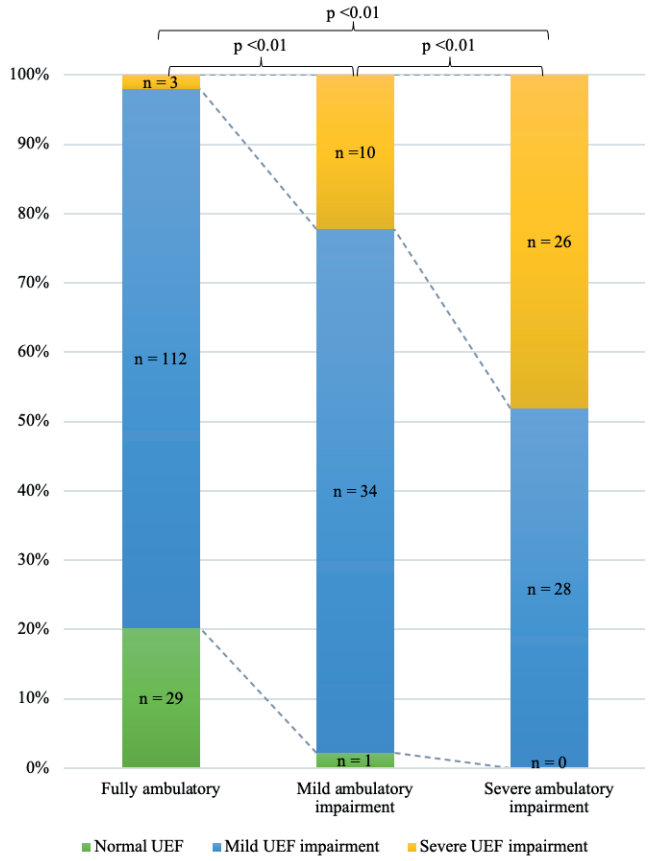


Figure 3, distribution of patients among the clinically different ambulatory and UEF impairment groups as defined by the EDSS- and 9HPT-benchmarks. Abbreviations: UEF = upper extremity function; EDSS = Expanded Disability Status Scale; 9HPT= nine-hole peg test. Fully ambulatory: EDSS 0 – 3.5; mild ambulatory impairment: EDSS 4.0 – 5.5; severe ambulatory impairment: EDSS 6.0 – 7.0; normal UEF 9HPT <18 seconds, mild UEF impairment 9HPT 18 seconds and <33.3 seconds; severe UEF impairment 9HPT ≥33.3 seconds

Distribution of upper extremity function groups

Figure 3 shows the results of the nine clinically different ambulatory and UEF impairment groups, as defined by the EDSS- and 9HPT-benchmarks. Four patients were excluded from the analysis since the 9HPT was missing due to reasons not related to MS: three patients in the 'fully ambulatory' group and one patient in the 'mild ambulatory impairment' group. Overall, the cohort consisted of relatively few patients with 'normal UEF' (n = 30) and all of these patients, except one, were 'fully ambulatory'. Most patients had mild UEF impairment (n=174) accounting for the largest proportion in all ambulation groups (77.8%, 75,6% and 51.9%, respectively). We observed a decline in

the proportion of patients with 'normal UEF' across the level of ambulatory impairment: 20.1% (fully ambulatory), 2.2% (mild ambulatory impairment) and 0% (severe ambulatory impairment); and a concurrent increase in the proportion of patients with 'severe UEF impairment'; 2.1%, 22.2% and 48.1%, respectively.

EDSS = Expanded Disability Status Scale; 9HPT= nine-hole peg test. Fully ambulatory: EDSS 0 – 3.5; mild ambulatory impairment: EDSS 4.0 – 5.5; severe ambulatory impairment: EDSS 6.0 – 7.0; normal UEF 9HPT <18 seconds, mild UEF impairment 9HPT ≥18 seconds and <33.3 seconds; severe UEF impairment 9HPT ≥33.3 seconds.

Combination of upper extremity function tests

Figure 4 shows the Venn diagrams, revealing the proportion of area of overlap between patients who performed abnormally on the 9HPT, FNT, PDT and CUP. Fifty-six patients were excluded from the analyses due to missing values on either one of the tests: 4 on the 9HPT, 23 on CUP, 30 on FNT and 17 on PDT (74 missing tests in total). Of the 191 remaining patients there were 165 (86.4%) who were restricted on the 9HPT, 105 (55.0%) patients performed abnormally on FNT, 49 (25.7%) on PDT and 91 (47.6%) on CUP. The vast majority of patients that were restricted on the 9HPT also performed abnormally on the other tests.

4 DISCUSSION

In this study, we aimed to explore the concurrent presence of UEF and ambulatory impairment in patients with MS, examining various aspects of UEF across different levels of ambulation. In a large representative cohort of MS patients with overall mild disability, UEF and ambulation showed distinct patterns of impairment. Although most patients exhibited only mild UEF deficits, this was already a prominent sign while ambulation was not obviously affected. Once ambulation was clearly impaired, deficits in UEF concurrently existed, affecting multiple aspects of function. Strikingly, in patients with markedly severe walking impairment (EDSS 6.0 – 7.0) there remained a large proportion of patients who displayed only mild UEF impairment. Our observations underscore the importance of the consideration of a patients' UEF impairment with multiple measures, independent of ambulation.

In the current study, more than 80% of patients exhibited some level of UEF impairment as measured with the 9HPT, which is in line with previous data on the magnitude of the problem.(27) To identify the relevant underlying constructs that comprised UE dysfunction, we examined the relationship between the 9HPT and the FNT, PDT and CUP test using Venn diagrams. Although the 9HPT is often used as a performance-based measure of manual

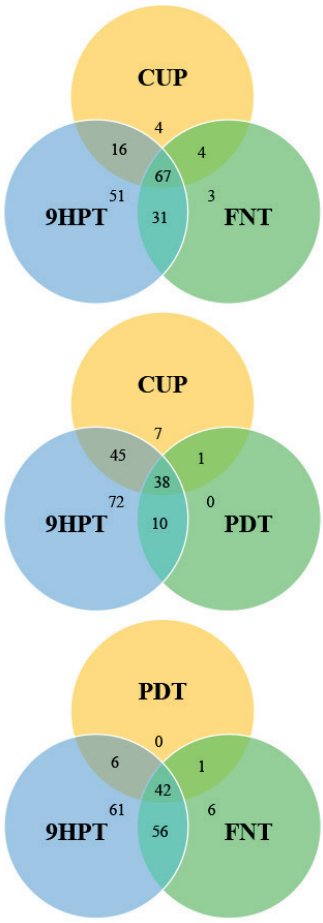


Figure 4, Venn diagrams of the 9HPT, FNT, PDT and CUP tests. Abbreviations: 9HPT= nine-hole peg test; FNT = finger-to-nose test; PDT = pronator drift test; CUP = drinking-from-a-cup. Data of 191 patients are included in the diagrams, excluding 56 patients because of missing values on either one of the tests (n= 74). Shown are schematic diagrams representing all possible relations between the patients that performed abnormally (score ≥1) on the 9HPT, FNT, PDT and CUP

dexterity, our results indicated that it may also be useful for identifying patients with pyramidal dysfunction, UE ataxia and the ability to perform ADL as the areas between patients performing abnormal on the 9HPT were largely overlapping with the clinical tests. This is in line with an earlier report on largely the same cohort of patients, which revealed that a large percentage of the variance of the 9HPT was explained by a combination of the UEF movements, of which CUP contributed most in all regression models.(19) However, additional studies are needed to determine the utility of the 9HPT as a performance-based measure of a variety of functional aspects in MS.

Previous studies already provided evidence that the 9HPT appeared to be a good outcome measure for differentiating the levels of severity of UE dysfunction.(25, 26, 28) Use has been recommended given its high test-retest reliability, sensitivity in discriminating MS patients from healthy subjects, and MS patients with different levels of UEF impairment. Furthermore, the test shows high convergent validity with other manual dexterity as well as more comprehensive UEF measures.(28) To date, it is still not clear which cut-off values on the 9HPT should be applied. For this current study, we used the proposed cut-off values published by Kierkegaard et al and Lamers et al for the lower and upper benchmark, because of their supposed relationship to real life anchors and functional independence.(29) Longitudinal and larger studies are needed to confirm the clinical utility and relevance of these proposed 9HPT benchmarks and to parse out whether there are additional benchmarks in the lower and higher ranges of performance.

Previous papers have reported the relative patterns of UEF and ambulation impairment in patients with MS.(27, 30, 31) A cross-sectional study explored UEF and ambulation across different levels of disability.(27) The 9HPT and T25FW were used to stratify patients into different UEF and ambulation groups (>1 SD worse than age-/sex-related norms). Disability was defined as 'mild' (EDSS 1.0 – 3.5), 'moderate' (EDSS 4 – 5.5) or 'severe' (EDSS 6 – 9.5), which resulted in a distribution of 71%, 92% and 97% patients with UEF impairment; and 22%, 89% and 100% patients with walking impairment, respectively. The authors concluded that the majority of patients with MS experienced several concurrent disabilities, which were independently associated with the perceived physical and psychological impact from the patients' perspective as measured with the MSIS-29. Although these findings are in line with our data, the authors did not provide a detailed insight into the broader aspects of UEF.

Another report provided a more detailed insight into the UEF of MS patients across different EDSS subgroups. In this cross-sectional study, various measures of UEF were included to assess strength, spasticity, sensation, and also the 9HPT and a PRO measure were incorporated. Findings indicated a concurrent deterioration of UEF on all aspects with disability accrual, which is in line with our data. However, a smaller proportion of patients was reported as being impaired on the 9HPT, which is probably due to the use of a different definition of "abnormal".(30)

Another study investigated patterns of UEF and ambulation deterioration longitudinally.(31) The primary goal of the authors was to improve the assessment of disability accrual in patients with SPMS using the 'EDSS-plus', a composite endpoint adding the T25FW and 9HPT to the EDSS. They found that once the T25WT deteriorated with 20% change or more, UEF deteriorated more gradually than ambulation in the subsequent two years, especially in

the most severely disabled group (EDSS 6.0-6.5). Although our data might suggest a similar trend, no inferences on a longitudinal relationship can be made.

Other studies confirm the presence of UEF impairment early in the disease course. Data from questionnaires of 35.000 patients collected in the North American Research Committee on Multiple Sclerosis registry revealed that in the first year after onset, 52% of patients reported minimal to mild UEF impairment, while 40% perceived no problems.(32) For ambulation, this was 35% and 50%, respectively. An important limitation of this study was that function was assessed with short 6- or 7-item questionnaires only, and additional physician- or performance-based measures were lacking. Two other studies on MS patients found subtle UEF impairments in the early disease course that were not detected with standard neurological examination.(33, 34) Abnormalities were found in movement smoothness, speed profile, and lifting tasks. These limitations were noticed by patients and had an impact on the ability to perform ADL.(33)

Several neurobiological explanations for different patterns of ambulatory and UE dysfunction can be given. One hypothesis is that changing functional networks that compensate for increasing structural damage are more robust in preserving UEF than in preserving ambulation. A similar phenomenon has been described in the preservation of cognition in patients with MS.(35) The "cognitive reserve hypothesis" postulates that genetic and environmental factors attenuate the negative effect of disease burden on cognitive decline.(36) One might assume that ambulation is more brittle than UEF (i.e. neural degeneration has a larger impact on function), but has a larger capacity for compensation. Under this scenario, ambulation remains intact while structural damage has already affected UEF. However, once the buffer capacity is exhausted, the accumulation of walking impairment will outpace the worsening of UEF. Another hypothesis is that of a central length-dependent axonopathy.(37) In this mechanism, longer neurons to the lower extremities are more vulnerable to accumulating focal damage, which causes secondary neurodegeneration. Clinically, this will lead to a faster deterioration of ambulation than UEF.

Study limitations

Our study also has some limitations. To define the level of ambulatory impairment, EDSS cut-off values were used. According to the Neurostatus-EDSS definitions,(18) a patient with unrestricted ambulation (able to walk >500 meters) can have an EDSS score between 0 and 3.5, depending on impairment of other functional systems. As a result, we cannot rule out the possibility that other patterns of disability contributed to the classification of ambulatory function in our MS patients. Furthermore, in subjects with an

EDSS lower than 2.0, walking and balance can already be impaired compared to healthy controls, which should be considered when interpreting these findings. Another limitation was that MS patients included in our study were relatively mildly disabled (median EDSS 3.5), which limits the generalisation of our data to more severely disabled patients. However, our sample was representative for a general MS

To conclude, this study emphasizes the importance of wider incorporation of performance-based and PRO measures that can be used for screening and assessment of UEF impairment in daily clinical practice and treatment studies. We have provided data on the relative patterns of UEF and ambulatory impairment and explored the concurrent presence of a variety of functional aspects of UEF. Further stratifying patients according to UEF, beyond ambulation, will enhance patient selection for future treatment and support the development of rehabilitation strategies specially targeted towards UEF impairment. Future studies are needed to explore the exact longitudinal relationship between accrual of ambulatory and UEF impairment, which will increase our understanding of the magnitude and impact of UEF impairment in MS patients and eventually improves patient care.

REFERENCES

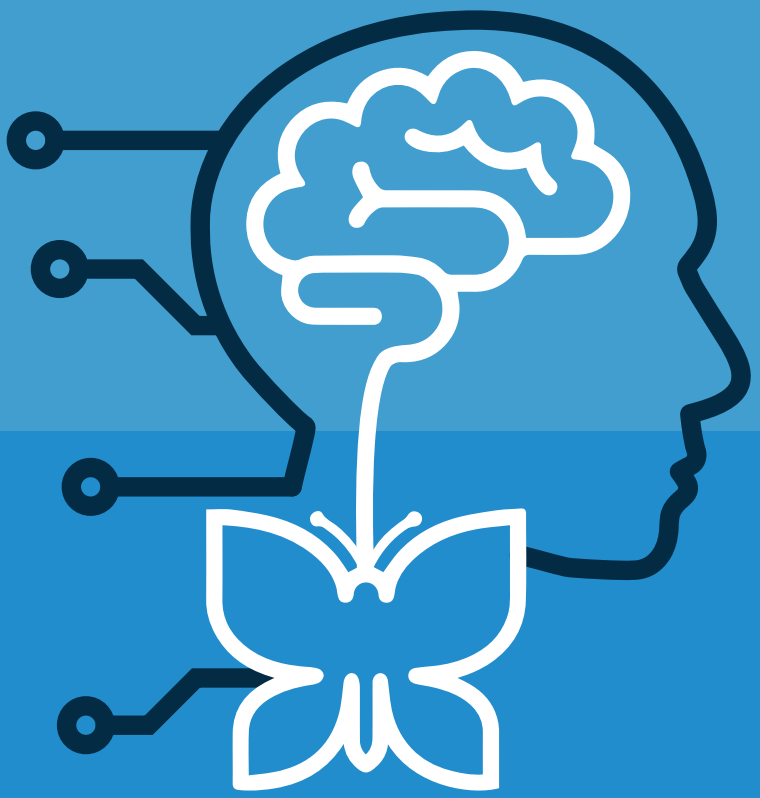
1. Kamm CP, Uitdehaag BM, Polman CH. Multiple sclerosis: current knowledge and future outlook. Eur Neurol. 2014;72(3-4):132-41.
2. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC Neurol. 2014;14:58.
3. Kraft GH, Amtmann D, Bennett SE, Finlayson M, Sutliff MH, Tullman M, et al. Assessment of upper extremity function in multiple sclerosis: review and opinion. Postgrad Med. 2014;126(5):102-8.
4. Yozbatiran N, Baskurt F, Baskurt Z, Ozakbas S, Idiman E. 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(1-2):117-22.
5. Koch MW, Murray TJ, Fisk J, Greenfield J, Bhan V, Jacobs P, et al. Hand dexterity and direct disease related cost in multiple sclerosis. J Neurol Sci. 2014;341(1-2):51-4.
6. van Munster CE, Uitdehaag BM. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs. 2017;31(3):217-36.
7. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain. 1999;122 (Pt 5):871-82.
8. Tur C, Moccia M, Barkhof F, Chataway J, Sastre-Garriga J, Thompson AJ, et al. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat Rev Neurol. 2018;14(2):75-93.
9. Kapoor R, Ho PR, Campbell N, Chang I, Deykin A, Forrestal F, 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.
10. Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. N Engl J Med. 2017;376(3):209-20.
11. Lamers I, Maris A, Severijns D, Dielkens W, Geurts S, Van Wijmeersch B, et al. Upper Limb Rehabilitation in People With Multiple Sclerosis: A Systematic Review. Neurorehabil Neural Repair. 2016;30(8):773-93.
12. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors. 2015;2(1):e11.
13. D'Souza M BJ, Kontschieder P. Automated quantification of motor dysfunction in multiple mclerosis using depth-sensing computer vision. Neurology. 2015;84(220):1526-632.
14. Kontschieder P, Dorn JF, Morrison C, Corish R, Zikic D, Sellen A, et al. Quantifying progression of multiple sclerosis via classification of depth videos. Med Image Comput Comput Assist Interv. 2014;17(Pt 2):429-37.
15. Sarkar A, Morrison, C, Dorn, JF. Setwise comparison: Consistent, scalable, continuum labels for computer vision. Proceedings of the 2016 CHI conference on human factors in computing systems. New York: Association for Computing Machinery; 2016. p. 261-71.

16. Burggraaff J, Dorn J, D'Souza M, Morrison C, Kamm CP, Kontschieder P, et al. Video-Based Pairwise Comparison: Enabling the Development of Automated Rating of Motor Dysfunction in Multiple Sclerosis. Arch Phys Med Rehabil. 2019.
17. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292-302.
18. Kappos L, D'Souza M, Lechner-Scott J, Lienert C. On the origin of Neurostatus. Mult Scler Relat Disord. 2015;4(3):182-5.
19. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Mult Scler. 2019;25(12):1673-81.
20. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BM. The Arm Function in Multiple Sclerosis Questionnaire (AMSQ): development and validation of a new tool using IRT methods. Disabil Rehabil. 2015:1-7.
21. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain. 2001;124(Pt 5):962-73.
22. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Mult Scler. 2018:1352458518796690.
23. D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. Mult Scler J Exp Transl Clin. 2018;4(3):2055217318792399.
24. Advait Sarkar CM, Jonas F. Dorn, Rishi Bedi, Saskia Steinheimer, Jacques Boisvert, Jessica Burggraaff, Marcus D'Souza, Peter Kontschieder, Samuel Rota Bulo, Lorcan Walsh, Christian P. Kamm, Yordan Zaykov, Abigail Sellen, Sian Lindley. Setwise Comparison: Consistent, Scalable, Continuum Labels for Computer Vision. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems

Pages. 2016:261-71.

25. Kierkegaard M, Einarsson U, Gottberg K, von Koch L, Holmqvist LW. The relationship between walking, manual dexterity, cognition and activity/participation in persons with multiple sclerosis. Mult Scler. 2012;18(5):639-46.
26. Lamers I, Cattaneo D, Chen CC, Bertoni R, Van Wijmeersch B, Feys P. 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(1):65-75.
27. Johansson S, Ytterberg C, Claesson IM, Lindberg J, Hillert J, Andersson M, et al. High concurrent presence of disability in multiple sclerosis. Associations with perceived health. J Neurol. 2007;254(6):767-73.
28. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult Scler. 2017;23(5):711-20.

29. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol. 2007;6(12):1094-105.
30. Bertoni R, Lamers I, Chen CC, Feys P, Cattaneo D. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with multiple sclerosis. Mult Scler. 2015;21(12):1566-74.
31. Cadavid D, Cohen JA, Freedman MS, Goldman MD, Hartung HP, Havrdova E, et al. The EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. Mult Scler. 2017;23(1):94-105.
32. Kister I, Bacon TE, Chamot E, Salter AR, Cutter GR, Kalina JT, et al. Natural history of multiple sclerosis symptoms. Int J MS Care. 2013;15(3):146-58.
33. Krishnan V, Jaric S. Hand function in multiple sclerosis: force coordination in manipulation tasks. Clin Neurophysiol. 2008;119(10):2274-81.
34. Solaro C, Brichetto G, Casadio M, Roccatagliata L, Ruggiu P, Mancardi GL, et al. Subtle upper limb impairment in asymptomatic multiple sclerosis subjects. Mult Scler. 2007;13(3):428-32.
35. Schoonheim MM, Geurts JJ, Barkhof F. The limits of functional reorganization in multiple sclerosis. Neurology. 2010;74(16):1246-7.
36. Sumowski JF, Leavitt VM. Cognitive reserve in multiple sclerosis. Mult Scler. 2013;19(9):1122-7.
37. Giovannoni G, Cutter G, Sormani MP, Belachew S, Hyde R, Koendgen H, et al. Is multiple sclerosis a length-dependent central axonopathy? The case for therapeutic lag and the asynchronous progressive MS hypotheses. Mult Scler Relat Disord. 2017;12:70-8.



CHAPTER 4

Tasks of activities of daily living (ADL) are moving more valuable than the classical neurological examination to assess upper extremity function and mobility in muliple sclerosis

Caspar E.P. van Munster, Marcus d'Souza, Saskia Steinheimer, Christian P. Kamm, Jessica Burggraaf, Jonas Dorn, Jacques Boisvert, Manuela Diederich, Kristina Kravalis, Frank Dahlke, Ludwig Kappos, Bernard M.J. Uitdehaag.

Multiple Sclerosis Journal. 2019 Oct; 25(12):1673-1681

ABSTRACT

Background
Accurate clinical assessment in multiple sclerosis (MS) is challenging. The Assess MS system is being developed to automatically quantify motor dysfunction in MS, including upper extremity function (UEF) and mobility.

Objective
To determine to what extent combinations of standardized movements included in the Assess MS system explain accepted measures of UEF and mobility.

Methods
MS patients were recruited at four European MS centres. Eight movements were selected, including tasks of activities of daily living (ADL) and classical neurological tests. Movements were recorded on video and rated by experienced neurologists (n=5). Subsequently, multivariate linear regression models were performed to explain the variance of the 9-Hole-Peg Test (9HPT), Arm Function in Multiple Sclerosis Questionnaire (AMSQ) and Timed-25 Foot Walk test (T25WT).

Results
In total 257 patients were included. The movements explained 62.9 to 80.1% of the variance of the 9HPT models, 43.3 and 44.3% of the AMSQ models, and 70.8% of the T25WT. In all models, tasks of ADL contributed most to the variance.

Conclusion
Combinations of movements are valuable to assess UEF and mobility. Incorporating ADL tasks into daily clinical practice and clinical trials may be more valuable than the classical neurological examination of UEF and mobility.

1 INTRODUCTION

Assessment of disability in multiple sclerosis (MS) is traditionally performed with the Expanded Disability Status Scale (EDSS), which is a physician-based method. However, there are several limitations to the EDSS. Some of these are related to the heterogeneous nature of MS, others are inherently a consequence of methodological aspects of the scale itself, e.g. a high inter- and intrarater variability and a disproportional impact of ambulatory function on the total score.(1) To improve the clinical assessment of MS disability, various performance-based tests were introduced. Widely accepted performance-based tests are the Nine-Hole Peg Test (9HPT),(2) the Timed 25-foot Walk Test (T25WT)(3) and the Symbol Digit Modalities Test.(4) Also, patient-reported outcome measures contribute to clinical assessment by giving insight into the patient-perspective of a certain aspect, such as upper extremity function (UEF) with the Arm Function in Multiple Sclerosis Questionnaire (AMSQ).(5)

A potential valuable improvement in clinical assessment would be the automatic quantification of disability with Machine Learning Algorithms (MLA). With this in mind, the Assess MS system is being developed to automatically quantify motor functioning by capturing standardized movements of patients recorded by the Microsoft Kinect® camera (Microsoft, Redmond, SA, USA) (6). Several of these movements are used for the assessment of UEF and mobility, which are important functional domains in MS since the majority of patients experience UEF and mobility impairment at some point in the course of their disease.(7, 8) Furthermore, impaired UEF and mobility can impact the ability to perform ADL, on general health perception,(9, 10) and on quality of life and social participation.(11, 12)

The standardized movements used to develop the Assess MS include several classical neurological tests and tasks of activities of daily living (ADL), which can easily be administrated in daily practice. However, it is unclear to what extent these tests contribute to determining UEF and mobility. Presumably, not all tests are required to assess these functions, and it is unclear how much each test contributes to UEF and mobility.

In the current study, we investigate to what extent combinations of standardized movements explain accepted measures of UEF and mobility.

2 METHODS

Patients

Patients were recruited at four large European MS centres in Amsterdam, Basel, Bern and Lucerne. Inclusion criteria were: aged older than 18 years, diagnosis of MS or a clinically isolated syndrome suspicious for MS according

to the 2010 revised McDonald criteria, EDSS score between 0 and 7. Exclusion criteria were: inability to follow procedures or read the informed consent due to psychological disorders, dementia or insufficient ability to speak the local language or English. Each patient provided written informed consent prior to study entry and the study was approved by the respective ethics committees.

Procedure

An example of the experimental setup is illustrated in figure 1. All patients were recorded with the Kinect® camera that simultaneously captures depth and colour videos.

Figure 1, experimental setup. In this example, a patient sits on a chair and performs the finger-to-nose test. The Assess MS machine is placed perpendicular to the patient and displays an audio-guided instruction video of the movement on the large screen. A physician on the other side operates the machine with a tablet that in this example has been turned towards the patient for demonstration purposes. After showing the instruction video, the patient performs the movement after a beep. This is recorded by the Kinect® camera and stored locally on the machine. The people seen in this picture are members of the study group that gave their consent.

Eight standardized movements covering trunk, upper and lower extremities, which are partly based on the classical neurological examination, and movements typical of ADL were chosen. Three movements covering UEF were performed: finger-to-nose test (FNT), pronator drift test (PDT), as classical neurological tests, and drinking from a cup (CUP), as an ADL movement. For CUP, patients had to take a sip from a standardized plastic cup that was

at least half-full of water. To assess mobility the following five movements were performed: Romberg test (ROM), tight-rope-walking (TRW), as classical neurological tests, and sit-to-stand (STS), turning-on-the-spot (TOS) and walking a distance of 25-foot (GAT), as ADL movements. For STS, patients were instructed to get up from a standardized chair without touching it. Schematic representations of the movements can be found in figure 2.

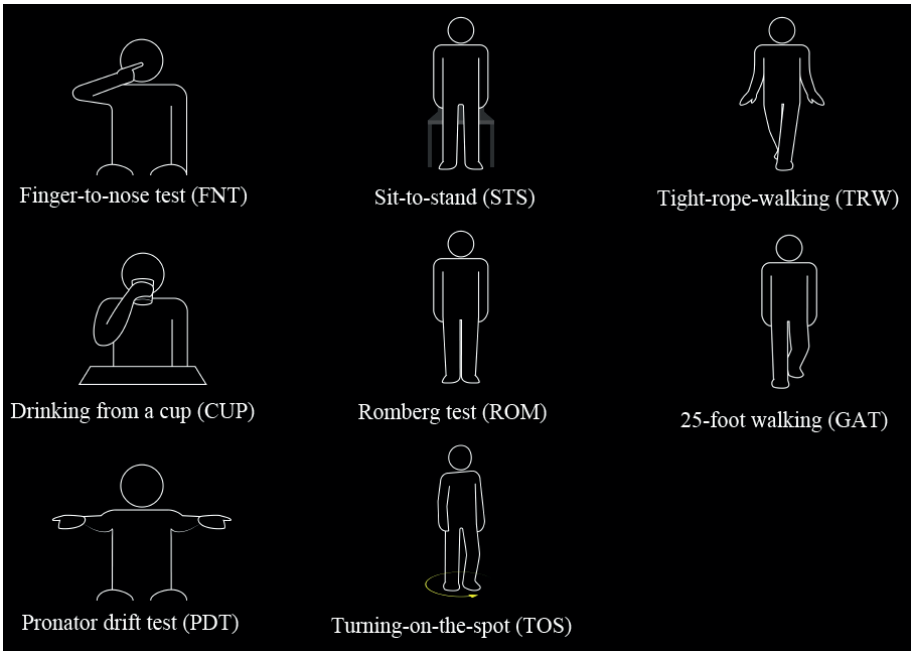


Figure 2, schematic representations of the standardized movements.

All colour videos were rated by two independent neurologists with experience in MS. Each patient video was given a score based on a predetermined rating scale (see table 1). Some of these scales (FNT, PDT and ROM) were derived from the functional system subscores from the Neurostatus-EDSS.(13) For the ADL movements a 0 to 4 scale was created, in which 0 is a normal performance, 1 mildly impaired (minor interference with function), 2 moderately impaired (clear interference with function), 3 severely impaired (severe interference with function) and 4 is impossible to perform. In addition, the videos were also presented as sets which the neurologists ordered from least affected to most affected. Using an algorithm similar to the one described by Sarkar et al(14) that takes into account individual rater bias, the videos were then assigned a consensus score. This consensus score was subsequently used in the statistical analysis. Videos of insufficient quality or videos that were not performed according to the protocol were excluded from further analysis. In

Table 1, rating scales of movements.

Movement	Grade	Severity	Description
Finger-to-nose test (FNT)	0	normal	
	1	signs only	
	2	mild	Tremor or clumsy movements easily seen, minor interference with function
	3	moderate	Tremor or clumsy movements interfere with function in all spheres
	4	severe	Most functions are very difficult
Drinking from a cup (CUP)	0	normal	
	1	mild	Discrete but clearly seen irregularities (problems grasping the cup, tremor, slow irregular movement, slowing in front of mouth, un-physiological posture of hand/arm (including holding the cup, cup gets impressed)). Minor interference with function (most of the movement is normal)
	2	moderate	Clear irregularities (problems grasping the cup, tremor, slow irregular movement, slowing in front of mouth, un-physiological posture of hand/arm (including holding the cup, cup gets impressed)). Clear interference with function (whole movement is affected)
	3	severe	Severe irregularities (problems grasping the cup, tremor, slow irregular movement, slowing in front of mouth, un-physiological posture of hand/arm (including holding the cup, cup gets impressed)). Severe interference with function (drinking from cup is very difficult including (almost) spilling of water)
	4	not possible	
Pronator drift test (PDT)	0	none	
	1	mild	
	2	evident	

Movement	Grade	Severity	Description
Sit-to-stand (STS)	0	normal	
	1	mild	Discrete but clearly seen irregularities. Minor interference with function, (most of the functions are normal)
	2	moderate	Clear irregularities (needs to lean forward, optimize sitting position, pushes with hands on thigh, instable). Clear interference with function (whole movement is affected)
	3	severe	Severe irregularities (needs to lean forward, optimize sitting position, pushes with hand on thigh, instable, multiple tries). Severe interference with function (standing up is very difficult)
	4	not possible	
Romberg test (ROM)	0	normal	
	1	mild	Mild instability with eyes closed
	2	moderate	Not stable with eyes closed
	3	severe	Not stable with eyes open
Turning-on-the-spot (TOS)	0	normal	
	1	mild	Discrete but clearly seen irregularities (ataxia, spasticity, limping, irregular/slow movement, widened range). Minor interference with function (most of the functions are normal)
	2	moderate	Clear irregularities (ataxia, spasticity, limping, irregular/slow movement, widened range). Clear interference with function (whole movement is affected)
	3	severe	Severe irregularities (ataxia, spasticity, limping, irregular/slow movement, widened range, tripping). Severe interference with function (movement is very difficult)
	4	not possible/ only with aid	

Movement	Grade	Severity	Description
Tight-rope-walking (TRW)	0	normal	
	1	mild	Discrete but clearly seen irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range). Minor interference with function (most of the movement is normal)
	2	moderate	Clear irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range). Clear interference with function (whole movement is affected)
	3	severe	Severe irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range, tripping). Severe interference with function (movement is very difficult)
	4	not possible/ only with aid	
25-foot walking (GAT)	0	normal	
	1	mild	Discrete but clearly seen irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range). Minor interference with function (most of the movement is normal)
	2	moderate	Clear irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range). Clear interference with function (whole movement is affected)
	3	severe	Severe irregularities (ataxia, spasticity, limping, irregular/ slow movement, widened range, tripping). Severe interference with function (movement is very difficult)
	4	not possible/ only with aid	

the current study, only the video ratings of movements were analysed. The development of MLA is part of another study that is currently being performed.

All patients received a standardized Neurostatus-EDSS assessment(13) on the day of recording, performed by another examiner than the before mentioned neurologists that rated the videos. Furthermore, the 9HPT and T25WT were administrated, as performance-based measures of UEF and mobility. All patients were asked to complete the Arm Function in Multiple Sclerosis Questionnaire (AMSQ)(5), as a patient-reported outcome measure for UEF.

Data analysis

Statistical analyses were performed in IBM SPSS Statistics for Macintosh, Version 24. A p-value of <0.05 was considered statistically significant. The normality of each variable was assessed using histograms and normality plots. For variables with a normal distribution mean values with standard deviation (SD) were calculated, and median values with interquartile ranges (IQR) for non-parametric distributions. For the movements that were performed multiple consecutive times (FNT three times for both sides, and 9HPT and T25WT two times), the best performance was used for statistical analyses. Spearman's rho correlation was used for assessing the relation between the 9HPT and AMSQ.

After confirming the absence of strong collinearity with partial regression (collinearity present if $r \geq 0.9$), combinations of the eight movements were used in stepwise multivariate linear regression models to determine how much the clinical ratings of the movements contribute to the variance of the 9HPT and the AMSQ in UEF, and the T25WT for mobility. For UEF, different models were used for the left and right side, and for the dominant and non-dominant hand. The rating scales were categorized into groups (i.e. dummy-variables were created), because the relation between the outcome variables and the rating scales of the movements was not linear.

3 RESULTS

In total 257 patients were included in this study of which 171 (66.5%) were women and the mean age was 46.6 years (SD 12.8). The mean disease duration was 14.9 years (SD 11.7). Clinical phenotypes were distributed as follows: clinically isolated syndrome 11 (4.3%), relapsing-remitting MS 186 (72.4%), secondary progressive MS 45 (17.5%) and primary progressive MS 15 (5.8%) patients. Twenty-four (9.3%) patients experienced a relapse within three months prior to inclusion. The median EDSS score was 3.0 (IQR 2.0). Baseline characteristics and results of the 9HPT, T25WT, and questionnaires are shown in table 2.

Table 2, baseline characteristics. SD = standard deviation; EDSS = Expanded Disability Status Scale; 9HPT = Nine-Hole Peg Test; T25WT = Timed 25-foot Walk Test; AMSQ = Arm Function in Multiple Sclerosis Questionnaire.

Total number of patients, N	**257**
Female, N (%)	171 (66.5)
Mean age, years (SD)	46.6 (12.8)
Mean duration, years (SD)	14.9 (11.7)
Disease type, n (%)	CIS: 11 (4.3) RRMS: 186 (72.4) SPMS: 45 (17.5) PPMS: 15 (5.8)
Relapse past 3 months, n (%)	24 (9.3)
Median EDSS (IQR)	3.0 (2)
Mean 9HPT right side, sec (SD)	24.2 (11.6)
Mean 9HPT left side, sec (SD)	24.8 (11.9)
Mean 9HPT dominant hand, sec (SD)	23.9 (11.0)
Mean 9HPT non-dominant hand, sec (SD)	25.2 (12.5)
Mean T25WT, sec (SD)	6.1 (3.8)
Mean AMSQ, sum (SD)	49 (24)

Correlation coefficients of the AMSQ, and 9HPT were 0.60 for the right side, 0.46 for the left side, 0.61 for the dominant hand, and 0.44 for the non-dominant hand. The video ratings of the eight movements are summarized in table 3.

Table 3, assessments of movements. FNT = finger-to-nose test; PDT = pronator drift test; CUP = drinking from a cup; ROM = Romberg test; TRW = tight-rope-walking; STS = sit-to-stand; TOS = turning-on-the-spot; GAT = walking a distance of 25-foot; n.a. = not applicable; *: two ambidextrous patients were defined as unrateable of which one CUP movements was unrateable.

Video rating score	0	1	2	3	4	Unrateable
Upper extremity function						
FNT right side	101	74	51	4	0	27
FNT left side	88	91	48	3	0	27
FNT dominant hand*	99	77	48	2	0	31
FNT non-dominant hand*	90	86	50	5	0	26
PDT	178	51	9	n.a.	n.a.	19
CUP right side	120	69	38	2	3	25
CUP left side	125	77	32	2	2	19
CUP dominant hand*	120	70	37	1	3	26
CUP non-dominant hand*	122	76	33	3	2	21
Mobility						
ROM	77	92	32	15	n.a.	41
TRW	64	52	37	19	38	47
STS	149	39	27	9	7	26
TOS right	59	57	43	25	5	68
TOS left	64	59	46	25	5	58
GAT	111	49	29	17	21	30

Regression models for upper extremity function and mobility

No co-linearity was found between the video ratings of the movements. Results of the regression models are displayed in table 4. CUP, PDT and FNT explained 73.2% of the variance of the right sided-9HPT, and 78.2% of the left sided-9HPT. CUP, PDT and FNT explained 80.1 and 62.9% of the variance in the dominant and non-dominant hand models of 9HPT, respectively. In all models CUP contributed most to the variance of the 9HPT, with only a minor contribution of PDT and FNT.

Table 4, regression models.

Stepwise regression model	Adjusted R^2	P-value	comment
Upper extremity function defined with the 9HPT			
Right side			
Step 1: CUP = 9HPT	0.631	P <0.001[a]	CUP contributed most to the variance
Step 2: CUP + PDT = 9HPT	0.686	P <0.001[b]	All variables contributed to the model
Step 3 (**final model**): CUP + PDT + FNT = 9HPT	0.732	P <0.001[b]	
Left side			
Step 1: CUP = 9HPT	0.709	P <0.001[a]	CUP contributed most to the variance
Step 2: CUP + PDT = 9HPT	0.750	P <0.001[b]	All variables contributed to the model
Step 3 (**final model**): CUP + PDT + FNT = 9HPT	0.782	P = 0.002[b]	
Dominant side			
Step 1: CUP = 9HPT	0.722	P <0.001[a]	CUP contributed most to the variance
Step 2: CUP + PDT = 9HPT	0.773	P <0.001[b]	All variables contributed to the model
Step 3 (**final model**): CUP + PDT + FNT = 9HPT	0.801	P = 0.004[b]	
Non-dominant hand			
Step 1: CUP = 9HPT	0.550	P <0.001[a]	CUP contributed most to the variance
Step 2: CUP + PDT = 9HPT	0.597	P <0.001[b]	All variables contributed to the model
Step 3 (**final model**): CUP + PDT + FNT = 9HPT	0.629	P = 0.004[b]	
Upper extremity function defined with the AMSQ			
Right and left side			
Step 1: CUP right = AMSQ	0.404	P <0.001[a]	CUP right side contributed most to the
Step 2 (**final model**): CUP right + FNT right = AMSQ	0.443	P = 0.001[b]	variance
Step 3: CUP right + FNT right+ CUP left = AMSQ	*0.430*	*P = 0.290[b]*	CUP left side, FNT left side and PDT did not
Step 3: CUP right + FNT right + FNT left = AMSQ	*0.432*	*P = 0.162[b]*	sign contribute to the model
Step 3: CUP right + FNT right + PDT = AMSQ	*0.439*	*P = 0.060[b]*	

Stepwise regression model	Adjusted R^2	P-value	comment
Dominant and non-dominant hand			
Step 1: CUP-D = AMSQ	0.400	P <0.001[a]	CUP dominant hand contributed most to the variance
Step 2 (**final model**): CUP-D + FNT-ND = AMSQ	0.433	P = 0.002[b]	
Step 3: CUP-D + FNT-ND + CUP-ND = AMSQ	*0.442*	*P = 0.144[b]*	CUP non-dominant hand, FNT dominant hand and PDT did not sign contribute to the model
Step 3: CUP-D + FNT-ND + FNT-D = AMSQ	*0.436*	*P = 0.261[b]*	
Step 3: CUP-D + FNT-ND + PDT = AMSQ	*0.443*	*P = 0.080[b]*	
Mobility defined with T25WT			
Step 1: STS = T25WT	0.615	P <0.001[a]	STS contributed most to the variance
Step 2: STS + GAT = T25WT	0.673	P <0.001[b]	All variables contributed to the model
Step 3: STS + GAT + ROM = T25WT	0.679	P <0.001[b]	
Step 4: STS + GAT + ROM + TOS right = T25WT	0.737	P <0.001[b]	
Step 5: STS + GAT + ROM + TOS right + TOS left = T25WT	0.760	P = 0.021[b]	
Step 6 (**final model**): STS + GAT + ROM + TOS right + TOS left + TRW = T25WT	0.708	P <0.001[b]	

9HPT = Nine-Hole Peg Test; T25WT = Timed 25-foot Walk Test; FNT = finger-to-nose test; PDT = pronator drift test; CUP = drinking from a cup; ROM = Romberg test; TRW = tight-rope-walking; STS = sit-to-stand; TOS = turning-on-the-spot; GAT = walking a distance of 25-foot; D = dominant hand; ND = non-dominant hand; □: p-value of Anova test; □: p-value of F-change.

In the AMSQ model in which CUP and FNT were stratified according to side (left and right side), 44.3% of the variance was explained by CUP and FNT of the right side. In the other AMSQ model in which CUP and FNT were stratified according to dexterity (dominant and non-dominant hand), 43.3% of the variance was explained with CUP from the dominant hand, and FNT of the non-dominant hand. In these models again, CUP contributed most to the variance of the AMSQ, and FNT contributed only to a minor proportion.

The six movements in the model for mobility, explained 70.8% of the variance of the T25WT. The STS contributed most to the variance in this model, and the other movements to a minor extent.

4 DISCUSSION

Combinations of standardized movements that are used in the Assess MS system explained UEF to a large extent as defined by the 9HPT as a measure of performance, and to a lesser extent as defined by the AMSQ as a measure of patient reported outcome. Mobility, as defined by the performance-based T25WT, was also explained to a large extent by a combination of movements. The ADL tasks CUP and STS contributed more to the variance of UEF and mobility, than classical neurological tests such as FNT and ROM.

The 9HPT was used as measure of UEF since it is the most widely used tool to assess UEF in MS studies so far.(15) It has good psychometric properties and clinical relevance concerning the ability to perform ADL tasks and quality of life.(2) Although the AMSQ has not yet been used as frequently as the 9HPT, it has good psychometric properties to assess UEF as well,(5, 16) and additionally gives insight into the patients' perspective of UEF.

A large percentage of the variance of the 9HPT was explained by a combination of movements, of which CUP contributed most in all models. Various explanations may be given for this. Firstly, CUP is a typical ADL movement, and the 9HPT is known to correlate with the ability to perform ADL tasks.(2) Secondly, the 9HPT primarily quantifies hand function (i.e. distal arm function), which is relevant for the ability to hold a cup and drink from it.(2) Lastly, one study found that approximately 53% of the variance of the 9HPT was explained by muscle strength, tactile sensitivity of the thumb and intention tremor,(17) which are all relevant in performing CUP.

The variance of the AMSQ could only be explained for 43.3 and 44.3%. The AMSQ covers a variety of patient-perceived ADL tasks, ranging from gross (such as holding a plate) to fine movements (such as using a keyboard), and covering both proximal and distal arm function.(5) Therefore, the AMSQ score probably represents more than what is covered with CUP, FNT and

PDT. The strong contribution of CUP, being a typical ADL movement, in these models is in line with the focus of AMSQ on patient-perceived ADL tasks.

Although the FNT and PDT are valuable in the neurological examination for localisation purposes, our results indicate that these tests are less sensitive to assess UEF, as defined with the 9HPT or AMSQ.

In our study, we found a lower correlation between the 9HPT and AMSQ (r = 0.44 to 0.61) than in another study (r = 0.77).(16) This supports the idea that different constructs were tested with the 9HPT and AMSQ. This is in line with our finding that combinations of movements explained different proportions of the variances of the 9HPT and AMSQ. The difference of correlation coefficients between left vs. right and non-dominant vs. dominant hand may be explained with the AMSQ being a measure of perceived upper extremity ADL tasks. Objective impairment of the dominant hand, which is most frequently the right hand, probably influences perceived UEF more strongly.

With regard to the assessment of mobility, the T25WT was chosen, because it has good psychometric properties to assess ambulatory function.(3) It is primarily a measure of walking speed, which seems clinically relevant, because walking speed relates to the capacity to perform outdoor activities important in daily life(18) and employment status.(19) However, since walking speed is often preserved in less disabled patients, measures of walking distance or endurance can better used for these patients.

The movements used to assess mobility in in Assess MS explained 70.8% of the variance of the T25WT. In previous studies, the T25WT correlated with the ability to perform ADL tasks(18), which is in line with our finding that STS contributed most to the variance. Furthermore, there are similarities between STS and the Timed Up & Go test (in which a patient gets up from a chair), which correlated strongly with the T25WT.(3)

The GAT also contributed significantly to the variance of the T25WT. Although these tests are very similar, GAT is principally a qualitative measure of ambulation (i.e. “how well does a patient walk?”) and the T25WT only measures walking speed. However, the relation of walking speed with spatial and temporal gait parameters has been previously described.(12)

A strong point of our study is the use of a combination of simple movements to assess UEF and mobility that can be performed in a short time, that can easily be done in clinical setting. Our study has some limitations. Firstly, patients included in our cohort were relatively mildly disabled with a median EDSS of 3.0, and this hampers generalisation to a more disabled population. This is also reflected in the distribution of assessments of the movements (table 3). Results of our models might have been different if more severely disabled

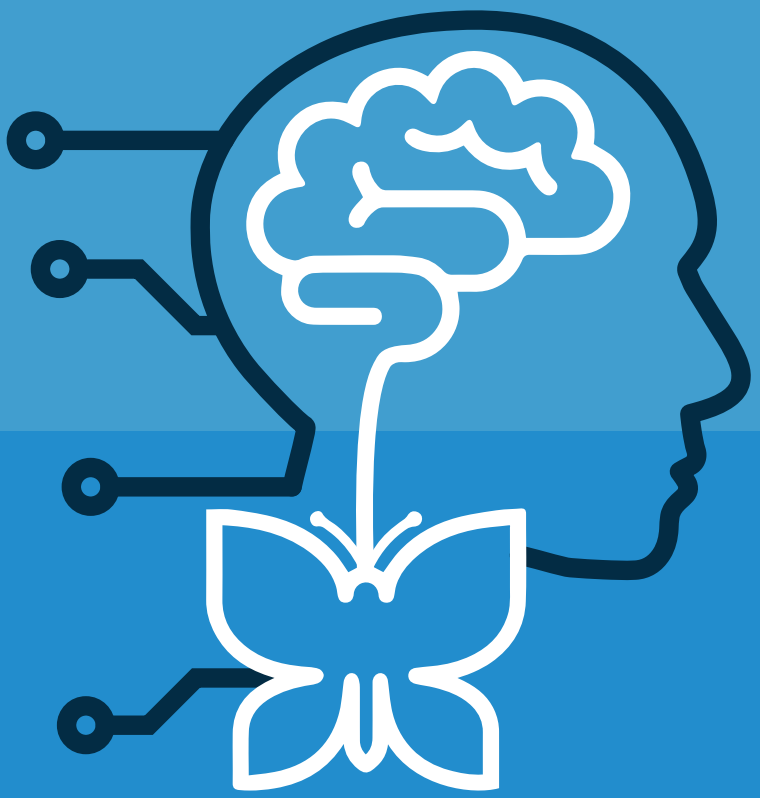
patients were included. This would particularly account for the T25WT, because of its limited sensitivity to detect abnormalities in patients with mild ambulatory impairment.(19) For these patients, it may be more appropriate to assess walking endurance with longer walking distances (e.g. with the 6-minute walking test).(20) Secondly, our construct of mobility is probably not entirely covered with the T25WT. Our construct includes standing up from a chair, turning on a spot, walking a straight line and the Romberg test. With the T25WT, only the time that a patient walks straight for a distance of 25 foot is measured. This explains why TRW and TOS did not contribute significantly to the model. Using another measure than only the T25WT as surrogate for the construct of mobility, would have likely given different results. Lastly, the rating scales of the ADL tasks have not been validated yet. Future research should consider the assessment of psychometric properties of these tests, such as validity and reliability. Nevertheless, the neurologists who performed the video rating, experienced that the ADL scales were much easier to apply than the scales derived from the Neurostatus-EDSS.

We conclude that UEF and mobility can be assessed with a combination of standardized movements. ADL tasks contributed most to these assessments, which indicates that including ADL tasks (such as drinking from a cup and standing up from a chair) in daily clinical practice, may be more valuable than the classical neurological examination (such as placing a finger on one's nose) to assess UEF and mobility. Also, incorporating ADL tasks in clinical trials may be valuable to assess motor functioning. Future research will have to determine whether these ADL movements have all the other psychometric properties that would make them valuable clinical assessments.

REFERENCES

1. van Munster CE, Uitdehaag BM. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs. 2017;31(3):217-36.
2. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult Scler. 2017;23(5):711-20.
3. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Mult Scler. 2017;23(5):704-10.
4. Ontaneda D, LaRocca N, Coetzee T, Rudick R, Force NMT. Revisiting the multiple sclerosis functional composite: proceedings from the National Multiple Sclerosis Society (NMSS) Task Force on Clinical Disability Measures. Mult Scler. 2012;18(8):1074-80.
5. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BM. The Arm Function in Multiple Sclerosis Questionnaire (AMSQ): development and validation of a new tool using IRT methods. Disabil Rehabil. 2015:1-7.
6. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors. 2015;2(1):e11.
7. Holper L, Coenen M, Weise A, Stucki G, Cieza A, Kesselring J. Characterization of functioning in multiple sclerosis using the ICF. J Neurol. 2010;257(1):103-13.
8. Johansson S, Ytterberg C, Claesson IM, Lindberg J, Hillert J, Andersson M, et al. High concurrent presence of disability in multiple sclerosis. Associations with perceived health. J Neurol. 2007;254(6):767-73.
9. Mansson E, Lexell J. Performance of activities of daily living in multiple sclerosis. Disabil Rehabil. 2004;26(10):576-85.
10. Yozbatiran N, Baskurt F, Baskurt Z, Ozakbas S, Idiman E. 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(1-2):117-22.
11. Larocca NG. Impact of walking impairment in multiple sclerosis: perspectives of patients and care partners. Patient. 2011;4(3):189-201.
12. Motl RW. Ambulation and multiple sclerosis. Phys Med Rehabil Clin N Am. 2013;24(2):325-36.
13. Kappos L, D'Souza M, Lechner-Scott J, Lienert C. On the origin of Neurostatus. Mult Scler Relat Disord. 2015;4(3):182-5.
14. Advait Sarkar CM, Jonas F. Dorn, Rishi Bedi, Saskia Steinheimer, Jacques Boisvert, Jessica Burggraaff, Marcus D'Souza, Peter Kontschieder, Samuel Rota Bulo, Lorcan Walsh, Christian P. Kamm, Yordan Zaykov, Abigail Sellen, Sian Lindley. Setwise Comparison: Consistent, Scalable, Continuum Labels for Computer Vision. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems Pages. 2016:261-71.
15. Lamers I, Kelchtermans S, Baert I, Feys P. Upper limb assessment in multiple sclerosis: a systematic review of outcome measures and their psychometric properties. Arch Phys Med Rehabil. 2014;95(6):1184-200.

16. van Leeuwen LM, Mokkink LB, Kamm CP, de Groot V, van den Berg P, Ostelo R, et al. Measurement properties of the Arm Function in Multiple Sclerosis Questionnaire (AMSQ): a study based on Classical Test Theory. Disabil Rehabil. 2017;39(20):2097-104.
17. Lamers I, Cattaneo D, Chen CC, Bertoni R, Van Wijmeersch B, Feys P. 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(1):65-75.
18. Kempen JC, de Groot V, Knol DL, Polman CH, Lankhorst GJ, Beckerman H. Community walking can be assessed using a 10-metre timed walk test. Mult Scler. 2011;17(8):980-90.
19. Goldman MD, Motl RW, Scagnelli J, Pula JH, Sosnoff JJ, Cadavid D. Clinically meaningful performance benchmarks in MS: timed 25-foot walk and the real world. Neurology. 2013;81(21):1856-63.
20. Goldman MD, Marrie RA, Cohen JA. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. Mult Scler. 2008;14(3):383-90.



CHAPTER 5

Minimal clinically important difference of improvement on the arm function in multiple sclerosis questionnaire (AMSQ)

Caspar E.P. van Munster, Levent Kaya,
Morgan Obura, Nynke F. Kalkers, Bernard M.J Uitdehaag.

Multiple Sclerosis Journal. 2020 Apr; 26(4):505-508

ABSTRACT

Background
The Arm function in Multiple Sclerosis Questionnaire (AMSQ) has been developed to assess upper extremity function of patients with multiple sclerosis (MS). A minimal clinically important difference (MCID) value has not been determined yet.

Objective
To determine an MCID for AMSQ.

Methods
We used the sensitivity- and specificity-based approach with dichotomized global perceived effect as an anchor.

Results
The receiver operating characteristic curve yielded an optimal threshold value of 14.5 (sensitivity 0.68 and specificity 0.79). The area under the ROC curve value was 0.77.

Conclusion
We identified an MCID of 15 points for the AMSQ (range 31 – 186).

1 INTRODUCTION

Traditionally, the focus of clinical assessment in multiple sclerosis (MS) has been on ambulation. However, other domains are being increasingly assessed in conjunction with ambulation. This includes the assessment of upper extremity function (UEF). Various measures are available of which the 9-hole peg test (9HPT) is considered as the gold standard for manual dexterity.(1) Nevertheless, a performance-based measure such as the 9HPT, does not provide any insight into the patient perspective of UEF. For this purpose, patient reported outcome measures (PROM) are valuable tools. To date, only one PROM is available that has been specifically developed to assess UEF in MS patients: the Arm function in Multiple Sclerosis Questionnaire (AMSQ).(2) The AMSQ is an unidimensional 31-item questionnaire with good psychometric properties.(3)

However, a minimal clinically important difference (MCID) has not been reported yet. An MCID defines the smallest amount of change on a scale that is important or meaningful to a patient.(4) Determining an MCID of a PROM is important, because a given change on the score generally does not have an obvious clinical importance to the clinician. The objective of this study is to determine an MCID of improvement on the AMSQ.

2 METHODS

Data were derived from patients that have been treated with fampridine. Fampridine increases axonal conduction velocity by selectively blocking potassium channels, which may lead to improvement of various motor functions, including UEF.(5) Effects generally occur within two weeks of treatment. Therefore, patients treated with fampridine are good subjects to assess change in AMSQ, and determine an MCID.

Patients

Patients were recruited in the VU Medical Centre in Amsterdam from an outpatient clinic that was specifically organised to assess eligibility for and efficacy of treatment with fampridine. All patients provided written informed consent prior to inclusion, and the study was approved by the local ethics committee. Patients were considered eligible if they complied to the official treatment label of fampridine.(6) Demographical data and MS characteristics, including an Expanded Disability Status Scale (EDSS), were collected for this study.

Arm function in Multiple Sclerosis Questionnaire (AMSQ)

Patients were asked to complete the AMSQ before treatment and during a follow up visit after a minimum of two weeks of treatment. The AMSQ consists of 31 items concerning activity limitations due to hand and arm functioning. A patient assigns a number to each item on a six-point Likert-scale ranging from "*not at all*" to "*no longer able to*". The sum score ranges from 31 to 186, with a higher score indicating more impairment. Change in the sum score was calculated by subtracting the AMSQ score of the follow up visit from the baseline value. Consequently, a positive change score indicates an improvement of UEF, and conversely a negative change score worsening of UEF. Questionnaires with more than two missing items were excluded from analysis. If one or two items were missing, the average of the other items was calculated and used as substitutes.

Determining the MCID

We used the sensitivity- and specificity anchor-based method to determine an MCID. In short, with an anchor-based approach the change in PROM score is being compared with change of another measure that is understandable and is considered as an anchor or external criterion.(7) As anchor, we used a global perceived effect (GPE) score that specifically addressed change in UEF and consisted of a five-point Likert-scale, including "*much deteriorated*" (1), "*deteriorated*" (2), "*unchanged*" (3), "*improved*" (4) and "*much improved*" (5). The GPE was determined by the treating physician on the follow-up visit by asking the patient how much the UEF was changed since the baseline visit. Because we used the sensitivity- and specificity approach, the GPE scores were dichotomized into "*improved*" or "*unchanged*". Since we wanted to address improvement of UEF, we excluded scores 1, 2 and 5 to minimize the impact of these scores on the MCID value.(8) With this method the GPE is considered the gold standard, and the AMSQ as a diagnostic test for which the sensitivity and specificity to discriminate between "*improved*" and "*unchanged*".(9) A receiver operating characteristics (ROC) curve was used to determine the MCID i.e. the AMSQ score that produces the greatest combined sensitivity and specificity (determined with the highest Youden's index) . Additionally, the area under the ROC (AUROC) was determined. This value represents the probability that scores will correctly discriminate between "*improved*" and "*unchanged*" UEF.(7) A value of 0.7 to 0.8 was considered acceptable and 0.8 to 0.9 excellent.(10) The correlation between AMSQ change and GPE was determined using Spearman's rank-order correlation statistics.

To investigate the statistical properties of the underlying distribution of change scores of the AMSQ, we calculated the Standard Error of Measurement (SEM).The SEM was calculated by multiplying the baseline standard deviation by the square root of one minus its reliability coefficient (i.e. the intraclass correlation coefficient).

Statistical analyses were performed in IBM SPSS Statistics for Macintosh, Version 24.

3 RESULTS

Data from 223 patients were analysed. The mean age was 51.3 years (standard deviation 10.5), with 57.4% females. Most patients had a progressive disease type (56.5%). The median (interquartile range) for disease duration was 11.4 years (4.4 – 16.6), and for EDSS was 6.0 (4.0 - 6.5). The correlation coefficient between AMQ and GPE was 0.37 ($p < 0.001$). The AMSQ thresholds from the ROC curve with corresponding sensitivity and specificity are displayed in table 1.

Table 1, threshold values from receiver operating characteristic curve with corresponding sensitivity, specificity and Youden's index.

Threshold value	Sensitivity	1 - Specificity	Youden's index
0.5	0.868	0.581	0.287
1.5	0.868	0.531	0.337
2.5	0.842	0.494	0.348
3.5	0.789	0.444	0.346
4.5	0.789	0.419	0.371
5.5	0.763	0.388	0.376
6.5	0.763	0.356	0.407
7.5	0.737	0.325	0.412
8.5	0.737	0.300	0.437
9.5	0.737	0.281	0.456
10.5	0.711	0.281	0.429
11.5	0.711	0.263	0.448
12.5	0.711	0.250	0.461
13.5	0.684	0.225	0.459
14.5	**0.684**	**0.213**	**0.472**
15.5	0.632	0.206	0.425
16.5	0.579	0.206	0.373
17.5	0.579	0.188	0.391
18.5	0.579	0.175	0.404
19.5	0.553	0.156	0.396
20.5	0.553	0.144	0.409
21.5	0.474	0.144	0.330
22.5	0.447	0.144	0.304
23.5	0.447	0.138	0.310
24.5	0.421	0.131	0.290
25.5	0.342	0.119	0.223
26.5	0.342	0.106	0.236
27.5	0.342	0.094	0.248

Threshold value	Sensitivity	1 - Specificity	Youden's index
29.5	0.316	0.081	0.235
32.0	0.289	0.081	0.208
34.0	0.289	0.075	0.214
35.5	0.289	0.069	0.221
37.0	0.289	0.063	0.227
38.5	0.289	0.056	0.233
39.5	0.289	0.050	0.239
40.5	0.263	0.050	0.213
43.0	0.237	0.044	0.193
48.0	0.211	0.044	0.167
51.5	0.211	0.038	0.173
54.5	0.184	0.031	0.153
60.0	0.158	0.031	0.127
63.5	0.158	0.025	0.133
66.0	0.132	0.025	0.107
69.0	0.132	0.019	0.113
71.5	0.105	0.019	0.087
75.0	0.053	0.019	0.034
79.5	0.053	0.013	0.040
82.5	0.026	0.013	0.014
84.5	0.026	0.006	0.020

A threshold value of 14.5 yielded the highest sensitivity (0.68) and specificity (0.79). The ROC curve is shown in figure 1. The AUROC value was 0.77. The SEM was 13.0.

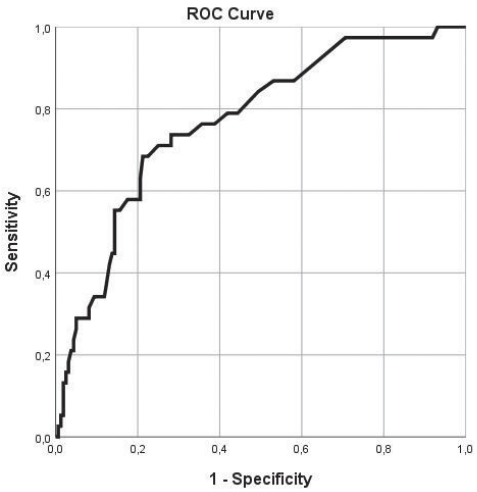


Figure 1, receiver operating characteristics (ROC) curve.

4 DISCUSSION

We found an MCID value of 14.5 with a sensitivity of 0.68,a specificity of 0.79, and an acceptable AUROC value. Since the AMSQ has no decimals, we rounded the threshold to 15 points.

There is no consensus about the preferred threshold value that determines sensitivity and specificity. Mostly, the threshold is chosen that jointly maximizes sensitivity and specificity in order to have the lowest overall misclassification. (11) We used this rational to determine the threshold.

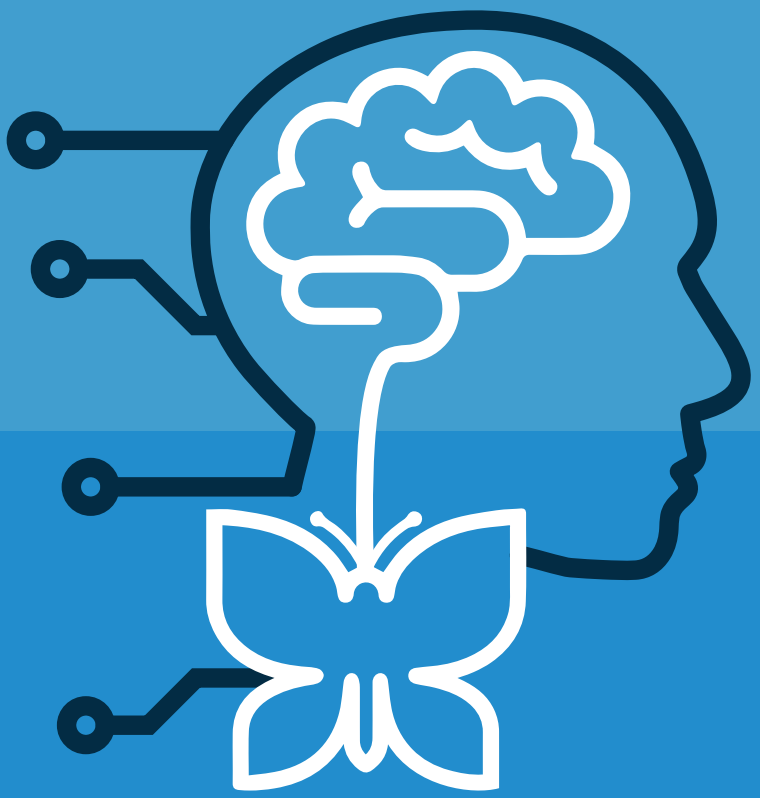
There is a certain degree of uncertainty in our findings. This is reflected in a weak, albeit sufficient, (8) correlation between AMSQ and GPE, a moderate sensitivity and specificity and rather similar threshold values around the MCID value. Therefore, our findings will have to be confirmed in future studies.

This is the first study to determine an MCID value for the AMSQ. The strength of our study lies in the large sample size and the normal distribution of AMSQ results. However, our study also has some limitations. Firstly, we used only one method to determine an MCID while there are others available.(4, 7) Secondly, we used a subjective anchor that is prone to recall bias in which case a patient may have recalled answers given at baseline that have subsequently influenced completion of the questionnaire at follow-up. Therefore, an MCID should also be assessed with an objective anchor for UEF, such as the 9HPT. Lastly, a placebo effect might have influenced our findings. This accounts particularly for the GPE, since a patient may have experienced improvement of UEF, while no improvement has been noticed on the ability to perform a task of UEF (as assessed with AMSQ). Inclusion of other measures that assess capacity of UEF objectively, such as the 9HPT, allows more certainty and magnitude of true improvement of UEF. Furthermore, additional objective measures contribute to a more detailed description of UEF of a group of patients.

In conclusion, our MCID estimate for AMSQ is 15 points based on a sensitivity- and specificity anchor-based method. Future studies should investigate reproducibility of this finding with similar and other methods, in a cohort with extensive assessment of different domains of UEF.

REFERENCES

1. van Munster CE, Uitdehaag BM. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs. 2017;31(3):217-36.
2. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BM. The Arm Function in Multiple Sclerosis Questionnaire (AMSQ): development and validation of a new tool using IRT methods. Disabil Rehabil. 2015:1-7.
3. van Leeuwen LM, Mokkink LB, Kamm CP, de Groot V, van den Berg P, Ostelo R, et al. Measurement properties of the Arm Function in Multiple Sclerosis Questionnaire (AMSQ): a study based on Classical Test Theory. Disabil Rehabil. 2017;39(20):2097-104.
4. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA. 2014;312(13):1342-3.
5. Savin Z, Lejbkowicz I, Glass-Marmor L, Lavi I, Rosenblum S, Miller A. Effect of Fampridine-PR (prolonged released 4-aminopyridine) on the manual functions of patients with Multiple Sclerosis. J Neurol Sci. 2016;360:102-9.
6. [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/contacts/not_yet_available.pdf.
7. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7(5):541-6.
8. de Vet HC, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RW, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. J Clin Epidemiol. 2010;63(1):37-45.
9. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis. 1986;39(11):897-906.
10. Hosmer DW LS. Applied logistic regression: New York: Wiley, Inc; 2000.
11. Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test-based threshold change value for an outcome measure: resolving the conflict. Phys Ther. 2012;92(10):1338-47.



CHAPTER 6

Responder rates to fampridine differ between clinical subgroups of multiple sclerosis patients and patient reported outcome influences treatment decision

Caspar E.P. van Munster, Levent Kaya, Ka-Hoo Lam,
Nynke F. Kalkers, Joep Killestein, Bernard M.J. Uitdehaag.

Multiple Sclerosis and Related Disorders. 2020 Feb; 38

ABSTRACT

Background
Fampridine is an effective treatment to improve ambulation for some multiple sclerosis (MS) patients. Remarkable discrepancies exist between responder rates in clinical trials and the proportion of patients continuing treatment in clinical practice. This may be related to clinical phenotypes of MS patients, and the influence of patient reported outcome (PRO) on treatment decision making.

Objective
To analyse responder rates to fampridine on ambulation and upper extremity function (UEF) and the influence on treatment decision making in different clinical subgroups in a real-world setting.

Methods
MS patients with ambulatory impairment treated with fampridine were included. Patients were subdivided based on disease duration, clinical phenotype, Expanded Disability Status Scale (EDSS), baseline walking speed, and presence of UEF impairment. Ambulatory response was assessed with the Timed 25-Foot Walk (T25FW, responder defined as ≥20% improvement) and with the MS Walking Scale (MSWS, responder defined as ≥8 points improvement) as a PRO. For patients also reporting impaired UEF, the Arm Function in MS Questionnaire (AMSQ, responder defined as ≥15 improvement) was the PRO of choice. Decision on treatment continuation was based on improvement of T25FW, MSWS and the clinicians' overall impression for improvement.

Results
In total 344 patients were included of which 75.3% continued treatment. More patients with a relapsing clinical phenotype continued treatment versus patients with a progressive phenotype (83.6 versus 68.6%, $p < 0.01$). A positive linear trend was found between severity of walking disability, as determined by baseline walking speed, and T25FW response ($p < 0.01$), while there was an inverse linear association between walking disability and MSWS response ($p = 0.03$). However, the proportion of patients continuing treatment was similar between subgroups of baseline walking speed. Impaired UEF was reported by 183 (66.5%) patients, of which 64 (39.3%) were AMSQ responders. Patients responding on AMSQ compared to non-responders, were also more frequently MSWS responders (82.8 vs 65.3%, $p = 0.02$), while response on T25FW was similar, and continued treatment more often (85.9 vs 70.7%, $p = 0.04$). This suggests an influence of PRO on treatment decision making.

Conclusion
Responder rates and treatment continuation of fampridine differed between clinical subgroups of MS. PROs influenced treatment decision making of fampridine in clinical practice, particularly in patients with mild ambulatory impairment or those reporting UEF impairment. To some extent, these findings explain discrepancies found between clinical trials and clinical practice, and support the importance of subgroup analyses and incorporation of PROs in clinical trials. For clinical practice, using PROs to assess patients experience in conjunction with performance measures helps in treatment decision making.

1 INTRODUCTION

Many patients with multiple sclerosis (MS) experience some level of ambulatory impairment during the course of the disease, and this can already be a prominent factor leading to early disability. With progression of ambulatory impairment, preservation of upper extremity function (UEF) becomes more important to maintain mobility and the ability to perform activities of daily living (ADL). Also, UEF correlates with quality of life,(1, 2) and employment status.(3) Therefore, preservation of UEF is a clinically relevant treatment goal during all disease stages.

Fampridine improves motor functioning in MS by selectively blocking potassium channels which increases axonal conduction velocity.(4) In a pooled analysis of two phase 3 trials of fampridine, walking speed, as assessed with the Timed 25-Foot Walk test (T25FW), improved ≥20% in 37.6% of patients. (5) Also, ambulation improved as reported by patients (assessed with the Multiple Sclerosis Walking Scale (MSWS)) compared with placebo. Treatment responses on UEF were less consistent, which may be due to heterogeneous study cohorts.(6-12)

Despite the approval by the European Medicines Agency (EMA) for treatment of ambulatory impairment in MS patients, in various European countries fampridine is not being reimbursed because the average group effects have been considered insufficient when balanced with the costs. In clinical practice however, patients may also report improvement in motor functions other than ambulation, such as UEF. In this way, clinicians are confronted with patients that benefit exceptionally, in which cases treatment might indeed be cost-effective.

Analyses of subgroups may facilitate identification of patients that benefit most from this treatment. In a pooled analysis of the phase 3 trials no differences were found in the proportion of responders (defined as an improvement of ≥20% on the T25FW) between subgroups of MS phenotype, Expanded Disability Status Scale (EDSS), disease duration, and baseline

T25FW performance.(5) However, patients treated in a real-world setting are always more heterogeneous than in trials with strict in- and exclusion criteria, which limits the external validity of these findings.(13)

We identified four studies with data of fampridine treatment from a real-world setting.(12, 14-16) Findings in these studies generally suggested that fampridine response should be assessed with multiple measures for ambulation. The physician's overall judgement of improvement was much larger (i.e. >70%) than was captured with a single ambulation test.(12, 14) Patient reported improvement was greater than improvement measured objectively, although we hypothesize this is partly related to a placebo-effect. Also, other MS-related symptoms improved, including UEF and fatigue. Only one of these studies stratified patients according to baseline walking speed (cut-off value T25FW of 8 seconds), and found better response on T25FW in the lower baseline walking speed.(12)

Although findings from these studies are valuable, the results offer insufficient insight into clinically relevant responses in subgroups of patients. A statistically significant difference does not necessarily mean that it is clinically relevant for a patient. For this, clinically relevant benchmarks and minimal clinically important differences have been determined.(17-21) These can be used as cut-off values to differentiate responders from non-responders to fampridine, and can subsequently be used to compare between subgroups. Furthermore, there is no data available of patient reported outcome (PRO) on UEF.

Therefore, in this study of patients from a real-world setting, we analysed fampridine response on ambulation according to clinically relevant responder definitions in various subgroups of MS patients, and PRO on UEF in patients also reporting problems in UEF. Furthermore, we aim to give insight into treatment decision making in clinical practice.

2 METHODS

Patients

Patients initiating fampridine treatment at VU University Medical Centre in Amsterdam were recruited. The study protocol was approved by the local ethics committee. All patients provided written informed consent prior to inclusion. Patients were enrolled between April 2016 and May 2017. As standard practice in the Netherlands, patients were assessed for eligibility and efficacy for treatment with fampridine according to the EMA recommendations,(22) which is adopted in the official treatment label of fampridine. Inclusion criteria were consistent with the therapeutic indication

as stated in the label: diagnosis of MS, minimum age of 18 years, and ambulatory impairment defined as an EDSS score between 4.0 and 7.0. Contraindications were handled as exclusion criteria.(22) As is recommended by the EMA, patients were assessed during a baseline and follow-up visit, planned after a minimum of two weeks of treatment. Demographical and disease characteristics were recorded, including an estimated EDSS by the clinician at baseline. Treatment (dis)continuation was recorded, which was based on T25FW performance, MSWS score and clinicians' best judgement of overall improvement of ambulation, as was instructed to the physician beforehand. Treatment was discontinued if benefit was not reported by patients. Of note, the EMA only recommends using the T25FW or MSWS to assess ambulation.

Assessment of ambulation

Ambulation was assessed at each visit with the T25FW and the MSWS. The T25FW was performed twice on both visits, as implemented in the Multiple Sclerosis Functional Composite (MSFC).(23) The shortest time to complete a trial was used in the analyses. The MSWS is a 12-item questionnaire with a five-point Likert-scale ranging from "not at all" to "extremely". It assesses the perceived impact of MS-related walking impairment across a range of functional activities.(24) A total score ranging from 12 to 60 was calculated and transformed to a 0 to 100 scale, in which a higher score indicates worse perceived ambulatory function. In cases with one or two items missing, the mean of all other items was used as substitute score. Cases with more missing items were excluded from analysis. Patients were dichotomized into ambulatory responders and non-responders to fampridine based on T25FW and MSWS. A T25FW responder was defined as an improvement of ≥20% at follow-up visit.(17) A MSWS responder was defined as an improvement of ≥8 points on the transformed scale at follow-up.(20)

Assessment of patients' perspective on UEF

UEF was assessed on each visit with the Arm Function in Multiple Sclerosis Questionnaire (AMSQ). The AMSQ is a PRO that has been designed specifically for MS patients for assessment of activity limitations due to impaired UEF. (25) The AMSQ consists of 31 items with a six-point Likert-scale ranging from "not at all" to "no longer able to." The sum score ranges from 31 to 186, with a higher score indicating more impairment. In cases with one or two items missing, the mean of all other items was used as substitute score. If more than two items were missing, the questionnaire was excluded from further analysis. The responder criterion for the AMSQ was defined as an improvement of ≥15 points at follow-up.(26)

Subgroup analyses

The proportion of responders on ambulation measures were compared between various subgroups. Clinical phenotype was dichotomized into relapsing-remitting and progressive types of MS. Disease duration was stratified into the following three subgroups: short (0 – 10 years), intermediate (10 – 20 years) and long (≥20 years). Disability severity was categorised into two subgroups according to EDSS score, as was defined as mild disability (EDSS 4.0 - 5.5) and severe disability (EDSS 6.0 - 7.0). Furthermore, the baseline T25FW performance in seconds was stratified into three clinically different ambulation subgroups: fully ambulatory (< 6 seconds), mild walking disability (6 – 8 seconds) and severe walking disability (≥ 8 seconds). The cut-off values represents clinically meaningful performance benchmarks: a T25FW performance of 6 to 8 seconds was associated with a change in occupation due to MS, occupational disability, walking with a cane, and needing "some help" with instrumental ADL; while a T25FW of ≥8 seconds was associated with collecting Supplemental Security Income and government health care, walking with a walker, and inability to do instrumental ADL.(27) The effect on patients perspective on UEF was assessed in a subset of patients reporting impaired UEF at baseline visit, which was determined by the physician by asking the patient if they experienced any impairment of UEF (answers "yes" or "no"). Subgroup analyses in this subset of patients were performed for clinical phenotype, disease duration and EDSS.

Data analysis

All statistical data analyses were performed using the Statistical Package for Social Science (SPSS) version 22 (SPSS, Chicago, IL). All p-values were 2-tailed and statistical significance was assumed for $p < 0.05$ in all tests. Histograms and q-q plots were used to assess normality of distribution. Normally distributed data were analysed with the paired t-tests and were reported with mean values and standard deviation (SD). Non-normally distributed data were analysed using the Wilcoxon rank test for matched pairs and were reported with median values and the interquartile range (IQR). Subgroup analyses were performed and checked for statistical significance with chi-square tests. For disease duration and T25FW performance on baseline subgroups, a chi-square (Cochran-Armitage) test for trend was used.

3 RESULTS

In total 344 patients were assessed for eligibility. Sixty-nine patients were excluded from analyses because of various reasons as summarized in figure 1. In total 275 patients were included for analyses.

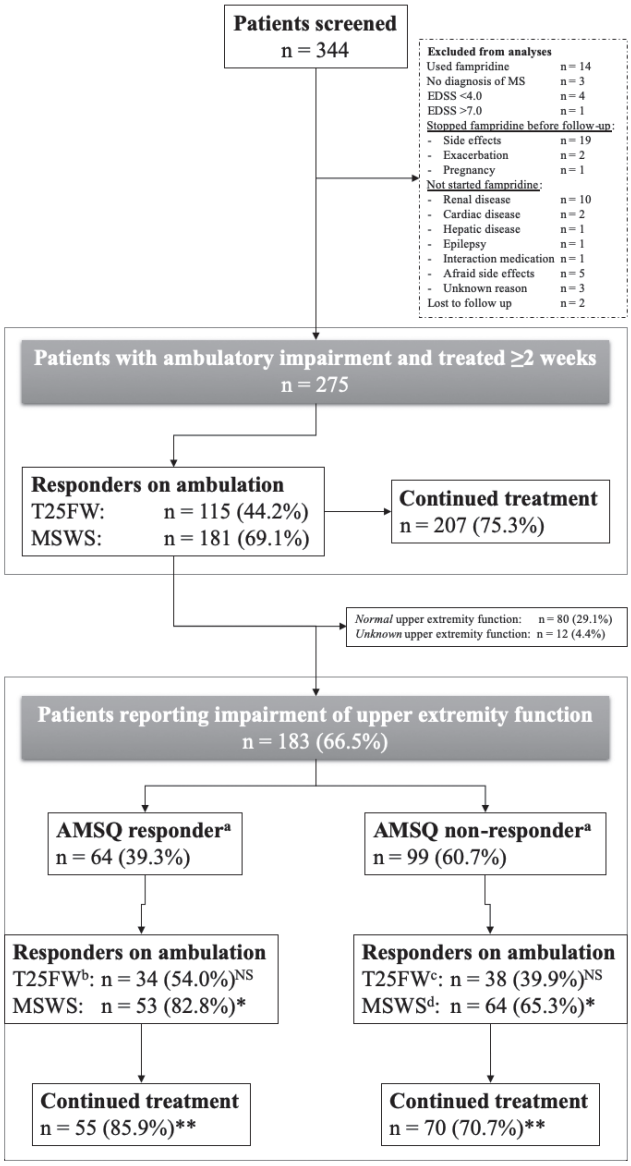


Figure 1, patient selection and schematic display of proportions of responders and continuation of treatment in patients with ambulatory impairment and patients reporting upper extremity impairment. Abbreviations: EDSS = Expanded Disability Status Scale; T25FW= timed 25-foot walk; MSWS = multiple sclerosis walking scale; Arm Function in MS Questionnaire (AMSQ); NS: not statistically significantly different; T25FW responder defined as an improvement of ≥20% at follow-up; MSWS responder defined as an improvement of ≥8 points at follow-up; AMSQ responder defined as an improvement of ≥15 points at follow-up; *: p = 0.02; **: p = 0.04; [a]measurements of 20 patients were missing; [b]measurement of 1 patient was missing; [c]measurements of 3 patients was missing; [d]measurement of 1 patient was missing.

Total cohort of patients with ambulatory impairment

Demographic and MS characteristics of the total cohort are presented in table 1. The median follow-up was 21 days (IQR 17 – 25). The mean age was 50.8 years (SD 10.6). The majority of patients were female (58.9%) and had a progressive MS phenotype (55.6%). The median disease duration was 10.3 years (IQR 5.0 – 16.6). The median EDSS was 6.0 (IQR 4.5 - 6.5). Both ambulation measures showed statistically significant improvement at follow-up ($p < 0.01$). The proportion of responders was higher when defined by the MSWS (69.1%) than by the T25FW (44.2%). In total 207 (75.3%) of patients continued treatment (figure 1).

Table 1, baseline characteristics and ambulation measures at baseline and follow-up (n = 275).

Demographic data & MS characteristics				
Age (in years)[a]	50.8 (10.6)			
Gender (female/ male)	162 / 113			
Disease phenotype (relapsing/ progressive)	122 / 153			
Disease duration (in years)[b]	10.3 (5.0 – 16.6)			
EDSS[b]	6.0 (4.5 – 6.5)			
Ambulation measures				
	Baseline	*Follow-up*	*P-value*	*Responders, n (%)[c]*
T25FW (in seconds)[b]	7.6 (5.9 – 10.6)	6.2 (5.2 – 8.1)	**<0.01**	115 (44.2)[d]
MSWS[b]	81 (69 – 90)	56 (42 – 75)	**<0.01**	181 (69.1)[e]

Abbreviations: Relapsing = relapsing-remitting multiple sclerosis; Progressive = secondary and primary progressive multiple sclerosis; EDSS = Expanded Disability Status Scale; T25FW= timed 25-foot walk; MSWS = multiple sclerosis walking scale; [a]Data are mean with standard deviation for normally distributed variables; [b]Because of non-normal distribution, median and interquartile range are provided; [c]T25FW responder defined as an improvement of ≥20% at follow-up, MSWS responder defined as an improvement of ≥8 points at follow-up; [d]measurements of 15 patients were missing; [e]measurements of 13 patients were missing; P-values in bold represent significant values.

The proportions of responders on the ambulation measures in subgroups are displayed in table 2. Only stratification by clinical phenotypes showed a significant difference between the proportion of patients continuing treatment: 83.6% of relapsing vs 68.6% of progressive patients. Stratification by disease duration or clinical phenotype did not show any statistically significant difference in proportion of responders. When stratified by EDSS subgroups, a higher proportion of T25FW responders was found in the subgroup with severe (EDSS 6.0 – 7.0) than mild (EDSS 4.0 – 5.5) disability (51.4 vs 36.4%, $p = 0.02$). Comparison of subgroups stratified by baseline T25FW performance revealed a positive linear trend ($p < 0.01$) between

walking disability severity and T25FW response: there were 18.3, 42.7, and 61.4% responders in the fully ambulatory, mild walking disability, and severe walking disability groups, respectively. However, an inverse trend however, was found for proportion of MSWS responders (p = 0.03): 81.2%, 69.0%, and 65.8% in the fully ambulatory, mild walking disability, and severe walking disability groups, respectively.

Table 2. proportion of responders on ambulation measures in subgroups (n = 275)[a].

Disease duration				
	0 – 10 yrs. (n = 137)	*11 – 20 yrs. (n = 85 / 87)[a]*	*≥21 yrs. (n = 38)*	*P-value*
T25FW responders, n (%)[b]	56 (40.9)	42 (49.4)	17 (44.7)	0.41
MSWS responders, n (%)[c]	99 (72.3)	52 (59.8)	30 (78.9)	0.91
Continued treatment, n (%)	107 (75.4)	69 (74.2)	31 (77.5)	0.88
Clinical phenotype				
	Relapsing (n = 116)		*Progressive (n = 144 / 146)[a]*	*P-value*
T25FW responders, n (%)[b]	51 (44.0)		64 (44.4)	1.00
MSWS responders, n (%)[c]	82 (70.7)		99 (67.8)	0.71
Continued treatment, n (%)	102 (83.6)		105 (68.6)	**<0.01**
Expanded Disability Status Scale				
	EDSS 4.0 – 5.5 (n = 121 / 117)[a]		*EDSS 6.0 – 7.0 (n = 138 / 144)[a]*	*P-value*
T25FW responders, n (%)[b]	44 (36.4)		71 (51.4)	**0.02**
MSWS responders, n (%)[c]	87 (74.4)		93 (64.6)	0.12
Continued treatment, n (%)	97 (80.2)		110 (71.9)	0.15
T25FW at baseline				
	<6 sec. (n = 71 / 69)[a]	*6 – 7.99 sec. (n = 75 / 71)[a]*	*≥8 sec. (n = 114/ 111)[a]*	*P-value*
T25FW responders, n (%)[b]	13 (18.3)	32 (42.7)	70 (61.4)	**<0.01**
MSWS responders, n (%)[c]	56 (81.2)	49 (69.0)	73 (65.8)	**0.03**
Continued treatment, n (%)	56 (78.9)	56 (73.7)	88 (75.9)	0.70

Abbreviations: Relapsing = relapsing-remitting multiple sclerosis; Progressive = secondary and primary progressive multiple sclerosis; T25FW= timed 25-foot walk; MSWS = multiple sclerosis walking scale; EDSS = Expanded Disability Status Scale; [a] 15 T25FW and 13 MSWS measurements were missing; [b] T25FW responder defined as an improvement of ≥20% at follow-up; [c] MSWS responder defined as an improvement of ≥8 points at follow-up; P-values in bold represent significant values.

Selection of patients reporting impairment of UEF

In total 183 (66.5%) patients reported impaired UEF at baseline visit (figure 1). Demographic and MS characteristics of this subgroup are summarized in table 3.

Table 3, baseline characteristics and upper extremity function measures at baseline and follow-up of *patients reporting impairment of upper extremity function* (n = 183).

Demographic data & MS characteristics				
Age (in years)[a]	50.1 (10.1)			
Gender (female/ male)	107 / 76			
Disease phenotype (relapsing/ progressive)	79 / 104			
Disease duration (in years)[b]	10.0 (5.2 – 15.9)			
EDSS[b]	6.0 (4.5 – 6.5)			
Upper extremity function measures				
	Baseline	*Follow-up*	*P-value*	*Responders, n (%)[c]*
AMSQ[b]	74 (49 – 99)	53 (39 – 81)	**<0.01**	64 (39.3)[d]

Abbreviations: Relapsing = relapsing-remitting multiple sclerosis; Progressive = secondary and primary progressive multiple sclerosis; EDSS = Expanded Disability Status Scale; Arm Function in MS Questionnaire (AMSQ); [a]Data are mean with standard deviation for normally distributed variables; [b]Because of non-normal distribution, median and interquartile range are provided; [c] AMSQ responder defined as an improvement of ≥15 points at follow-up; [d]measurements of 20 patients were missing; P-value in bold represent a significant value.

The median AMSQ improved significantly at follow up (p <0.01), and 64 (39.3%) patients were classified as AMSQ responder. Stratification by disease duration, clinical phenotype or EDSS did not reveal any differences in proportion of AMSQ responders (table 4). However, similarly as in the total cohort, relatively more patients continued treatment with a relapsing than a progressive clinical phenotype (84.8 vs 68.3% of patients). More MSWS responders were found in the AMSQ responder group compared with the AMSQ non-responder group (82.8 vs 65.3%, p <0.02; figure 1), while the proportion of T25FW responders was similar. Furthermore, more AMSQ responders continued treatment than non-responders (85.9 vs 70.7%, p <0.04).

Table 4, proportion of responders on AMSQ in subgroups of *patients reporting impairment of upper extremity function* (n = 183)[a].

Disease duration				
	0 – 10 yrs. (n = 86)	*11 – 20 yrs. (n = 55)*	*≥21 yrs. (n = 22)*	*P-value*
AMSQ responders, n (%)[b]	31 (36.0)	26 (47.3)	7 (31.8)	0.31
Continued treatment, n (%)	72 (75.8)	49 (77.8)	17 (68.0)	0.63
Clinical phenotype				
	Relapsing (n = 67)		*Progressive (n = 96)*	*P-value*
AMSQ responders, n (%)[b]	29 (43.3)		35 (36.5)	0.48
Continued treatment, n (%)	67 (84.8)		71 (68.3)	**0.02**
Expanded Disability Status Scale				
	EDSS 4.0 – 5.5 (n = 76)		*EDSS 6.0 – 7.0 (n = 87)*	*P-value*
AMSQ responders, n (%)[b]	31 (40.8)		33 (37.9)	0.83
Continued treatment, n (%)	66 (80.5)		72 (71.3)	0.21

Abbreviations: AMSQ = Arm function in Multiple Sclerosis Questionnaire, responder defined as an improvement of ≥10 points at follow-up; UEF = upper extremity function; Relapsing = relapsing-remitting multiple sclerosis; Progressive = secondary and primary progressive multiple sclerosis; EDSS = Expanded Disability Status Scale; Arm Function in MS Questionnaire (AMSQ); [a] AMSQ measurements of 20 patients were missing; [b] AMSQ responder defined as an improvement of ≥15 points at follow-up.

4 DISCUSSION

In our cohort that was derived from a real-world setting, treatment with fampridine resulted in a statistically significant improvement of all measures of ambulation and UEF which led to continuation of treatment in the vast majority of patients, in a greater proportion of patients with a relapsing than progressive disease course. In patients with more severely impaired ambulation, a greater proportion showed improvement on walking speed than in patients with mild impairment. Conversely, the proportion of patients with MSWS improvement was larger in patients with less impaired ambulation. Furthermore, in the subset of patients with UEF impairment, the proportion of patients with improvement of AMSQ were more likely to show improvement on MSWS and continue treatment, than patients without improvement of AMSQ.

Responder rates and clinical subgroups

The proportion of T25FW responders in our study (defined as a ≥20% improvement) is in line with findings from a pooled data analyses of the

phase 3 trials of fampridine,(5) and two real-world studies.(12, 15) Another real-world study reported only 11% of patients improving ≥20% on T25FW. Since only 39 of 221 records were selected for inclusion in that study, there is a high chance of selection bias.(16) Considering response on MSWS, these studies all reported improvement on fampridine.(5, 12, 14-16) However, none used responder criteria as we used in our study which hampers comparison.

Our findings suggest that the T25FW might be a more appropriate measure for assessing response to fampridine in patients with worse ambulatory impairment, and, conversely, the MSWS for patients with less impaired ambulation. This may be explained by measurement properties of T25FW, since it is not sensitive enough to detect abnormalities in patients with mild ambulation impairment, i.e. a ceiling effect.(19) In these patients PROs, such as the MSWS, are more suitable to assess change of ambulation as is experienced in daily life. Alternatively, fampridine may just be not effective to improve walking speed in patients with relatively preserved ambulation. A similar finding concerning T25FW improvement was presented in a real-world study.(12) However, no difference of MSWS was found between subgroups of T25FW, which may be a result of using a single cut-off value for baseline T25FW performance (i.e. 8 seconds), or a more stringent definition of MSWS improvement (i.e. ≥15%).

Comparison of the subgroups of disease duration and clinical phenotype in our study did not reveal different treatment response on ambulation, and this is in line with pooled analyses of the phase 3 trials.(5) Three out of four real-world studies did not perform analyses of subgroups.(14-16)

In the subset of patients reporting impaired UEF as well, approximately 40% of patients were classified as AMSQ responders. To our knowledge, only one study assessed PRO on UEF with the ABILHAND questionnaire in a cohort that was selected on impairment in UEF.(6) The ABILHAND improved by 16% after one month, and 31% after three months of treatment. In contrast, in another study no response on ABILHAND was found.(27) This might be due not selecting patients on impaired UEF, which probably attenuated the impact on ABILHAND. Also, the ABILHAND may be less suitable to detect change of UEF in MS patients, because it was not *specifically* designed for MS patients. Small sample sizes (i.e. 25 and 26 patients) are other limitations of these studies.

Treatment decision making

In our total cohort, 75% of patients continued treatment based on the physicians' judgment, which concurs with other real-world data (ranging from 70 – 87%).(12, 14-16) In one study, fampridine response was assessed in 120 patients and 74% were classified as responders when defined as a 15%

improvement of at least one of the three walking tests (i.e. T25FW, 2-min walk test or MSWS).(12) This is a more liberate definition of response than used in pre-marketing studies, and possibly better reflects the physician's overall judgement in clinical practice. However, no data were presented on factors that influenced treatment decision making.

We found two studies in which treatment decision making was analysed. The first study included 189 patients in which the EMA recommendations for continuing treatment were compared with the physicians' overall judgement of improvement, similarly to our study.(14) They found that 70% of patients continued treatment, and that a combination of T25FW and MSWS offered the best sensitivity and specificity for determining response, based on a receiver operating characteristic (ROC) curve analyses. The other study found that baseline performance of the 6-miute walk test and T25FW predicted responder status (defined as ≥15% improvement in at least one clinical walking test) with an accuracy of 85.5%.(28) The MSWS did not contribute significantly as a predictor for response.

In our impaired UEF cohort, 39.3% of patients showed improvement of AMSQ, and in this subgroup 82.8% also reported improvement of MSWS and continued treatment in 85.9% of patients, which is a larger proportion (70.7%) than in the subgroup without AMSQ improvement. Therefore, it is likely that patient reported benefits of treatment contributed to the physicians' decision to continue treatment. Obviously, physician-based factors (such as experience, training, influenceability) and local factors (such as reimbursement policies) influence treatment decision making as well. It nevertheless illustrates the gap between clinical trials and practice, and the challenges physicians are being faced with the real-world.

The inclusion of specific PRO measures in clinical trials are valuable for enhancing information on the functional impact of various aspects of impairment.(29) In clinical practice, standardized measurement of PRO provides valuable insight into the patient perspective. For example, treatment success for a patient might be more influenced by adverse events than a physician perceives or deduces from other outcome measures. All in all, assessment of PRO is indispensable for accurate understanding of clinical aspects of a disease or treatment.

Limitations

A strength of our study is the large cohort size of patients derived from a real-world setting. Although randomized controlled trials are the gold standard, they inherently have a limited degree of external validity.(13) Another strength is that our study is the first study in which the patient's self-perceived effects of fampridine on UEF was assessed with a PRO measure specifically designed

for MS patients (i.e. the AMSQ). Also, the use of responder definitions that are based on established clinically relevant changes, gives a more robust insight into clinically meaningful treatment effects than merely statistical differences of the crude scores of the outcome measures. Our study also has some limitations. Firstly, a placebo or learning effect of the measures used cannot be ruled out without comparison with a placebo group. Secondly, we do not have long-term follow-up data to assess the persistence of efficacy, particularly on PRO measures. Thirdly, fampridine response on UEF might be underestimated due to patient selection based on the presence of ambulatory impairment. Analysis of a cohort primarily selected on impaired UEF may give different results.

5 CONCLUSIONS

We conclude that responder rates, and treatment continuation of fampridine differed between clinical subgroups of MS. PROs influenced treatment decision making of fampridine in clinical practice, particularly in patients with mild ambulatory impairment or those reporting UEF impairment. To some extent, these findings explain discrepancies found between clinical trials and clinical practice, and support the importance of subgroup analyses and incorporation of PROs in clinical trials. For clinical practice, using PROs to assess patients experience in conjunction to performance measures helps in treatment decision making.

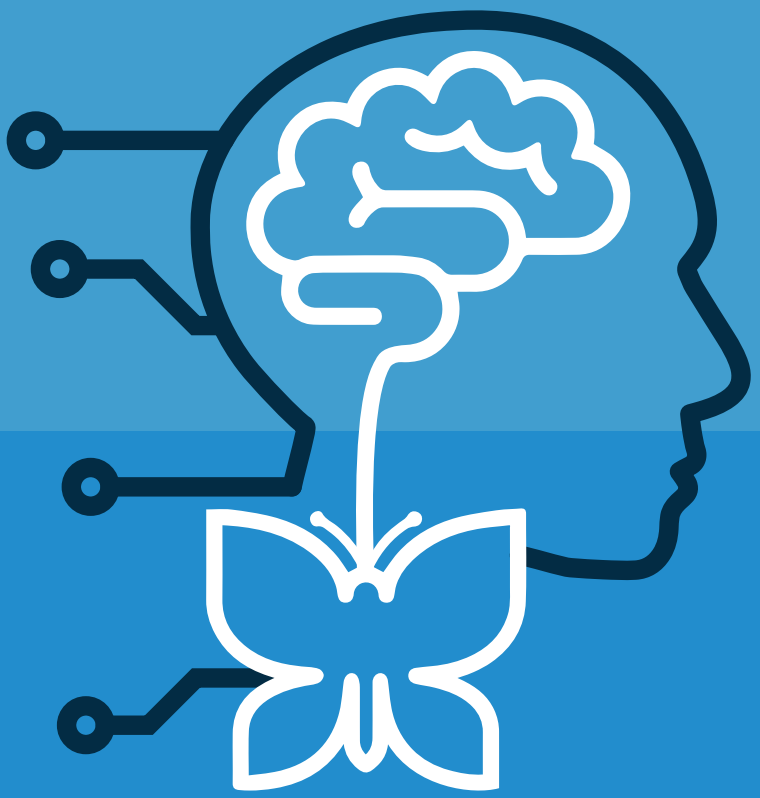
REFERENCES

1. Kamm CP, Heldner MR, Vanbellingen T, Mattle HP, Muri R, Bohlhalter S. Limb apraxia in multiple sclerosis: prevalence and impact on manual dexterity and activities of daily living. Arch Phys Med Rehabil. 2012;93(6):1081-5.
2. Salter AR, Cutter GR, Tyry T, Marrie RA, Vollmer T. Impact of loss of mobility on instrumental activities of daily living and socioeconomic status in patients with MS. Curr Med Res Opin. 2010;26(2):493-500.
3. Julian LJ, Vella L, Vollmer T, Hadjimichael O, Mohr DC. Employment in multiple sclerosis. Exiting and re-entering the work force. J Neurol. 2008;255(9):1354-60.
4. Hayes KC. Impact of extended-release dalfampridine on walking ability in patients with multiple sclerosis. Neuropsychiatr Dis Treat. 2011;7:229-39.
5. Goodman AD, Brown TR, Schapiro RT, Klingler M, Cohen R, Blight AR. A pooled analysis of two phase 3 clinical trials of dalfampridine in patients with multiple sclerosis. Int J MS Care. 2014;16(3):153-60.
6. Savin Z, Lejbkowicz I, Glass-Marmor L, Lavi I, Rosenblum S, Miller A. Effect of Fampridine-PR (prolonged released 4-aminopyridine) on the manual functions of patients with Multiple Sclerosis. J Neurol Sci. 2016;360:102-9.
7. Ruck T, Bittner S, Simon OJ, Gobel K, Wiendl H, Schilling M, et al. Long-term effects of dalfampridine in patients with multiple sclerosis. J Neurol Sci. 2014;337(1-2):18-24.
8. Pavsic K, Pelicon K, Ledinek AH, Sega S. Short-term impact of fampridine on motor and cognitive functions, mood and quality of life among multiple sclerosis patients. Clin Neurol Neurosurg. 2015;139:35-40.
9. Lo AC, Ruiz JA, Koenig CM, Anderson BM, Olson KM, Triche EW. Effects of dalfampridine on multi-dimensional aspects of gait and dexterity in multiple sclerosis among timed walk responders and non-responders. J Neurol Sci. 2015;356(1-2):77-82.
10. Jensen HB, Nielsen JL, Ravnborg M, Dalgas U, Aagaard P, Stenager E. Effect of slow release-Fampridine on muscle strength, rate of force development, functional capacity and cognitive function in an enriched population of MS patients. A randomized, double blind, placebo controlled study. Mult Scler Relat Disord. 2016;10:137-44.
11. Jensen H, Ravnborg M, Mamoei S, Dalgas U, Stenager E. Changes in cognition, arm function and lower body function after slow-release Fampridine treatment. Mult Scler. 2014;20(14):1872-80.
12. Allart E, Benoit A, Blanchard-Dauphin A, Tiffreau V, Thevenon A, Zephir H, et al. Sustained-released fampridine in multiple sclerosis: effects on gait parameters, arm function, fatigue, and quality of life. J Neurol. 2015;262(8):1936-45.
13. Mahajan R. Real world data: Additional source for making clinical decisions. Int J Appl Basic Med Res. 2015;5(2):82.
14. Rodriguez-Leal FA, Haase R, Thomas K, Eisele JC, Proschmann U, Schultheiss T, et al. Fampridine response in MS patients with gait impairment in a real-world setting: Need for new response criteria? Mult Scler. 2018;24(10):1337-46.
15. Prugger M, Berger T. Assessing the long-term clinical benefit of prolonged-release fampridine tablets in a real-world setting: a review of 67 cases. Patient Relat Outcome Meas. 2013;4:75-85.

16. Klineova S, Farber R, Friedman J, Farrell C, Lublin FD, Krieger S. Objective and subjective measures of dalfampridine efficacy in clinical practice. Mult Scler J Exp Transl Clin. 2018;4(3):2055217318786742.
17. Schwid SR, Goodman AD, McDermott MP, Bever CF, Cook SD. Quantitative functional measures in MS: what is a reliable change? Neurology. 2002;58(8):1294-6.
18. Goldman MD, Ward MD, Motl RW, Jones DE, Pula JH, Cadavid D. Identification and validation of clinically meaningful benchmarks in the 12-item Multiple Sclerosis Walking Scale. Mult Scler. 2017;23(10):1405-14.
19. Goldman MD, Motl RW, Scagnelli J, Pula JH, Sosnoff JJ, Cadavid D. Clinically meaningful performance benchmarks in MS: timed 25-foot walk and the real world. Neurology. 2013;81(21):1856-63.
20. Mehta L, McNeill M, Hobart J, Wyrwich KW, Poon JL, Auguste P, et al. Identifying an important change estimate for the Multiple Sclerosis Walking Scale-12 (MSWS-12v1) for interpreting clinical trial results. Mult Scler J Exp Transl Clin. 2015;1:2055217315596993.
21. Bosma LV, Kragt JJ, Brieva L, Khaleeli Z, Montalban X, Polman CH, 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(7):862-7.
22. [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/contacts/not_yet_available.pdf.
23. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain. 1999;122 (Pt 5):871-82.
24. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). Neurology. 2003;60(1):31-6.
25. van Leeuwen LM, Mokkink LB, Kamm CP, de Groot V, van den Berg P, Ostelo R, et al. Measurement properties of the Arm Function in Multiple Sclerosis Questionnaire (AMSQ): a study based on Classical Test Theory. Disabil Rehabil. 2017;39(20):2097-104.
26. van Munster CEPK, L; Obura, M; Kalkers, N.F.; Uitdehaag, B.M.J. minimal clinically important difference of improvement on the Arm function in Multiple sclerosis Questionnaire (AMSQ). Mult Scler. 2018(Accepted for publication 11-12-2018).
27. Pickering H, Murray J, Lin CS, Cormack C, Martin A, Kiernan MC, et al. Fampridine treatment and walking distance in multiple sclerosis: A randomised controlled trial. Clin Neurophysiol. 2017;128(1):93-9.
28. Filli L, Werner J, Beyer G, Reuter K, Petersen JA, Weller M, et al. Predicting responsiveness to fampridine in gait-impaired patients with multiple sclerosis. Eur J Neurol. 2018.
29. Tur C, Moccia M, Barkhof F, Chataway J, Sastre-Garriga J, Thompson AJ, et al. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat Rev Neurol. 2018;14(2):75-93.

PART 3

Video-assisted assessment of motor functioning



CHAPTER 7

Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis

Marcus D`Souza, Saskia Steinheimer, Jonas Dorn, C. Morrison, Jacques Boisvert, Kristina Kravalis, Jessica Burggraaff, Caspar E.P. van Munster, Manuela Diederich, A. Sellen, Christian P. Kamm, Frank Dahlke, Bernard M.J. Uitdehaag, Ludwig Kappos.

Multiple Sclerosis Journal: Experimental, Translational and Clinical. 2018; Aug 9; 4(3)

ABSTRACT

Motor dysfunction, particularly ataxia, is one of the predominant clinical manifestations in patients with multiple sclerosis (MS). Assessment of motor dysfunction suffers from a high variability. We investigated whether the clinical rating of ataxia can be improved through the use of reference videos, covering the spectrum of severity degrees as defined in the Neurostatus-Expanded Disability Status Scale. Twenty-five neurologists participated. The variability of their assessments was significantly lower when reference videos were used (SD = 0.12; range = 0.40 vs SD = 0.26; range = 0.88 without reference videos; p = 0.013). Reference videos reduced the variability of clinical assessments and may be useful tools to improve the precision and consistency in the clinical assessment of motor functions in MS.

1 INTRODUCTION

In multiple sclerosis (MS) clinical assessment scales—mainly the Expanded Disability Status Scale (EDSS)—are used to quantify impairment and disability. The EDSS is known for a low inter- and intrarater reliability and suffers from a high variability, especially at lower EDSS scores.(1) Motor dysfunction and particularly ataxia is one of the predominant clinical manifestations in patients with MS and a major contributor to disability progression.(2) Thus, reliable and consistent rating of ataxia is crucial for the follow up of patients with MS.

2 OBJECTIVE

The objective of this report is to investigate whether reference videos (RVs) exemplifying degrees of ataxia severity can reduce the variability of motor dysfunction assessment in MS.

3 METHODS

Study design and participants

This study was a subproject of "Assess MS," a study approved by the local ethics committees.(3) All patients gave their written informed consent to the video recordings. Twenty-five raters (neurologists) from the university hospitals in Bern and Basel rated 60 videos based on 43 MS patients performing the finger-to-nose test (FNT). The videos were recorded with a Microsoft Kinect® 1 camera and chosen out of >2000 videos recorded for the Assess MS study, with the constraint to have coverage for all limb ataxia grades of the Neurostatus- EDSS definitions.(4,5) According to these definitions there are five grades of limb ataxia: 0 = no ataxia, 1 = signs only, 2 = tremor or clumsy movements easily seen, minor interference with function, 3 = tremor or clumsy movements interfere with function in all spheres and 4 = most functions are very difficult. The ratings were performed at baseline and six weeks later ("retest"), to assess the long-term intrarater agreement. In both rating sessions 10% of the videos were presented twice for short-term intrarater agreement.

Forty-one RVs, different from the videos used for rating, were chosen by experienced neurologists of the Assess MS study. They also showed MS patients performing the FNT, with different degrees of limb- ataxia severity, based on the Neurostatus-EDSS definitions.(4,5) The raters were randomized into two groups: one group assessing videos based only on the written Neurostatus-EDSS definitions,(5) without simultaneous access to the RVs (Setting 1), and the other, with simultaneous access to the RVs (Setting 2).

The characteristics of the raters are summarized in Table 1. There was no difference in experience with MS patients between the groups (Setting 1 vs Setting 2).

Table 1, characteristics of patients and neurologists participating in this study.

Patients		
Mean Age, y [range]	42.79 ± 12.09 [23-77]	
Gender (female), f/m	29/14	
Disease duration (y), mean ± SD [range]	13.25 ± 8.38 [0.5-40]	
Median EDSS [range]	3.5 [1-6.5]	
MS type, n (%)	N (43)	
RRMS	39	
SPMS	3	
PPMS	1	
Neurologists	**Group Setting 1**	**Group Setting 2**
Gender (female), f/m	7/6	5/7
Years of experience with MS, mean [range]	5.5 [0.5-12]	5.8 [0.2-12]

EDSS: Expanded Disability Status Scale; MS: multiple sclerosis; PPMS: primary progressive multiple sclerosis; RRMS: relapsing–remitting multiple sclerosis; SPMS: secondary progressive multiple sclerosis.

Patient-performance

For FNT videos, MS patients were instructed by the recording neurologists of the Assess MS study to close their eyes and abduct their arms to 90 degrees at the shoulder in full extension, before touching the nose with the tip of their index finger, first with the dominant, then with the nondominant side (Figure 1).

Video rating

Videos were presented for rating on a touchscreen. Setting 2 allowed for simultaneous presentation of RVs on the right part of the screen (Figure 1). Horizontal swipe allowed for viewing RVs of different limb-ataxia severity degrees; vertical swipe for viewing alternative RVs of the same severity degree. In Setting 1 this part of the screen remained black. Raters were allowed to view each video as often as required for scoring.

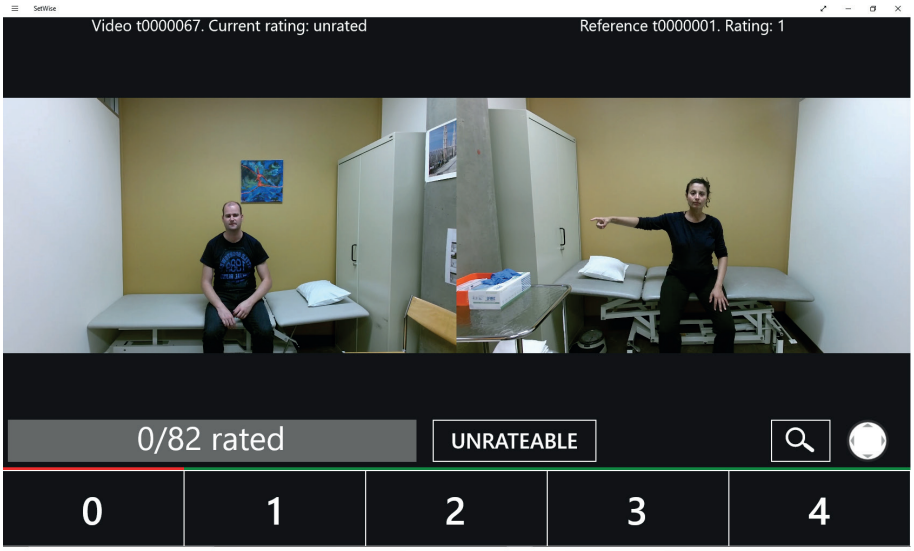


Figure 1, reference videos on the right, videos to be rated on the left, below fields for scoring the appropriate severity of the performance using ataxia grades of the Neurostatus-Expanded Disability Status Scale definitions. According to these definitions there are five grades of limb ataxia: 0 = no ataxia, 1 = signs only, 2 = tremor or clumsy movements easily seen, minor interference with function, 3 = tremor or clumsy movements interfere with function in all spheres and 4 = most functions are very difficult. People shown are not patients and gave written consent to be shown.

Statistics

The analysis was conducted using Matlab R2014b (MathWorks, Natick, MA, USA). *F* test was used to compare the variability of the ratings between the two rater groups (Setting 1 vs Setting 2). Interrater agreement was calculated as intraclass correlation coefficient (ICC) for single measurements and absolute agreement.(6) Intrarater agreement was calculated as the percentage of identical ratings.

4 RESULTS

The variability of ratings was significantly lower in Setting 2 (standard deviation (SD) = 0.12; range = 0.40) than in Setting 1 (SD = 0.26; range = 0.88, F test; p = 0.013), as illustrated in Figure 2. The ICC for interrater agreement was numerically slightly higher in Setting 2 (0.816 (95% confidence interval (CI): 0.756–0.871) vs 0.756 (95% CI: 0.674–0.829) in Setting 1) but this difference was not significant. Short-term and long-term intrarater agreement were similar across settings (Setting 1: 79 ± 18% and 69 ± 11%; Setting 2: 75 ± 22% and 68 ± 9%, not significant).

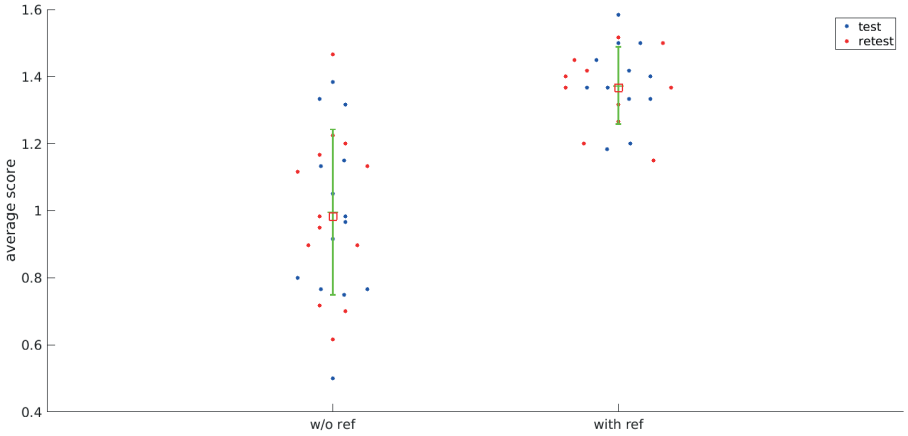


Figure 2, on the left the ratings of Setting 1, i.e. the group without reference videos are shown without ((w/o) ref) and on the right, those from Setting 2, i.e. the group with reference videos ("with ref"). Mean and standard deviation (SD) are shown in green, median in magenta. The variability of ratings was significantly lower in Setting 2 (SD = 0.12; range = 0.40) than in Setting 1, w/o) reference videos (SD = 0.26; range = 0.88, F test; p = 0.013). Each dot represents the average of all ratings of one neurologist (blue at baseline and red six weeks later).

The average score of limb ataxia (according to the Neurostatus-EDSS definitions) was slightly higher in Setting 2, with RVs (mean score (test and retest after six weeks): 1.4 ± 0.1 in Setting 2, vs 1 ± 0.3 in Setting 1, $p < 0.0001$), as illustrated in Figure 2. No significant interaction was found between intrarater agreement, raters' experience with MS or EDSS assessments, or the centers (data not shown).

5 DISCUSSION

As pars pro toto," the results of this study show that using preselected RVs can reduce the rating variability in the assessment of limb ataxia of MS patients. The use of such videos can be easily implemented and does not require an additional/new scale, since we used the already well-established Neurostatus- EDSS definitions.(4) Whether this approach can also be used for assessments other than limb ataxia remains to be shown.

We found a small but statistically significant difference of the average severity level obtained in the two settings with higher ratings in the setting with RVs. As the ataxia degrees were assigned to the RVs by neurologists with special expertise in clinical ratings, this may have contributed to stricter interpretation of the grade definitions. A further limitation in our study was the low number of severely affected patients (ataxia grades 3 and 4). In daily routine, however, rating of lower-severity grades is more challenging than

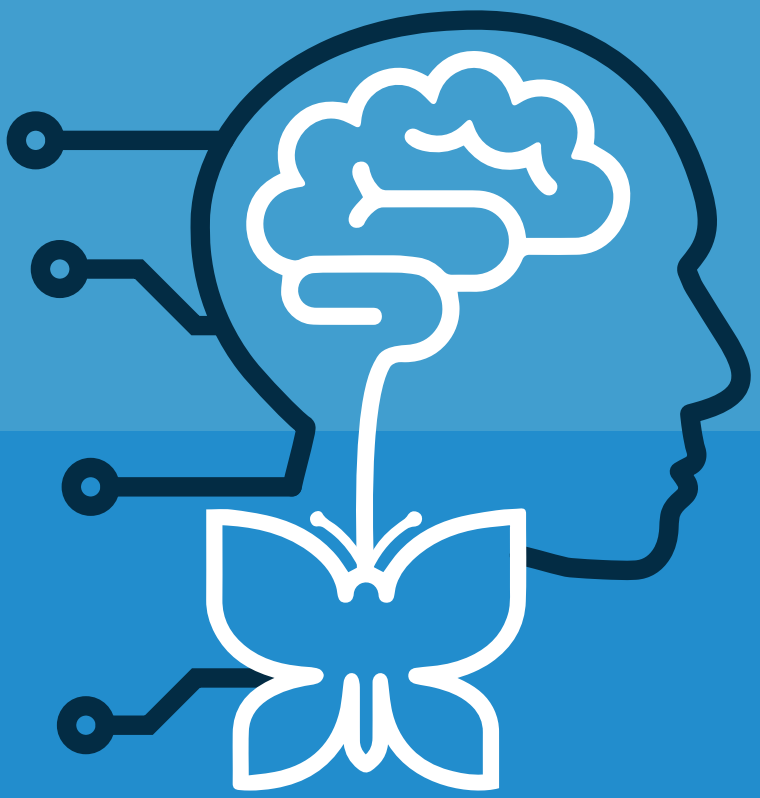
higher grades. Our RV approach may also have a role in training machine-learning algorithms (MLAs). Such an example is the Assess MS system, a potentially finer-grained tool to measure motor dysfunction in MS.(3) This system uses advanced MLAs to analyze three-dimensional-depth-sensor recordings of MS patients performing standard tests of motor function, like the FNT. Reducing the variability of clinical assessments that are used to train MLAs should also contribute to improved algorithms that are derived from machine learning.

6 CONCLUSIONS

The use of RVs may represent a simple method to reduce variability in the assessment of motor dysfunction in MS. This method could be particularly useful in the context of clinical research, and to train MLAs.

REFERENCES

1. Meyer-Moock S, Feng YS, Maeurer M, et al. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC Neurol. 2014; 14(58): 1-10.
2. Weinshenker BG, Issa M, Baskerville J. Long-term and short-term outcome of multiple sclerosis: a 3-year follow-up study. Arch Neurol 1996; 53: 353-358.
3. Morrison C, D'Souza M, Huckvale K, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Human Factors 2015; 2; iss. 1; e11.
4. D'Souza M, Yaldizli O, John R, et al. Neurostatus e-Scoring improves consistency of E xpanded Disability Status Scaleassessments: a proof of concept study. Mult Scler 2017; 23: 597–603.
5. Kappos L. Neurostatus scoring definitions (Version 04/10.2), 2011. http://www.neurostatus.net
6. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psych. Meth. 1996; 1:30–46.



CHAPTER 8

Video-assisted rating improves the detection of change in motor function in multiple sclerosis

Caspar E.P. van Munster, Ka-Hoo Lam, Marcus D'Souza, Saskia Steinheimer, Christian Kamm, Jessica Burggraaff, Matthew Johnson, Yordan Zaykov, Jonas Dorn, Frank Dahlke, Ludwig Kappos, Joep Killestein, Bernard M.J. Uitdehaag.

Submitted for publication

ABSTRACT

Background
Assessing motor function is important to monitor the disease course of multiple sclerosis (MS). Video-capturing and -rating of the performance of classical neurological tests and tasks of activities of daily living (ADL) may improve the detection of change in motor function. We investigated the additional value of video-assisted composite measures for the detection of change in mobility and upper extremity function (UEF).

Methods
Data were collected from 43 MS patients who started fampridine treatment and performed video-recorded composites of tests prior to and after at least two weeks of treatment. Patients were classified as improved and not improved on mobility (MOB-COM) and UEF (UEF-COM) composites based on neurologists' video-ratings of the tests. The proportional agreement between the composites and the Timed 25-Foot Walk test (T25FW) and Nine-Hole Peg Test (NHPT) was determined and compared to clinically relevant improvement with Venn diagrams. Clinically relevant improvement was determined using patient-reported ratings of change in mobility and UEF.

Results
Agreement between MOB-COM and T25FW was 79.5% and 82.1% for UEF-COM and NHPT. 26 patients had perceived clinical improvement on mobility, in which MOB-COM identified two (7.7%) patients with improvement that was not detected with T25FW. In 13 patients with perceived UEF improvement, three (23.1%) patients improved on UEF-COM without improving on NHPT.

Conclusion
The video-assisted composites of ADL tasks and classical neurological tests detected motor improvement in patients with perceived improvement who were missed with conventional measures only. This may improve the detection of treatment effects in clinical practice and trials in MS.

1 INTRODUCTION

Clinical assessment of patients with multiple sclerosis (MS) is notoriously difficult largely due to the multifaceted aspect of disability caused by the disease.(1) With increasing numbers of effective disease modifying therapies and expanding treatment windows, selection of outcome measures that adequately detect clinical change is increasingly important.(2) This was demonstrated by two large MS trials in which no effect on disease progression was found based on the primary endpoint that was heavily weighted on ambulation, while positive effects were found on upper extremity function (UEF).(3, 4) The accurate assessment of disability is also relevant in the evaluation of symptomatic treatment options such as fampridine.(5) Fampridine improves walking speed and UEF in a subset of MS patients, albeit patient-perceived improvement does not fully correspond with improvement measured with clinical instruments.(6-8)

The Expanded Disability Status Scale (EDSS) is the most widely used outcome measure in MS trials, despite having a high inter- and intra-rater variability and a disproportional impact of ambulation on the total score.(9, 10) To improve the assessment of commonly affected functional domains in MS, the Multiple Sclerosis Functional Composite (MSFC) was introduced to assess ambulation (timed 25-foot walk test, T25FW), UEF (nine-hole peg test, NHPT), and cognition.(11, 12) Despite generally good psychometric properties, the individual components have several shortcomings.(13, 14) For instance, the T25FW may not be sensitive enough to detect abnormalities in patients with mild ambulatory impairment.(15) Furthermore, the T25FW and NHPT measure only a certain aspect of ambulation and UEF: respectively walking speed and manual dexterity, which do not fully capture the broader aspects of functioning. Because of these shortcomings possible clinically relevant treatment effects might be missed.

By recording movements of patients on video, multiple standardised functional tests can be assessed at different time points and still be rated simultaneously. This video-assisted approach allows the assessment of tests of the neurological examination with less intra-rater variability as there is no recall bias.(16, 17) Moreover, standardised tasks of activities of daily living (ADL) can be added to the assessment, which is valuable in measuring the patients' functioning.(18) All in all, such video-assisted composite measure taking into account multiple aspects of motor function, may achieve a more complete assessment of (change in) motor performance. In this study we investigated the additional value of a multidimensional video-assisted composite measure compared to conventional measures and aim to enhance the detection of change in mobility and UEF in patients with MS.

2 PATIENTS AND METHODS

Data used in this study were part of a multicentre project to develop the Assess MS system, performed in four large European MS centres located in Amsterdam, Basel, Bern and Lucerne.(19, 20) The aim was to develop a consistent and fine-grained system that automatically quantifies motor function in MS patients. Standardised movements of MS patients were recorded using a 3D depth sensing and colour camera (i.e. the Microsoft Kinect®) with the aim of training machine learning algorithms to automatically quantify motor function. Written informed consent was obtained from all subjects prior to study participation and the study was approved by the respective ethics committee.

Subject recruitment

Patients were recruited at Amsterdam UMC, location VU Medical Center between August 2017 and April 2018. Patients initiating treatment with fampridine were eligible for study participation, as fampridine can have positive treatment effects on ambulatory function and UEF.(6-8) Inclusion criteria were, conforming to the official treatment label of fampridine, diagnosis of MS according to the revised 2010 McDonald's criteria,(21) age above 18 years, and ambulatory impairment defined as an estimated EDSS score between 4.0 and 7.0 without additional diseases that contributed to disability. Contraindications as stated in the fampridine product label were handled as exclusion criteria.(22) Patients were assessed at baseline prior to fampridine treatment and at follow-up after at least two weeks of fampridine treatment. Patient and disease characteristics were collected during the baseline visit: age, sex, EDSS,(9) disease type, and disease duration.

Conventional measures

As conventional measure for mobility the T25FW was used, and for UEF the NHPT. The T25FW was performed twice on both visits, as implemented in the MSFC.(11) For each visit two trials were averaged for each patient. A decrease of at least 20% of baseline walking time was used to indicate significant improvement in mobility function.(14, 23, 24) The NHPT was performed twice for both hands on both visits, as implemented in the MSFC.(11) The trials from the dominant and non-dominant hand were averaged into one value. Similarly to the T25FW, a decrease in completion time of at least 20% between visits was used as cut-off for significant improvement of UEF.(23)

Video-assisted composite measures

In addition to the conventional measures, patients performed a set of standardised movements consisting of ADL tasks and tests from the

neurological examination, see figure 1A. Three classical neurological tests and four ADL tasks assessed mobility and comprised the mobility composite (MOB-COM). UEF was assessed with three classical neurological tests and two ADL tasks and were combined into the UEF composite (UEF-COM). The movements comprising MOB-COM and UEF-COM were recorded using the Microsoft Kinect® camera, as shown in figure 1B. Each colour video was rated by two neurologists, who were blinded for the visit type (baseline or follow up) and the patients' performance on the conventional measures. The classical neurological tests were quantified according to predetermined rating scales based on the Neurostatus-EDSS functional system scoring definitions.(25) For the ADL tasks, a five-point Likert scale was created ranging from 0 ('normal') to 4 ('unable to perform due to disability'). Video rating was performed by using reference videos.(17) Using an algorithm that takes into account individual rater bias, the videos were then assigned a consensus score.(26) This consensus score was subsequently used in the statistical analyses. Videos of insufficient quality or inaccurately performed movements were excluded from the analysis. A more comprehensive description of the standardised movements, video rating, and score calculation can be found elsewhere.(17, 18) Significant improvement on MOB-COM and UEF-COM was defined as two or more tests improving (i.e. ≥1 point decrease) and no more than one test worsening (i.e. ≥1 point increase) during follow-up compared to baseline.

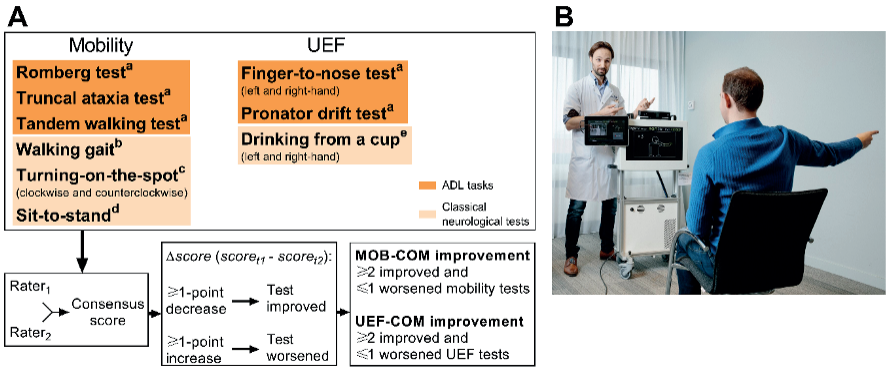


Figure 1, schematic overview of the video-assisted composite measures (A). Setup of the video-assisted composite measure (B). Consent was given by the people in the figure. [a]From the classical neurological examination, scored according to the Neurostatus EDSS. [b]Normal walking gait over 25 foot, scoring: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), 4 (not possible). [c]Turning 360 degrees while standing, scoring: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), 4 (not possible). [d]Getting up from a chair without using the arms, scoring: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), 4 (not possible). [e]Taking a sip from a cup, scoring: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), 4 (not possible)Abbreviations: UEF = Upper Extremity Function; ADL = Activities of Daily Living; t_1 = Baseline visit; t_2 = Follow-up visit; MOB-COM = Video-assisted Mobility Composite; UEF-COM = Video-assisted Upper Extremity Function Composite.

Clinically relevant improvement

Clinically relevant improvement was externally anchored by using Global rating of change (GRC) scales.(27) At the follow-up visit, patients rated their perceived change on mobility and UEF at compared to baseline. GRC mobility was rated on a 7-point Likert scale ranging from -3 (i.e. very much deteriorated) to +3 (i.e. very much improved) and GRC UEF was rated on a 5-point scale ranging from -2 (i.e. much deteriorated) to +2 (i.e. much improved). Patients with positive GRC scores were classified as having clinically relevant improvement, whereas GRC of 0 or negative scores were classified as no clinically relevant improvement.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics version 24. Categorical variables were summarised as frequencies with percentages. The mean and standard deviation were used to summarise continuous variables that were normally distributed, otherwise the median and interquartile range were used. The proportional agreement between the video-assisted composite and conventional measures was determined using 2 by 2 contingency tables for ambulation (T25FW and MOB-COM) and UEF (NHPT and UEF-COM). To determine and visualise the added value of the composite measures, Venn diagrams were drafted showing the overlap in patients who improved on the various outcomes for mobility and UEF.

3 RESULTS

A total of 43 MS patients initiating fampridine treatment were included and completed baseline measurements. After baseline measurements, 2 patients discontinued treatment prematurely due to side-effects, and 2 patients refused follow-up assessment due to time constraints. The baseline characteristics of the 39 patients that completed the study are summarised in table 1.

Table 1, baseline patient characteristics (n = 39).

Age, y, mean (SD)	54.0 (10.4)
Sex, n (%)	
Male	21 (53.8)
Female	18 (46.2)
Disease type, n (%)	
RRMS	17 (43.6)
SPMS	10 (25.6)
PPMS	12 (30.8)

Age, y, mean (SD)	54.0 (10.4)
Disease duration, y, mean (SD)	13.8 (8.1)
EDSS, median (IQR)	5.5 (4.0 – 6.0)
Fampridine treatment duration, d, mean (SD)	24.5 (8.5)
T25FW, s, median (IQR)	6.9 (5.8 – 9.3)
MSWS, mean (SD)	73.2 (17.5)
NHPT, s, median (IQR)	27.3 (23.7 – 42.2)
AMSQ, median (IQR)	63.0 (45.0 – 94.0)

Abbreviations: RRMS = Relapsing Remitting Multiple Sclerosis; SPMS = Secondary Progressive Multiple Sclerosis; PPMS = Primary Progressive Multiple Sclerosis; EDSS = Expanded Disability Status Scale; T25FW = Timed-25 Foot Walk test; MSWS = Multiple Sclerosis Walking Scale; NHPT = Nine-Hole Peg Test; AMSQ = Arm function in Multiple Sclerosis Questionnaire.

Based on the conventional measures and the video-assisted composite measures, the number and percentage of patients with significant improvement of mobility and UEF are shown in figure 2.

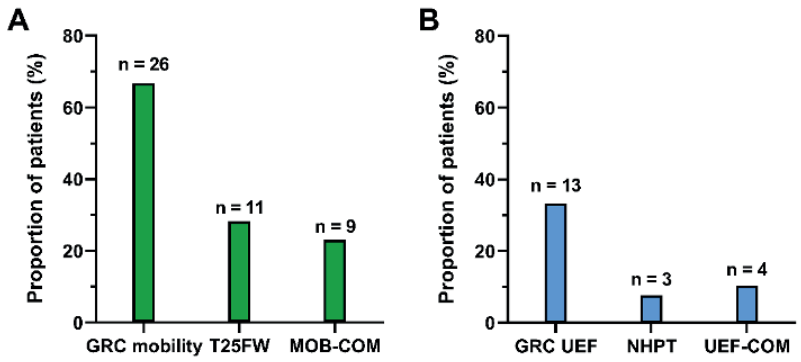


Figure 2, bar graphs depicting the counts and percentages of patients with significant improvement at follow-up for mobility (A) and upper extremity (B) function. Abbreviations: GRC = Global Rating of Change; T25FW = Timed 25-Foot Walk test; MOB-COM = Video-assisted Mobility Composite; NHPT = Nine-Hole Peg Test; UEF-COM = Video-assisted Upper Extremity Function Composite.

Of the patients improving on MOB-COM, the proportion of improvement on ADL tasks and classical neurological tests was equal. This was similar for UEF, only slightly favouring neurological tests (54.5%) compared to ADL tasks (45.5%). See figure 3.

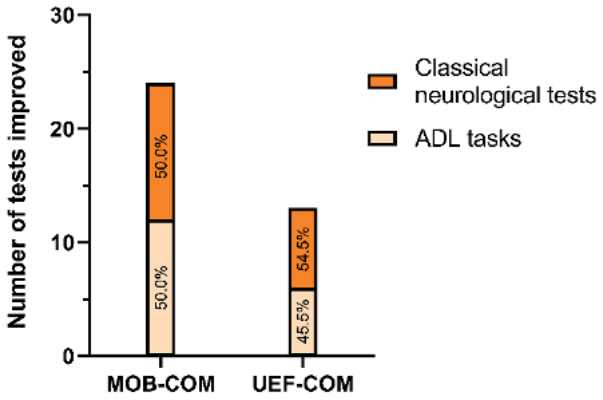


Figure 3, stacked bar graphs depicting the number of tests contributing to MOB-COM and UEF-COM improvement split between ADL tasks and Classical neurological tests. Abbreviations: MOB-COM = Video-assisted Mobility Composite; UEF-COM = Video-assisted Upper Extremity Function Composite; ADL = Activities of Daily Living.

Comparison of conventional and video-assisted composite measures

Table 2 shows the proportion of patients who improved as stratified by the conventional measures and video-assisted composite measures for ambulatory function. Three (7.7%) patients improved on the MOB-COM without improvement on T25FW. Conversely, five (12.8%) patients improved on T25FW without improvement on MOB-COM. The remaining patients (79.5%) were stratified similarly by both measures.

Table 2, cross tabs between T25FW and MOB-COM.

	MOB-COM Improved[a]	Not improved
T25FW		
Improved[b]	6 (15.4)	5 (12.8)
Not improved	3 (7.7)	25 (64.1)

Data are expressed as n (%). [a]≥ 2 improvement and ≤ 1 worsening of movements on the mobility composite, [b]≥ 20% increase of walking speed at follow-up compared to baseline. Abbreviations: MOB-COM = Video-assisted Mobility Composite; T25FW = Timed 25-Foot Walk test

For UEF, 32 (82.1%) patients were stratified similarly by NHPT and UEF-COM. Four (10.3%) patients improved on the UEF-COM with no improvement on the NHPT. Conversely, three (7.7%) patients did not improve on UEF-COM, and improved on NHPT. See table 3.

Table 3, cross tabs between NHPT and UEF-COM.

	UEF-COM Improved[a]	Not improved
NHPT		
Improved[b]	0 (0.0)	3 (7.7)
Not improved	4 (10.3)	32 (82.1)

Data are expressed as n (%). [a]≥ 2 improvement and ≤ 1 worsening of movements of the upper extremity function composite, [b]≥ 20% decrease of time to complete the Nine-Hole Peg Test at follow-up compared to baseline. Abbreviations: UEF-COM = Video-assisted Upper Extremity Function Composite; NHPT = Nine-Hole Peg Test.

Detection of clinically relevant improvement

Of the 26 patients who reported improvement on ambulatory function on the GRC, 10 also showed significant improvement on T25FW compared to 8 on MOB-COM. One patient significantly improved on T25FW without reporting improvement on GRC. Two of three patients that improved on MOB-COM and not on the T25FW, also reported improvement of GRC. The overlap between improvement on GRC, T25FW and MOB-COM is depicted in figure 4. There was no overlap of patients reporting improvement on GRC (n = 13) and showing improvement on 9HPT (n = 3). Three of four patients that improved on UEF-COM and not on the 9HPT, also reported improvement of GRC. The overlap between improvement on GRC, 9HPT and UEF-COM is depicted in figure 5.

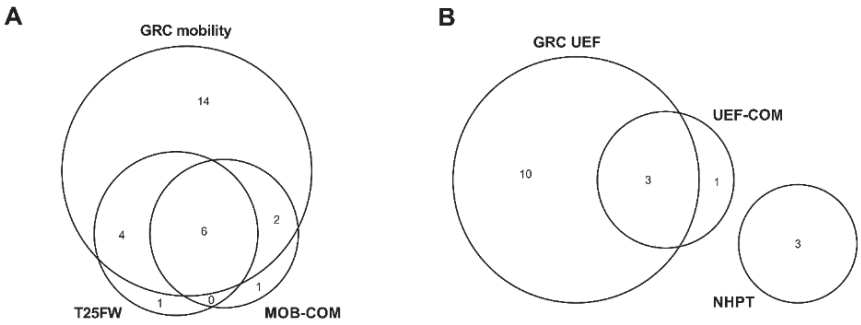


Figure 4, Venn diagrams depicting the overlap between patient-perceived improvement (GRC) and improvement on the conventional (T25FW and NHPT) and video-assisted composite (MOB- and UEF-COM) measures for mobility (A) and upper extremity function (B). Abbreviations: GRC = Global Rating of Change; T25FW = Timed 25-Foot Walk test; MOB-COM = Video-assisted Mobility Composite; UEF-COM = Video-assisted Upper Extremity Function Composite; NHPT = Nine-Hole Peg Test.

4 DISCUSSION

In this study we investigated the additional value of two multidimensional video-assisted composites in relation to conventional measures to detect change in motor function. The composites used here cover various aspects of motor function: ADL tasks and classical neurological tests. In patients who reported perceived improvement, the composites for mobility function detected two (7.7%) and UEF three (23.1%) patients with improvement which were not found using the conventional measures.

With single modality measures alone, improvement of motor function may be missed. For instance, a patient might walk more coordinated despite needing the same amount of time to walk 25 feet compared to a patient who is able to walk faster, but walks more impaired. In addition, for the T25FW both ceiling and floor effects have been pointed out in patients with EDSS >6.5 and patients in the lower EDSS ranges, respectively.(14, 15). With the highest EDSS of 6.5 in our cohort, only floor effects of the T25FW are likely to have incurred in the patients who had close to 'normal' short distance walking speed. These patients are unlikely to achieve an even faster walking speed to satisfy the clinically meaningful improvement criterion of 20%, but may improve in other domains than walking speed. This is supported by a larger cohort of patients treated with fampridine where lower baseline walking speed was matched with lower rates of 20% improvement on T25FW.(28) Similarly, floor effects have also been reported for the NHPT together with the suggestion of the use of multiple clinical tests to overcome the deficiency.(29)

The composite measures also incorporated tasks of ADL, which we found to be equally important, proportionally wise, as neurological examination tests when assessing improvement of ambulation and UEF. A previous study with similar methodology in a cross-sectional setting suggests ADL tasks to be important, perhaps even more valuable than neurological tests, for the assessment of motor function in MS.(18) This study confirms the role of ADL functioning in assessment of change in a longitudinal setting. Strikingly, there was no overlap between NHPT improvement and patient-perceived UEF improvement. Whereas 75% of the patients with improvement on the UEF composite also had perceived UEF improvement, further supporting the role of ADL functioning.

Several advantages may be pointed out regarding the use of video-assistance in the assessment of motor function. This includes the ability to directly compare motor performance across patients and within patients in a longitudinal setting, which may reduce intra- and interobserver variability. Which can be potentially further reduced when aided by the use of reference videos or machine-learning algorithms in the rating of movements. In addition,

video-assisted assessment of motor function allows disease monitoring from physical distance. The importance of assessment from physical distance is emphasized in the current COVID-19 pandemic, especially in patients with MS who may be at higher risk due to immunomodulatory therapies. On the other hand, limitations of this study include the small sample size. The studied composite measures were based on consensus scores derived from ratings of only two neurologists. For the assessment of clinically relevant change potential placebo effects are not accounted for. Our findings also provide limited insight in patients in the lower and upper ranges of EDSS scores as disability in our cohort of patients ranged from EDSS of 3.5 to 6.5. Lastly, potential learning effects in performing the video-assisted composite at follow-up were not studied, although we assume this is not a major bias since the composites were assessed once each visit with an interval of at least 2 weeks.

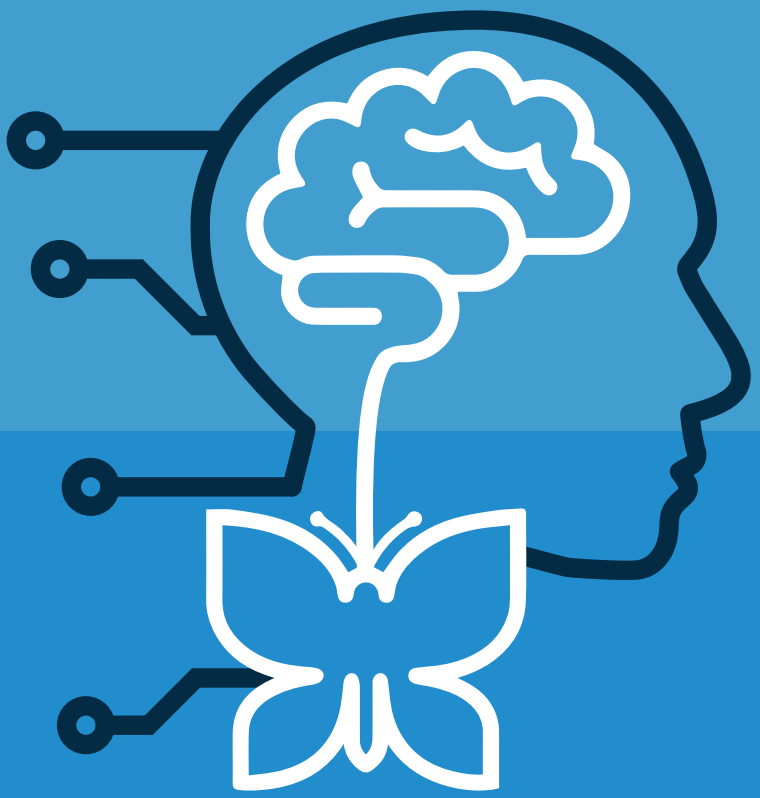
CONCLUSIONS

The composite measures allow assessment of multiple domains of physical functioning, including ADL which is not conventionally measured. The video-based rating method also enables direct comparison of motor function across patients and within patients in a longitudinal setting. Through the video-assisted composites, previously undetected objective improvement can be found in patients with self-perceived improvement. This video-assisted method, enabling assessment from physical distance, enhances the detection of improvement of mobility and upper extremity function in MS.

REFERENCES

1. Scalfari A, Neuhaus A, Daumer M, Deluca GC, Muraro PA, Ebers GC. Early relapses, onset of progression, and late outcome in multiple sclerosis. JAMA neurology. 2013;70(2):214-22.
2. Faissner S, Plemel JR, Gold R, Yong VW. Progressive multiple sclerosis: from pathophysiology to therapeutic strategies. Nature reviews Drug discovery. 2019;18(12):905-22.
3. Kapoor R, Ho PR, Campbell N, Chang I, Deykin A, Forrestal F, 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. The Lancet Neurology. 2018;17(5):405-15.
4. Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. The New England journal of medicine. 2017;376(3):209-20.
5. Hayes KC. Impact of extended-release dalfampridine on walking ability in patients with multiple sclerosis. Neuropsychiatric disease and treatment. 2011;7:229-39.
6. Goodman AD, Brown TR, Schapiro RT, Klingler M, Cohen R, Blight AR. A pooled analysis of two phase 3 clinical trials of dalfampridine in patients with multiple sclerosis. International journal of MS care. 2014;16(3):153-60.
7. Savin Z, Lejbkowicz I, Glass-Marmor L, Lavi I, Rosenblum S, Miller A. Effect of Fampridine-PR (prolonged released 4-aminopyridine) on the manual functions of patients with Multiple Sclerosis. Journal of the neurological sciences. 2016;360:102-9.
8. Klineova S, Farber R, Friedman J, Farrell C, Lublin FD, Krieger S. Objective and subjective measures of dalfampridine efficacy in clinical practice. Multiple sclerosis journal - experimental, translational and clinical. 2018;4(3):2055217318786742-.
9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33(11):1444-52.
10. van Munster CE, Uitdehaag BM. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS drugs. 2017;31(3):217-36.
11. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain : a journal of neurology. 1999;122 (Pt 5):871-82.
12. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC neurology. 2014;14:58.
13. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2017;23(5):711-20.
14. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2017;23(5):704-10.
15. Goldman MD, Motl RW, Scagnelli J, Pula JH, Sosnoff JJ, Cadavid D. Clinically meaningful performance benchmarks in MS: timed 25-foot walk and the real world. Neurology. 2013;81(21):1856-63.

16. Steinheimer S, Dorn JF, Morrison C, Sarkar A, D'Souza M, Boisvert J, et al. Setwise comparison: efficient fine-grained rating of movement videos using algorithmic support - a proof of concept study. Disability and rehabilitation. 2019:1-7.
17. D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. Multiple sclerosis journal - experimental, translational and clinical. 2018;4(3):2055217318792399.
18. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2018:1352458518796690.
19. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR human factors. 2015;2(1):e11.
20. Kontschieder P, Dorn JF, Morrison C, Corish R, Zikic D, Sellen A, et al. Quantifying progression of multiple sclerosis via classification of depth videos. Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention. 2014;17(Pt 2):429-37.
21. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302.
22. European Medicines Agency. Fampyra: EPAR - Summary for the public [Internet]. 2011 [updated 2017 August 4; cited 2019 July 24]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/fampyra.
23. van Winsen LM, Kragt JJ, Hoogervorst EL, Polman CH, Uitdehaag BM. Outcome measurement in multiple sclerosis: detection of clinically relevant improvement. Multiple Sclerosis Journal. 2010;16(5):604-10.
24. Hobart J, Blight AR, Goodman A, Lynn F, Putzki N. Timed 25-foot walk: direct evidence that improving 20% or greater is clinically meaningful in MS. Neurology. 2013;80(16):1509-17.
25. Kappos L. Neurostatus scoring definitions (Version 04/10.2) 2011 [Available from: https://www.neurostatus.net.
26. Sarkar A, Morrison C, Dorn JF, Bedi R, Steinheimer S, Boisvert J, et al. Setwise Comparison: Consistent, Scalable, Continuum Labels for Computer Vision. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; San Jose, California, USA: Association for Computing Machinery; 2016. p. 261–71.
27. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-15.
28. van Munster CEP, Kaya L, Lam KH, Kalkers NF, Killestein J, Uitdehaag BMJ. Responder rates to fampridine differ between clinical subgroups of MS patients and patient reported outcome influences treatment decision making. Multiple sclerosis and related disorders. 2020;38:101489.
29. Solaro C, Cattaneo D, Brichetto G, Castelli L, Tacchino A, Gervasoni E, et al. Clinical correlates of 9-hole peg test in a large population of people with multiple sclerosis. Multiple sclerosis and related disorders. 2019;30:1-8.



CHAPTER 9

Autoencoder: a new method for keeping data privacy when analysing videos of patients with motor dysfunction - a proof of concept study

Caspar E.P. van Munster, Marcus d'Souza, J. Dorn, A. Dorier, C.P. Kamm, S. Steinheimer, F. Dahlke, B.M.J. Uitdehaag, L. Kappos, M. Johnson.

Journal of Medical Internet Research. 2020 May 8; 22(5)

ABSTRACT

Background
In chronic neurological diseases, especially in multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients. Traditional scales are not sensitive enough to detect slight changes. Video recordings of patient performance are more accurate and increase the reliability of severity ratings. When these recordings are automated, quantitative disability assessments by machine learning algorithms can be created. Creation of these algorithms involves non–health care professionals, which is a challenge for maintaining data privacy. However, autoencoders can address this issue.

Objective
The aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

Methods
In this study, 20 pre-rated videos of patients performing the finger-to-nose test were recorded. An autoencoder created encoded frame vectors from the original videos and decoded the videos again. The original and decoded videos were shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether the 200 videos were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-Expanded Disability Status Scale definitions of limb ataxia. Furthermore, the neurologists tested whether ratings were equivalent between the original and decoded videos.

Results
In total, 172 of 200 (86.0%) videos were of sufficient quality to be ratable. The intrarater agreement between the original and decoded videos was 0.317 (Cohen weighted kappa). The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The interrater agreement between the original videos was 0.459 and that between the decoded videos was 0.302. The agreement was higher when no deficits or very severe deficits were present.

Conclusions
The vast majority of videos (172/200, 86.0%) decoded by the autoencoder contained clinically relevant information and had fair intrarater agreement with the original video. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non–health-care professionals are involved.

1 INTRODUCTION

In chronic neurological diseases, especially multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients.(1) Traditional scales used to assess MS, such as the Expanded Disability Status Scale (EDSS), are not sensitive enough to detect slight changes in motor performance.(2) Video recordings of patient performance are more accurate and increase the reliability of severity ratings.(3,4) Moreover, when these recordings are automated, quantitative disability assessments by machine learning algorithms (MLA) can be created.(5) Machine learning algorithms are potentially more sensitive in detecting small changes between images; however, they require high-resolution images because of the high dimensionality of the data.(6,7) Creation of these algorithms usually involves non–health care professionals, which is a potential challenge for maintaining data privacy. Autoencoders can address this issue. They embed visual information into a lower-dimensional latent space that preserves information needed for algorithm development but is not visually interpretable by humans.(6) An autoencoder consists of an encoder that creates encoded videos by creating a sequence of coded frame vectors and a paired decoder that transforms the coded frame vectors back into the original video. Videos encoded in this way can be shared with non–health care professionals, while the decoder can be used to verify if the essential information from the video has been captured. However, it is unknown whether the condensed data in the coded frame vectors contain clinically relevant data. Therefore, the aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

2 METHODS

Study design and participants

This study was a subproject of the ASSESS MS study and was approved by the local ethics committees.(6) All participants gave their written informed consent prior to inclusion. In the ASSESS MS study, 9 standardized movements were recorded on video; these movements covered overall motor function, including upper extremity function, truncal stability, and mobility. A detailed description of the movements can be found elsewhere.(8) For this study, we used recordings of the finger-to-nose test. The execution of the finger-to-nose test was standardized using a detailed protocol: Each participant was instructed to close their eyes and abduct their arms to 90° at the shoulder in full extension before touching their nose with the tip of their index finger. Both sides were tested. Original and decoded videos of 20 participants were

shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether these 200 videos in total were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-EDSS definitions of limb ataxia (subscore grade 0=no ataxia; grade 1=signs only; grade 2=tremor or clumsy movements easily seen, minor interference with function; grade 3=tremor or clumsy movements that interfere with function in all spheres; and grade 4=most functions are very difficult).(9) The decoded videos were shown firstly, and after an interval of 2-3 weeks, the original videos were shown in the same order to minimize recall bias. The neurologists tested whether these videos were human-readable after decoding.

Autoencoder

A variational autoencoder was trained on 2230 videos comprising the 9 standardized motor performances included in the ASSESS MS study. The autoencoder was structured so that the frames of each video were encoded into a lower-dimensional space and then decoded into their original form. The key property of interest to us was that when a frame is in its coded form, it is computationally prohibited to decipher it without access to the decoder. (6) An autoencoder as described above reduces the dimensionality of the input data (in our case, videos) by passing the data through an "information bottleneck".(14) The resulting coded, or latent, space sufficiently describes the data in a way that allows an accurate partial reconstruction. The shared latent embedding is optimized to represent the salient information that is similar across frames of multiple videos (in our case,: the movement), whereas dissimilar aspects (eg,. background aspects, details of physical features) are less well conserved. Neural networks are a machine learning approach that is inspired by biological neuronal computation; these networks have demonstrated exceptional performance in complex image-related tasks in recent years.(15-17) Given this success, in this study, we used a neural net approach called a variational autoencoder [18]. A variational autoencoder has at its center a coded vector of vastly reduced dimensionality. This is because the decoder requires millions of floating-point values to be set precisely before the coded vector can be successfully decoded into an image. At the same time, the coded vector contains all the information necessary to reconstruct that frame; interestingly, due to the variational constraints during training, the frame has semantically meaningful cosine distances to other visually similar frames. This property is very useful for machine learning tasks that operate upon these coded vectors because the coded frames can be used in place of the original video frames without the possibility that a human could use it to recognize the depicted participant.

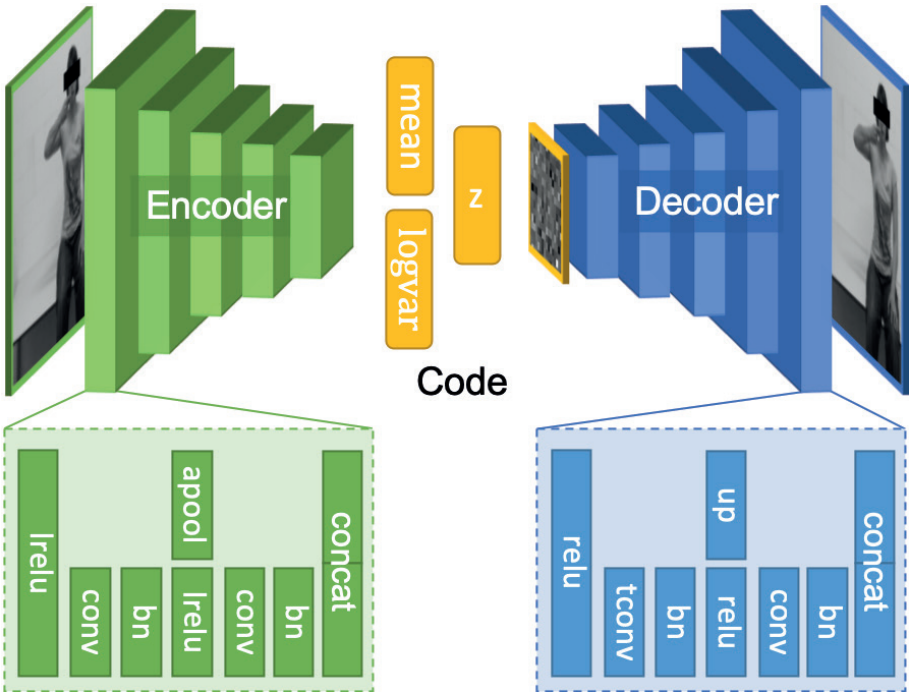


Figure 1, depicts the structure of the autoencoder.(10) An encoder network was presented with a single frame from the video without further context. The frame passed through 5 encoding blocks. In each block, the input was processed in a block inspired by a densely connected convolutional network,(11) wherein a skip connection was provided between the input and output layers in addition to a convolutional layer/batch normalization sequence. Each block halved the resolution of the image and doubled the feature depth. This network predicted the mean and variance of a normal distribution, which was then sampled to produce a code. The code was presented to a second network that consisted of 5 decoding blocks. Each decoding block consisted of a skip connection (which performed a simple upsampling process) and a transposed convolutional block like that used in a deep convolutional generative adversarial network.(12) Each block doubled the resolution and halved the feature depth. The network was trained using a multi-scale structural similarity–based perceptual loss function with Kullback-Leibler regularization as per Kingma and Welling.(10,13) The input images were 256×256 RGB-D images with a code length of 256. The training hyperparameters were as follows: the learning rate was 0.001, the convolutional kernel size was 5, and the number of initial filters was 8. The model was trained for 400 epochs.

Statistics

Intrarater agreement between the ratings of the original and the decoded videos was assessed using the Cohen weighted kappa with linear weights (ie, disagreements of 1, 2, and 3 were weighted by factors of 1, 2, and 3, respectively). A Cohen kappa of 0 corresponds to chance agreement; 0-0.2, to slight agreement; 0.21-0.4, fair agreement; 0.41-0.6, to moderate agreement; 0.61-0.8, to substantial agreement; and 0.81-1, to almost perfect agreement. (19) All analyses were performed in MATLAB (MathWorks, Inc).

3 RESULTS

The characteristics of the study population and the participating neurologists are summarized in Table 1. In total, 172/200 (86.0%) videos were of sufficient quality to be ratable.

Table 1, characteristics of patients and neurologists.

Patients	
Mean Age, years (95% CI)	44.4 [27-74]
Gender female/ male	12/7
Mean disease duration, years (95% CI)	13.2 [1-40]
Median EDSS [range]	3.5 [0-6.5]
MS type, n RRMS SPMS	 19 1
Neurologists	
Gender (female), f/m	5/5
Years of experience in Neurology, mean [range]	8.8 [3->30]

CI = confidence interval; EDSS = Expanded Disability Status Scale; MS = Multiple Sclerosis; RRMS = Relapsing Remitting MS; SPMS = Secondary Progressive MS.

In total, 172/200 (86.0%) videos were of sufficient quality to be ratable. The Cohen weighted kappa indicating intra-rater agreement between the original and decoded videos was 0.317. The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The inter-rater agreements of the original and decoded videos were 0.459 and 0.302, respectively. As depicted in Figure 2, agreement was higher when no deficits (grade 0) or very severe deficits (grade 4) were present. Note that most videos that were not ratable were judged so by neurologists 2 and 5.

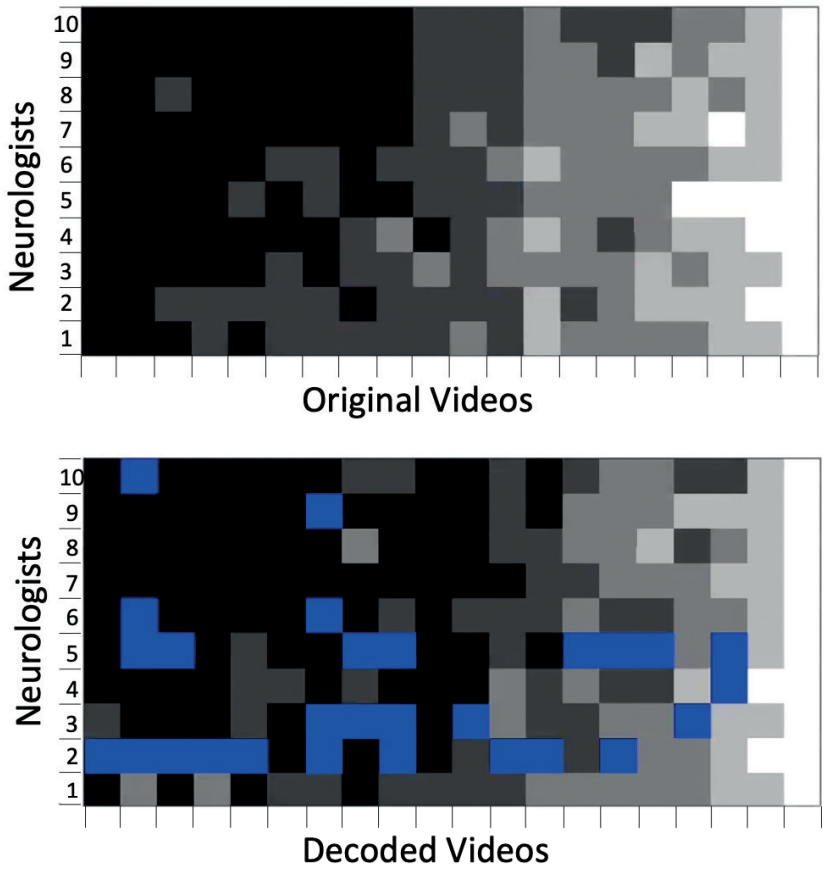


Figure 2, ratings by 10 neurologists of the original and decoded videos. The colored squares represent the different grades for limb ataxia of the finger-to-nose-test according to the Neurostatus-Expanded Disability Status Scale subscores: black=0, dark grey=1, grey=2, bright grey=3, and white=4. The blue squares represent videos that were judged as not ratable by the neurologists.

4 DISCUSSION

Principle Findings

In this proof-of-concept study, 172/200 (86.0%) of the decoded videos were of sufficient quality to be ratable. We found fair intrarater agreement between the original and decoded videos. The agreement was better for minor and severe deficits in motor function.

Data security and privacy are increasingly requested by health care professionals for data capture, analysis, and storage.(20) At the same time, the use of machine learning algorithms and deep neuronal network techniques as subdomains of artificial intelligence is increasingly infiltrating all areas of health care.(21,22) The use of new technologies and electronic tools for capture and automated analysis of clinical data generally requires the involvement of non–health care professionals, which creates challenges regarding data privacy. To our knowledge, this is the first study to use an autoencoder to allow the analysis of patient videos while preserving data privacy.

Patients with MS may present with slight changes in motor performances over their disease course. Clinical assessment of these changes is notoriously difficult. Video analysis of motor performances allows automated analyses and quantification of disability by using machine learning algorithm–based analysis systems such as those used in the ASSESS MS study; however, it requires a huge data set.(5) Since the creation of machine learning algorithms usually involves non-medical collaborators, encoding of these videos is essential. The intra-rater agreement of original and decoded videos in this study was fair. It is unclear whether this is due to accordance of the video quality or the test-retest reliability of the finger-to-nose test. To our knowledge, no data are available regarding this psychometric property of the finger-to-nose test.

Limitations

A limitation of this proof-of-concept study is the class imbalance of the patient videos according to the four grades of limb ataxia for the finger-to-nose test.(9,21) Further iterations of the deep neural network are necessary to increase the intrarater reliability.

Conclusions

In this proof-of-concept study, we have shown that the vast majority (172/200, 86.0%) of videos decoded by an autoencoder contained clinically relevant information regarding upper extremity motor performance represented by the finger-to-nose test and had fair intrarater agreement. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non–health care professionals are involved.

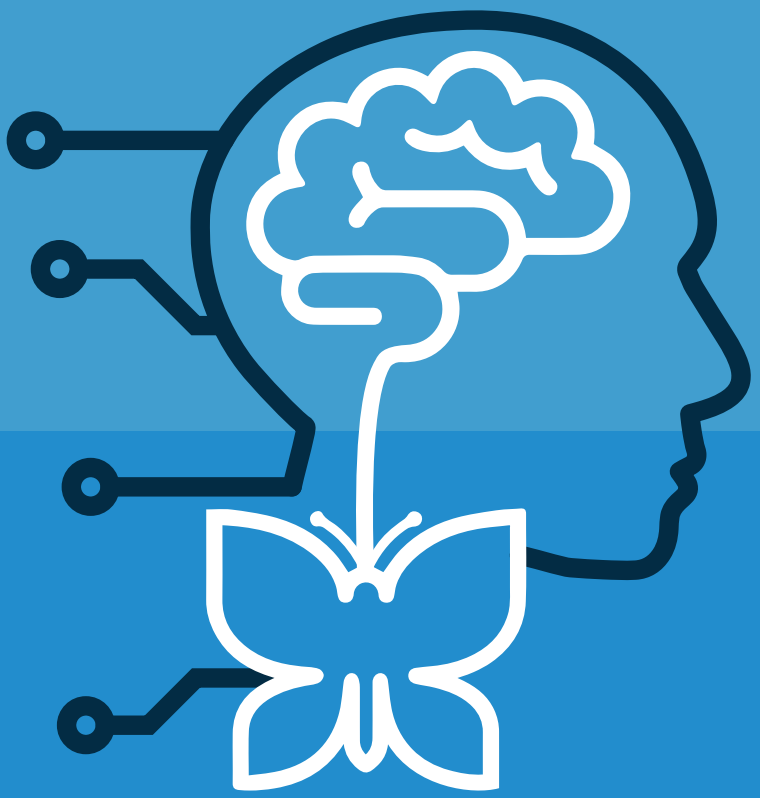
REFERENCES

1. Cohen JA, Reingold SC, Polman CH, Wolinsky JS. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. Lancet Neurol 2012 May;11(5):467-476. [doi: 10.1016/s1474-4422(12)70059-5] [Medline: 22516081]
2. van Munster CEP, Uitdehaag BMJ. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs 2017 Feb 9;31(3):217-236. [doi: 10.1007/s40263-017-0412-5] [Medline: 28185158]
3. Burggraaff J, Dorn J, D'Souza M, Morrison C, Kamm CP, Kontschieder P, et al. Video-Based Pairwise Comparison: Enabling the Development of Automated Rating of Motor Dysfunction in Multiple Sclerosis. Arch Phys Med Rehabil 2020 Feb;101(2):234-241. [doi: 10.1016/j.apmr.2019.07.016] [Medline: 31473205]
4. D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. Mult Scler J Exp Transl Clin 2018 Aug 09;4(3):205521731879239. [doi: 10.1177/2055217318792399] [Medline: 30116550]
5. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors 2015 Jun 24;2(1):e11. [doi: 10.2196/humanfactors.4129] [Medline: 27025782]
6. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci Biobehav Rev 2017 Mar;74:58-75. [doi: 10.1016/j.neubiorev.2017.01.002] [Medline: 28087243]
7. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large□scale multi□sample study. Hum Brain Mapp 2018 Oct 11;40(3):944-954. [doi: 10.1002/hbm.24423] [Medline: 30311316]
8. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Mult Scler 2018 Aug 31;25(12):1673-1681. [doi: 10.1177/1352458518796690] [Medline: 30168739]
9. Kappos, L. https://www.neurostatus.net/. 2011. Neurostatus Scoring Definitions; Version 04/10.2 URL: https://www. neurostatus.net/
10. Kingma DP, Welling M. An Introduction to Variational Autoencoders. Found Trends Mach Learn 2019;12(4):307-392. [doi: 10.1561/2200000056]
11. Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. Convolutional Networks with Dense Connectivity. IEEE Trans Pattern Anal Mach Intell 2019 May 23:1-1. [doi: 10.1109/tpami.2019.2918284] [Medline: 31135351]
12. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2016 Presented at: International Conference on Learning Representations; 2016 May 2-4; San Juan, Puerto Rico p. 1-16 URL: https://arxiv.org/abs/1511.06434

13. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. 2004 Presented at: Thirty-Seventh Asilomar Conference on Signals, Systems and Computers; 2003 Nov 9-12; Pacific Grove, CA, USA p. 1402-1402. [doi: 10.1109/acssc.2003.1292216]
14. Liou C, Huang J, Yang W. Modeling word perception using the Elman network. Neurocomputing 2008 Oct;71(16-18):3150-3157. [doi: 10.1016/j.neucom.2008.04.030]
15. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: IEEE. USA: IEEE; 2016 Dec 12 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA; 27-30 June 2016; as Vegas, NV, USA p. 770-778 URL: http://ieeexplore.ieee.org/document/7780459/ [doi: 10.1109/cvpr.2016.90]
16. Cao Z, Simon T, Wei S, Sheikh Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In: IEEE. Honolulu, HI: IEEE; 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; 21-26 July 2017; Honolulu, HI, USA p. 7291-7299 URL: http://ieeexplore.ieee.org/document/8099626/ [doi: 10.1109/cvpr.2017.143]
17. Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. 2018 Presented at: International Conference on Learning Representations; 2018 April 30-May 3; Vancouver, BC, Canada.
18. Kingma D, Welling M. Auto-Encoding Variational Bayes. 2014 Presented at: International Conference on Learning Representations; 2014 April 14-16; Banff, Canada p. 1-14 URL: https://arxiv.org/abs/1312.6114
19. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
20. Beinke JH, Fitte C, Teuteberg F. Towards a Stakeholder-Oriented Blockchain-Based Architecture for Electronic Health Records: Design Science Research Study. J Med Internet Res 2019 Oct 7;21(10):e13585. [doi: 10.2196/13585] [Medline: 31593548]
21. Rowe M. An Introduction to Machine Learning for Clinicians. Acad Med 2019;94(10):1433-1436. [doi: 10.1097/acm.0000000000002792] [Medline: 31094727]
22. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286. [doi: 10.2196/12286] [Medline: 30950797]

PART 4

Discussion, summaries
and appendices



CHAPTER 10

Discussion and
future perspectives

Introduction

The clinical assessment of disability in patients with Multiple Sclerosis (MS) is not straightforward. This is largely due to the heterogeneous nature of the disease, the variable contribution of symptoms and signs to disability, the slow accumulation of disability and confounding factors. Although many clinical measurements exist to aid disability assessment, each measurement has specific limitations and caveats, which are described in **chapter 2**.

The traditional and most widely used outcome measure to assess disability in MS trials is the Expanded Disability Status Scale (EDSS), which is regularly used in daily practice as well (schematic representation in figure 1). The EDSS intends to capture the disability of MS patients based on neurological examination, ambulation and ability to carry out activities of daily living (ADL). An overall score is given on an ordinal scale ranging from 0 (normal neurological examination) to 10 (death due to MS). Despite many limitations and caveats, the EDSS is still the most widely used measure. **Chapter 2** describes the attempts that have been made to (i) improve (the practical use of) the EDSS, and (ii) develop other measures to enhance the ability to assess disability in MS.

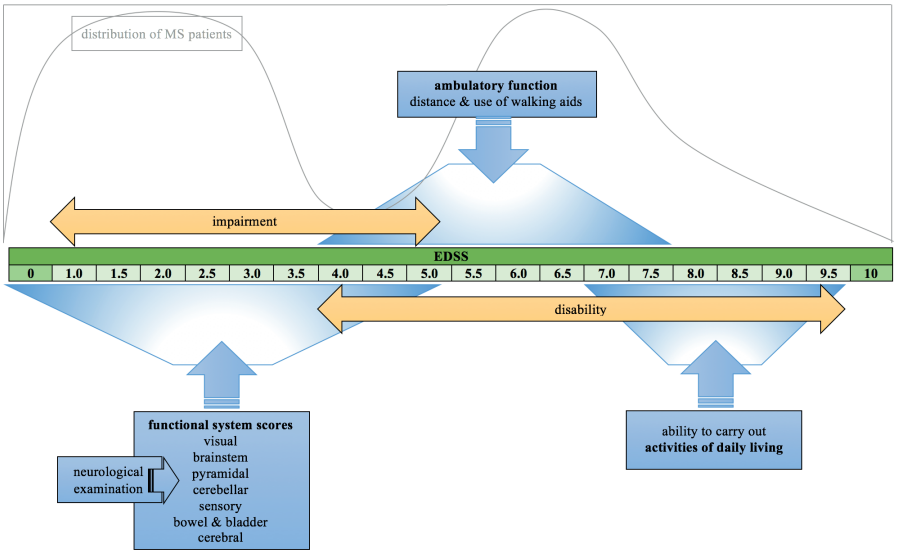


Figure 1, schematic representation of the Expanded Disability Status Scale (reproduced from van Munster et al. with permission from Springer Nature).(1)

Clearly, good clinical assessment is pivotal for (i) early and adequate diagnosis of MS, (ii) assessment of disease activity and progression to evaluate treatment indication and response, and (iii) choosing the right treatment for the right clinical phenotype. As the range of treatment options continues to expand, treatment goals are becoming more ambitious (with the ultimate objective in future being no evidence of disease activity (NEDA)). Consequently, clinical assessment in adjunct to paraclinical measures such as MRI, has become increasingly relevant. In this thesis, novel insights into the clinical assessment of disability in patients with MS are presented.

Main findings

Part II – Clinical assessment of upper extremity function and mobility

The studies described in this part of the thesis aimed to improve the assessment of disability in the functional domains of upper extremity function (UEF) and mobility. It has been demonstrated that assessment of disability in MS can be improved when (i) UEF and ambulation are assessed independently (**chapter 3**), and (ii) tasks of ADL are used in conjunction to other measures (**chapter 4**). In addition, a minimal clinically important difference (MCID) of improvement was calculated for the Arm Function in Multiple Sclerosis Questionnaire (AMSQ), which is a patient-reported outcome measure (PROM) of UEF that was specifically developed for MS patients (**chapter 5**). Lastly, improved evaluation of treatment effects was demonstrated using multimodal assessment of UEF and mobility, and subgroup analyses (**chapter 6**).

Independent assessment of UEF and ambulation

Historically, the focus of clinical assessment of disability in MS patients has been on ambulatory impairment. However, although many patients experience impaired ambulation, often already early in the disease course, other functional domains such as UEF, cognition and vision are frequently affected as well.(2-4) Fortunately, the importance of assessing these initially under-recognized domains is now generally acknowledged. Identifying deficits in these functional domains, and characterizing their magnitude and impact is important in MS management. However, the ways in which UEF and ambulation are exhibited and interact during the disease course are largely unknown. To develop the understanding of this, various aspects of UEF across different levels of ambulation were cross-sectionally examined in 247 patients. These results are presented in detail in **chapter 3**, and key findings are summarized as follows.

Over 80% of patients showed some impairment of UEF, even when ambulation was still unaffected. Patients with worse ambulation displayed

worse performance on all UEF measures, including ADL tests and a PROM. Remarkably, most patients showed only mild UEF impairment, even when ambulation was severely impaired (EDSS 6.0 – 7.0). To conclude, UEF and ambulation show distinct patterns of impairment, affecting multiple aspects of UEF. These findings underscore the importance of adopting multimodal measurement of UEF in the assessment of disability in MS, independent from ambulation. A previous study supports this by showing a concurrent deterioration of various aspects of UEF with disability accrual.(5)

Although these data are derived cross-sectionally, the high prevalence of mild UEF impairment in all EDSS groups may suggest that UEF is affected early on in the disease course, and deteriorates more slowly than ambulation. This mirrors findings from an earlier study which describes the high prevalence of UEF impairment in patients with low EDSS scores is described in one earlier study.(3) Another study revealed that patients frequently report some level of UEF impairment within the first year of onset.(6) Findings from this longitudinal study suggest an early, and more gradual deterioration of UEF relative to ambulatory impairment, which deteriorated more rapidly.(7) Although the methodology of this study differed from the study in **chapter 3**, the data are indicative of a similar trend.

What aspects of UEF are affected, and when these effects occur during the disease course, is important because it potentially widens the treatment window and the opportunity for rehabilitation strategies targeted towards impairment. This is clearly illustrated by several studies that evidence treatment efficacy on UEF, including in SPMS patient with natalizumab,(8) in PPMS patient with ocrelizumab,(9) and in rehabilitation studies with various training programmes.(10)

Assessment of UEF and mobility with ADL tasks

In the study presented in **chapter 3**, patients were asked to perform several movements to evaluate UEF and mobility, including a task of ADL and several classical neurological tests. Patients were selected from an international multicentre study, which developed the *Assess MS system* with the aim to automatically quantify motor functioning using Machine Learning Algorithms (MLA). **Chapter 4** outlines a study that evaluated to what extent a combination of movements can contribute to the assessment of UEF and mobility.

Multivariate linear regression models were used to determine to what extent the (combinations of) movements explained the variance in the Nine-Hole Peg Test (9HPT) and AMSQ for UEF, and the Timed 25-foot Walk Test (T25FW) for mobility. The variances in 9HPT and T25WT scores were largely explained by the combinations of movements, and, importantly, ADL tasks contributed

most to the variances. This finding indicates that including ADL tasks (such as drinking from a cup, or standing up from a chair) in daily practice or clinical trials may be more valuable than classical neurological test (such as placing a fingertip on one's nose) in terms of assessing MS. This result also highlights a caveat of the EDSS. Whilst classical neurological examination plays a central role in the lower EDSS range (0 – 4.0), and ADL in the higher range (7.0 – 9.5), clinically relevant impairment of UEF and mobility remains probably undetected. This reasoning favours the use of multiple measures, rather than one-size-fits-all approaches to assess disability.

At present, the 9HPT and T25FW are often used in clinical practice and in trials as secondary outcome measures. Both correlate with the ability to perform ADL tasks.(11, 12) However, the ADL movements used in the study presented here can be administrated more quickly, and more easily. Further, the neurologists who rated the movements found the scales used for the ADL tasks much easier to apply than those of the classical neurological examination (personal communication). Nevertheless, the psychometric properties of the ADL tasks should be examined before implementing them in clinical practice or trials. This will require further testing in a more disabled population than was included in this study.

MCID for AMSQ

Performance measurements such as those described in **chapter 3 and 4**, are valuable tools in disability assessment, however they do not give any insight into the patient's perspective of impairment. Therefore, complementary PROMs can be used to fulfil this purpose. Many PROMs are being used in the MS field, with some being developed specifically for use with MS patients (**chapter 2**). The AMSQ, for example, was designed to assess the impact of MS on patient-perceived UEF. The questionnaire covers a variety of ADL tasks, ranging from gross to fine movements, and has valuable psychometric properties.(13, 14) Notably, PROMs generally detect clinically meaningful change, discarding changes with no clinical relevance.

As described in **chapter 5**, the MCID was determined to be 15 points. A cohort of patients that were treated with fampridine was used in a sensitivity and specificity anchor-based approach. We acknowledge that future studies will have to confirm our findings and dispel remaining uncertainties, preferably with other statistical methods.(15, 16) Nonetheless, this MCID value was used in the study presented in **chapter 6**.

Multimodal assessment of UEF and ambulation, and subgroup analyses

Building on the findings in **chapters 3 through 5**, the treatment response to fampridine was examined in a cohort of MS patients, who were selected from a real-world setting using a process described in **chapter 6**. It is particularly interesting to assess UEF and mobility independently, using multiple measures, including a PROM for UEF, in a cohort such as this. Firstly, it may give insight into the discrepancy between a physicians' overall judgement of improvement of motor functioning in cases where, re-imbursement policies dictate insufficient efficacy. Secondly, subgroups of patients may be identified that have the largest treatment effects on specific motor functions. Clinically relevant benchmarks and MCID values were used as cut-off points to determine such subgroups. Using this method, it may be possible to discriminate between patients with and without clinically relevant improvement. Lastly, scrutinizing a real-world cohort with such methods may give insight into treatment decision-making in clinical practice.

In total, 275 patients reporting ambulatory impairment were analysed. At a group level, patients treated with fampridine showed statistically significant improvements on the T25FW and MS walking scale ([MSWS] which is a PROM to assess perceived difficulties on several ambulation tasks in daily life). This result justified the continuation of fampridine treatment in the vast majority of patients. Notably, response rates in terms of improved T25FW score were higher amongst patients who were more severely disabled (EDSS 6.0-7.0) or exhibited more impaired ambulation. This might suggest a higher efficacy in more severely affected patients. Alternatively, it may also merely reflect a floor effect of the T25FW.(17) In contrast, the proportion of patients with MSWS improvement was larger amongst patients with less impaired ambulation. This may reflect a certain placebo effect, likely to have also influenced treatment decision-making, because no further proportional difference between subgroups was found amongst patients continuing treatment. Nonetheless, the MSWS may still be useful in complement with the T25FW to evaluate ambulation in less impaired patients.

A subgroup of 183 patients who also reported UEF impairment were subsequently examined. On average, a statistically significant improvement was found in AMSQ while on treatment, and approximately 40% of cases were classified as an AMSQ responders. In this subset of the cohort, patients with AMSQ improvement were more likely to show improvement on MSWS and continue treatment, than patients without improvement of AMSQ. This finding suggests that there is an influence of patient-reported outcome on treatment decision making in daily practice. Obviously, physician-based factors (such as experience, training, influenceability) and local factors (such as reimbursement policies) influence treatment decision making as well. It nevertheless illustrates the gap between clinical trials and practice, and the challenges physicians are being faced with in the real-world.

These observations, together with other data, contributed to the revision of the initially negative decision on reimbursement for fampridine made by regulatory authorities in the Netherlands. Reimbursement is now being granted for patients with more severely impaired ambulation, meaning a T25FW performance of 6 seconds or worse, and an EDSS of 4.0 – 7.0. A strict improvement of at least 20% on the T25FW is a condition for reimbursement. Before these revisions, reimbursement criteria were less strict with the main selection criterion being an EDSS between 4.0 and 7.0, which included patients with less severe ambulatory impairment. As described in the study described in **chapter 6**, efficacy on ambulation in this subgroup of patients is less clear, and patient-reported outcome probably influenced the decision to continue treatment. Consequently, treatment was continued for a certain proportion of patients who likely only had a placebo-effect, leading to cost-inefficacy. The study in **chapter 6** therefore exemplifies the practical and policy-relevance, as well as the challenges of good clinical assessment of UEF and mobility in MS patients.

Part III – Video-assisted assessment of motor functioning

Part III describes three studies in which several aspects of video-assisted assessment of motor functioning are evaluated. Video-assisted assessment has the potential to improve evaluation of disability in several ways. It may improve certain psychometric properties, such as intra- and interobserver variability, of performance-based measures (e.g. finger-to-nose test (FNT)), and promotes data quality by allowing a visual check on compliance relative to the standardization of performance. Furthermore, digital video-data can be used for building MLA that automatically quantify motor performance. Ultimately, automated video-assisted assessment of disability can be performed by patients themselves, at their preferred location and time, without needing to be physically present in a hospital or research clinic. The COVID-19 pandemic, which imposed strict protocols on 1.5-meter social distancing to safely assess patients, combined with a preference for home-based assessments, particularly for vulnerable patients with chronic diseases or immunosuppression, emphasises the relevance this flexibility. video-assisted assessment can also contribute to shorter administration times and associated administrative burdens, relative to other assessments, particularly the EDSS, which takes at least 15 – 20 min to obtain.

In this part, it is established that the use of reference videos can reduce the variability of motor functioning assessment (**chapter 7**). Furthermore, it is determined the detection of change in UEF and mobility can be improved when video-assisted composite measures are used in conjunction with conventional measures (**chapter 8**). Lastly, a proof of concept study demonstrates that autoencoders are a potentially valuable method for the analyses of patient videos while preserving data privacy (**chapter 9**).

Reference videos

The EDSS is known for its high intra- and interrater variability, especially in the lower ranges if the score.(18) This is partly a consequence of variability in performing a neurological examination, including the FNT. The study presented in **chapter 7** investigated whether the rater variability of such a test can be reduced by using reference videos. These are videos of patients performing a movement according to specific test instructions, where the movements have been rated according to a predefined rating scale. In the case of the FNT, for example, a patient is instructed to perform the test with eyes closed and holding the arm 90 degrees abducted in the shoulder, and subsequently putting the tip of the index finger on the top of the nose. The movement is rated according to the Neurostatus-EDSS definitions. (19) Collectively, a set of reference videos showcases all possible degrees of performance on the test., from high to low. As such, a set can be used as a benchmark to assist the rater in accurately rating the clinical videos presented.

As expected, the use of reference videos substantially reduced the intra- and interrater variability of video-assessment of FNT, as compared with the sole use of the textual rating scale. These findings confirm the common impression of participating rating neurologists, including the author of this thesis, that rating of videos was much easier to do when using reference videos (personal communication). Further, the software used to display the reference videos was user-friendly and easily integrated into practice. To conclude, reference videos improve clinical assessment of motor functioning, and could be present a valuable addition if integrated in daily practice and clinical trials.

Video-assisted composite measures to detect change in UEF and mobility

Building on the finding that using reference videos facilitated video-assisted rating of motor performance, it is likely that multidimensional assessment (including multiple tests) can more broadly be facilitated using videos. Combining multiple tests into a composite measure allows a more complete assessment of change in motor performance. The difficulties of documenting clinical change in MS has been repeatedly illustrated throughout this thesis. Therefore, the study presented in **chapter 8** investigated the value of multidimensional video-assisted composite measures to enhance detection of change in UEF and mobility. For this purpose, a cohort of 43 patients treated with fampridine was analysed. Fampridine can quickly improve several motor functions (within two weeks of treatment), and thus offers the possibility to observe change in the scope of a short follow up. By contrast, performing such a study in a natural disease course cohort would take several years.

Composite measures were determined for UEF and one for mobility, and compared the detection of improvement with the conventional measures of

9HPT and T25FW. The composite measures consisted of tasks of ADL and classical neurological tests, which were performed by patients performed who were recorded on video. These videos were subsequently rated by a neurologist, using reference videos such as described in **chapter 7**. To confirm whether the improvement in composite measures that was not detected with the conventional measures was clinically relevant, a global rating of change scales was used to assess if patients perceived the change.

When used in conjunction with conventional measures (9HPT and T25FW in this study), the video-assisted composite measures identified additional patients who exhibited clinically relevant change. This implies that the use of a series of videos of patients performing movements can be useful to monitor disease course and evaluate treatment outcome. Novel methods that improve clinical disease monitoring in progressive phenotypes are particularly welcome, considering the recent introduction of disease modifying therapies (DMT) for these patients. Further, these results indicate that assessing motor functioning on video is feasible, and might even have the potential to replace conventional measures. As such, clinical assessment would no longer have to be confined to specific time slots or locations that primarily suit the physician or researcher. However, these conclusions to date are based on findings in a small cohort with no control, and thus future studies will have to replicate and re-investigate the hypothesis before video-based composites can be justifiably integrated in daily practice or clinical trials.

Autoencoders for preserving data privacy

Video-assisted assessment has the potential to improve the assessment of motor functioning in MS by several ways, such as increased accuracy,(20) reliability (**chapter 7**) and sensitivity to change (**chapter 8**). Moreover, the video data can be used for building MLA which can be used for example to automatically assess motor functioning. In the current digital age, MLA and artificial intelligence (AI) are increasingly used in various areas of healthcare research.(21) However, if implemented, this implies the need for collaboration with IT specialists and other non-healthcare workers, which may pose a threat to data privacy. The issue of data security and privacy remains pivotal in modern society, as reflected by contemporary public debates. The use of autoencoders may help to overcome this barrier.

Autoencoders are a digital technique that embed visual information into a latent space, which essentially means that the visual data are compressed into a form in which similar data points are closer together in space. The compressed data preserve information needed for MLA development, but is not visually interpretable by humans. Simply put, an autoencoder contains an encoder that compresses the data, coupled with a paired decoder that transforms the compressed data back into the original video. A schematic

representation is depicted in figure 2. However, the question remains as to whether the compressed data, which can be used for MLA building, still retains the clinically relevant data when decompressed.

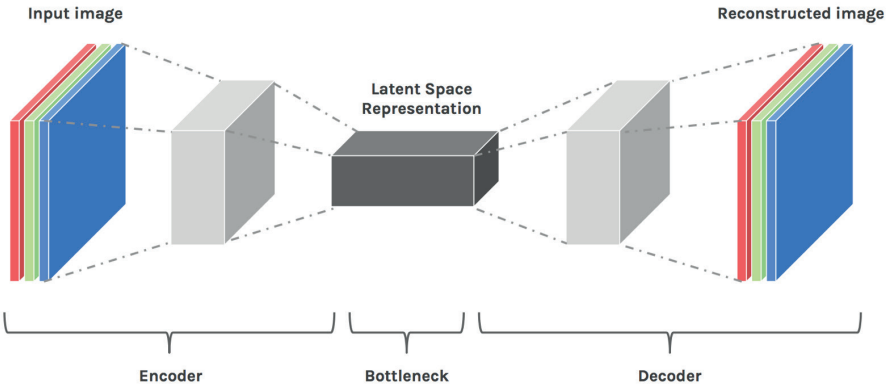


Figure 2, schematic representation of an autoencoder.(22)

In the "proof of concept" study presented in **chapter 9**, this issue has been assessed. Videos of patients performing the FNT were rated by 10 neurologists before encoding and after decoding, and these ratings were subsequently compared to see how closely they agreed. The vast majority of videos decoded by the autoencoder contained clinically relevant information and had a fair intra-rater agreement with the original video. To conclude, autoencoders are a potentially valuable method in healthcare research to preserve data privacy, although iterative testing is required to confirm this conclusion prior to implementation.

Future perspectives

What determines the appropriateness of clinical outcome measurements for daily practice and clinical trials in MS? The right choice of measurements is central to clinical decision making in both diagnosis and treatment, and as well as for assessing the efficacy of novel DMT. Consequently, the measures chosen will influence patient prognosis in daily practice, and determine if a drug can proceed to the next step of the development process and obtain regulatory approval.

Many of the outcome measures used in clinical trials are also implemented in daily practice. However, the levels of standardization and quality control of assessments, both of which are mandatory in trials, are rarely met in daily practice.(23) Efforts have been made in trials to implement outcome measures that are commonly used in daily practice, particularly to monitor treatment response.(24, 25) The well-known EDSS, which is traditionally used

in trials, is also widely used in daily practice, although it is more challenging to apply consistently than might be assumed at first encounter. EDSS also has the advantage that it is immediately understood by clinicians, which is not always the case for novel measures. Nevertheless, novel measures are generally superior in detecting more subtle changes in disability. Other important considerations for measures to be implemented in daily practice include ease of use (e.g. administration time of EDSS is inconveniently long, requiring at least 15 – 20 minutes), the clinical utility and ease of interpretation of the data collected (e.g. what is the clinical relevance of a given change score?).

The findings of this thesis contribute to the resolution of these clinical challenges, specifically:

- Clinical assessment of disability can be improved when (i) UEF and ambulation are assessed independently (chapter 3), (ii) tasks of ADL are used in conjunction to other measures (chapter 4), and (iii) a value for MCID of improvement for AMSQ is determined (chapter 5)
- Multimodal assessment of UEF and mobility, and subgroup analyses improve evaluation of treatment effects (chapter 6)
- Reference videos reduce the variability of motor functioning assessment (chapter 7)
- Detection of change in UEF and mobility can be improved when video-assisted composite measures are used in conjunction with conventional measures (chapter 8)
- Autoencoders are a valuable method to preserve data privacy in analyses of patient videos (chapter 9).

In the last part of this chapter possible future perspectives of the clinical assessment of MS patients in daily practice and clinical trials of MS will be elaborated on. A special emphasis is placed on the potential of biosensor-based technology. Finally, the benefits of collecting data in multidisciplinary data infrastructures will be highlighted.

Clinical assessment in daily practice and clinical trials of MS

As elaborated previously, the complexity of clinical assessment in MS is largely a consequence of its heterogeneous symptomatology. A logical consequence is that patients cannot be fully evaluated with a single, universal and all-encompassing measure. Therefore, multiple measures are needed to cover all relevant clinical aspects of MS. Combining multiple measures to better capture (multidimensional) change is the core concept of composite measures. A composite measure can be tailored to a specific domain of interest or broadened to cover multiple functional areas. Therewith, detection

of infrequent events such as relapses, or small changes such as disease progression, can be improved. This is particularly relevant for patients under (highly effective) treatment and with progressive MS phenotypes, in which small and gradual treatment effects can be expected. Furthermore, trial duration and size may be reduced through the use of composite measures. Notably, it is important to differentiate composite measures from composite scores. A composite score is a statistical method to combine the results of several tests into a single score. The resulting score can be difficult to interpret and is not always widely accepted. Other limitations and caveats of composite measures are summarized in table 6 of **chapter 2**.

What are the minimum functional clinical domains that should be assessed to monitor MS? The most commonly affected domains in MS are ambulation, UEF, cognition and vision. Problems in these domains can be found in approximately three-quarter of patients.(2-4) It would therefore make sense to include these domains in a basic set of tests. Measures with good psychometric properties and are already well-established in the MS field should also be considered for inclusion in this basic set. The inclusion of a PROM is particularly important because it gives healthcare professionals insight into the patient's perspective of a certain aspect of the disease or matter of interest. This aspect should not be underestimated, because the physicians' view does not necessarily agree with the patients' view. Patient and physician perspectives may diverge on, *inter alia*, what symptoms are fundamental for perception of health,(26, 27) and preference for specific attributes of DMT.(28) Some PROMs that focus on a specific functional domain, such as UEF (AMSQ) or ambulation (MSWS), assess perceived difficulties on performing several ADLs in that domain. The aspect of ADL may be valuable, because the ability to perform (instrumental) ADL tasks can discriminate between different levels of disability and predict disability progression.(29) To reduce the patient burden of filling in multiple PROMs, computer adaptive testing should be further developed for all PROMs.(30) However, disadvantages and caveats of PROMs should be taken into account (summarized in table 3 of **chapter 2**) when including them into the basic set.

With these considerations in mind, a set of valuable measurements is proposed in table 1 that may be considered as a basic clinical composite to monitor disease course and treatment efficacy. This set can be extended based on the individual characteristics of a patient, preferably with measures that are designed or well established for MS patients (see **chapter 2**). In the end, the measurements chosen for application in daily practice should be tailored to the individual situation to achieve a holistic view of a patient. For example, the AMSQ might me more important to use in patients with severely impaired ambulation, since UEF becomes more important to maintain self-dependence when ambulatory function deteriorates. On the other hand, one could argue that a yearly assessment of quality of life should be applied to all patients.

Table 1, proposition of a basic set of measurement to monitor disease course and treatment efficacy

Functional domain	Measure
Ambulation	Timed 25-Foot Walk test, to assess walking speed for patients with more severe ambulatory impairment, or the 6 Minute Walking Test to assess walking speed in patients with relatively mild ambulatory impairment.(31-33) *PROM*: MS walking scale.(34)
Upper extremity function	Nine-Hole Peg Test (NHPT) to assess manual dexterity.(11) *PROM*: Arm Function in Multiple Sclerosis Questionnaire.(13)
Cognition	Symbol Digit Modalities Test to assess information processing speed.(35)
Vision	Low-contrast letter acuity test to assess visual acuity.(36)

PROM = patient-reported outcome measure.

In addition to the clinical assessment, complementary paraclinical methods should be used to produce a holistic assessment of MS patients. In addition to basic MRI measures such as (gadolinium enhancing) lesion load, other valuable measures that are broadly implemented in clinical trials as a secondary endpoint are slowly infiltrating daily practice. Examples are optical coherence tomography and biomarkers in body fluids such as neurofilament light chain. Grey matter pathology and volumetric assessment of the brain and spinal cord are also potentially valuable, as they are associated with physical and cognitive functioning.(37) The exact value of these measures in daily practice have yet to be determined. If they are clinically meaningful, they may be particularly useful in disease monitoring of progressive MS, because the identification of progression or neurodegenerative changes remains very challenging. Further elaboration about paraclinical measures is beyond the scope of this thesis.

Clinical monitoring of the progressive disease course in SPMS and PPMS cases is particularly challenging. This is increasingly important because the treatment window is opening up for these phenotypes.(38) The main reason that the assessment of clinical progression is so difficult is that it generally accumulates slowly. The duration of MS trials is commonly limited to 2-3 years, which is insufficient time to effectively assess progression (and certainly improvement), particularly in progressive phenotypes. Further, the development trajectories of impairment differ across functional domains, as does the clinical relevance of certain functional changes over the disease course (**chapter 3**). For example, UEF might deteriorate in the early stages of the disease, but it becomes most relevant when ambulation becomes impaired, since the ability of wheelchair-bound patients to perform ADL tasks is more dependent on UEF. This means that the clinical relevance of different functional domains differs throughout the disease course. If these facets are

not taken into account, possible positive treatment effects may be missed.(8, 9) On this note, the choice of inadequate clinical outcome measures has been cited as a possible explanation for negative trails in the MS field.(39)

What we deem necessary to include in the clinical assessment of patients with MS also depends on how much we want to raise the bar. For example, the concept of NEDA as the ultimate treatment goal of a *disease-activity-free status* can be handled as the ultimate treatment goal. The term was been originally described as the absence of relapses, EDSS progression and disease activity on MRI.(40) One important argument for setting this as the objective is its high predictive value for achieving the absence of disability progression in the long run. Specifically, the predictive value was greater for NEDA than for any of its individual components.(41) Notably, the timing of assessing NEDA after the initiation of treatment is crucial to its interpretation.(1) Raising the bar higher by adding more components to NEDA (e.g. with a PROM or brain volume measures) will give a more sensitive composite to detect change. A downside is that this will also reduce the number of patients fulfilling NEDA, which may lead to the rejection of highly effective DMT, as well as possibly more severe side-effects.

In order to achieve more holistic clinical assessment, a paradigm shift from physician/centre-based care to patient-based care is necessary. In centre-based care, the healthcare professional is put at the centre of the management process. Via various clinical tests, usually performed within a hospital, information is conveyed to the healthcare professional, and choices concerning periodical assessment, pharmacological treatment and rehabilitation strategies are made. In a patient-based care system, the patient is actively involved in this process. In addition to clinical tests performed at the centre, the patient can report about certain aspects of the disease via PROMs, and modern methods such as biosensor-based technology (described in the next paragraph) deliver personalised, on-demand data to the team, which supports treatment decision making based on specific needs of a patient. Importantly, in this system of care a patient should never feel alone. Instead, there is a need for more education and supervision to support the patient.(42) As we have recently experienced in the COVID-19 pandemic, e-health (tele-health) consulting is acceptable to many patients and can be more easily used than was previously perceived. It also has several advantages over classical physical contacts in the hospital, such as clinical sorting of care and more frequent contacts.(43)

Biosensor-based technology

The clinical measurement of MS patients discussed so far is largely based on the classical neurological examination, which can be seen as the *ground truth* for diagnosis and assessment of the disease course. This elegant examination

dates back as far as the end of the 19th century. However, modern 21st century technologies potentially offer new ways to capture more subtle clinical signs and changes over time. Biosensors are used in these techniques to monitor and convert biological human signs into electrical signals and digital data, to be subsequently analysed using different methods (such as MLA for AI) for a specified goal.

Biosensors can be broadly categorized into (i) small molecular sensors that are used inside tissue or externally, and (ii) wearable/ wireless sensors.(44) In the MS field the first category has not yet emerged fully. On the contrary, wearable/wireless biosensors are increasingly becoming the subject of research, with some already being used in daily practice. The most extensively evaluated – and perhaps most promising - wearable biosensors in MS are accelerometers, either with or without gyroscopes. Other less well examined biosensors are grip sensors, electrodermal sensors and surface/ portable electromyography.(42)

Accelerometers can track various motor functions. Continuous step counts and measurements of physical activity are already widely used in normal daily life, as an integral part of smartphones and -watches. The addition of gyroscopes to accelerometers improves their accuracy to detect falls and movement to stand up from a chair, as well as quantify gait, balance and tremor.(42, 44) Various devices have been tested in MS patients.(42) Examples are the Floodlight Open,(45) and Multiple Sclerosis Performance Test.(46) A strong trait of accelerometers is that they allow continuous assessment of motor functioning during a person's daily life. Thus, an enormous amount of data can be collected for analyses. More extensive longitudinal studies of accelerometers are needed to assess predictive value of long-term disability and accuracy to detect small changes, which is particularly important for progressive MS phenotypes.

A more high-tech biosensor technique that was recently introduced is based on wireless radio reflection signals.(47) No wearable sensor is needed. This method uses a device that emits low-power radio frequency signals, which is placed at home or any other location. The waveforms are changed when they come into contact with a body moving in space. These changes are detected by the device and subsequently analysed using MLA and AI. In this way a person's motion, sleep pattern and falls can be analysed. This futuristic biosensor technique sounds appealing, although it may raise substantive privacy issues, which will have to be resolved before it can be implemented in daily practice or clinical trials.

Video-recording of movements is another sensor-based technique. Several chapters of this thesis use data from the multinational project *Assess MS*, which aimed to develop a system to assess motor functioning in MS by using a colour and 3D-depth camera (Kinect from Microsoft).(20, 48-52)

The hardware and software used in *Assess MS* was usable and acceptable to both patients and healthcare professionals, and generated high quality data suitable for clinical analysis.(51) This technique has the potential to provide a more nuanced measure for motor functioning. The *Assess MS* study aimed to develop a system that could automatically assess motor functioning by developing MLA for AI. A major advantage is that the assessment could be performed by patients themselves, at a preferred location and time, without the need to be physically present in a hospital or research clinic. Unfortunately, the study has not yet succeeded to develop a MLA that was able to assess motor functioning sufficiently. The amount of data in colour- and 3D-depth videos was too large for the MLA to quantify movements at the level required in daily practice and clinical trials. In this case, the human brain still seems to prevail over AI.

The implementation of biosensor-based technology for the automatic measurement of physical signs will improve the ability to assess disability in MS. Biosensors can provide additional information to complement conventional clinical measurements, which will give a finer-grained and richer presentation of disability. During the coming decades, these techniques are likely to penetrate daily practice and clinical trials. However, the clinical significance, reliability, sensitivity to change and other psychometric properties first have to be established. Other factors such as data privacy, disruption to daily life, costs and user-friendly software will also have to be clarified.

Using real-world data for clinical research: multidisciplinary data infrastructures

The clinical data obtained from MS patients in daily practice and clinical trials are complemented by various paraclinical techniques, and primarily serves to improve patient healthcare and answer scientific questions. However, looking beyond these primary objectives, these hard-earned data may contribute to other goals too.(53) (i) Regulatory authorities can use these data for making reimbursement policies, and to check effectiveness and safety in the real-world. (ii) Pharmaceutical or health-care technology organisations may use data for development of new DMT or health-care improving techniques. (iii) Researchers can increase their knowledge about the disease. (iiii) Neurologists can improve the diagnostic accuracy, prognostication and personalization of treatment. To achieve these, and other, goals, the multidimensional measurements of MS patients will have to be merged into large multidisciplinary data infrastructures.

It is important in MS to pursue multidisciplinary data infrastructures for several key reasons.(53) Firstly, MS is a complex disease for which no universal all-encompassing description exists or will ever be sufficient. Secondly, several aspects of MS symptomatology, treatment and rehabilitation effects are

interdependent, and should therefore be interpreted together. Thirdly, to truly reach personalized healthcare in MS, a measure of that person's fundamental biology is needed, as reflected by a dynamic cluster of technological markers such as body fluid biomarkers and genomics. Lastly, the inclusion of PROMs in the multivariable dataset facilitates the understanding of the patient perspective, putting the patient at the heart of healthcare. These arguments are elaborated in more detail elsewhere.(53)

The number of MS data initiatives in the form of MS databases and (national) registries is growing. In the Netherlands the *MS kwaliteitsregister* started in 2017, and currently more than 60 centres participate which led to the inclusion of data from more than 800 MS patients. Other well-known real-world registries are the international MSBase consortium, North American Research Committee on MS Registry (NARCOMS), German MS Register (GMSR), and United Kingdom MS (UK-MS) Register. Various technological improvements have contributed to facilitate data collection, input and communication with MS patients.(54-56) Looking to the future, multidisciplinary data infrastructures could be linked to establish a foundation for (big) data-analyses. In a recent topical review, three joint actions were proposed to increase the chance of getting to that point.(53) In short, these are the inclusion of multidisciplinary staff for data collection, agreement on a minimal set of relevant measurements, and the implementation of protocols for measurements and common data models (CDM) in data collection procedures. The principle of CDM is that standard data-formats are used across different applications or systems (e.g. opening or reading a word file with a PDF viewer). Recent evidence shows that harmonization across different MS registries is feasible.(57)

Final conclusion

The clinical assessment of MS patients is an exciting field grounded on illustrious historical foundations and deepened through contemporary technological innovation. Further improvement and integration into multidisciplinary data infrastructures will eventually lead to a better understanding and control of MS, which ultimately improves the quality of the life of our patients.

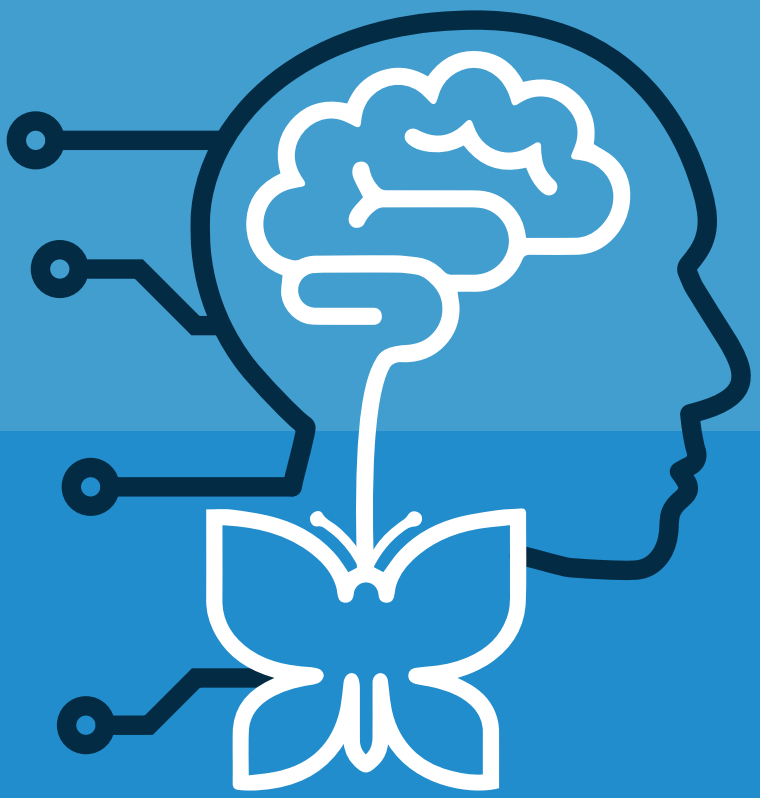
REFERENCES

1. van Munster CE, Uitdehaag BM. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs. 2017;31(3):217-36.
2. Chiaravalloti ND, DeLuca J. Cognitive impairment in Multiple Sclerosis. Lancet Neurol. 2008;7(12):1139-51.
3. Johansson S, Ytterberg C, Claesson IM, Lindberg J, Hillert J, Andersson M, et al. High concurrent presence of disability in Multiple Sclerosis. Associations with perceived health. J Neurol. 2007;254(6):767-73.
4. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in Multiple Sclerosis. I. Frequency, patterns, and prediction. Neurology. 1991;41(5):685-91.
5. Bertoni R, Lamers I, Chen CC, Feys P, Cattaneo D. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with Multiple Sclerosis. Mult Scler. 2015;21(12):1566-74.
6. Kister I, Bacon TE, Chamot E, Salter AR, Cutter GR, Kalina JT, et al. Natural history of Multiple Sclerosis symptoms. Int J MS Care. 2013;15(3):146-58.
7. Cadavid D, Cohen JA, Freedman MS, Goldman MD, Hartung HP, Havrdova E, et al. The EDSS-Plus, an improved endpoint for disability progression in secondary progressive Multiple Sclerosis. Mult Scler. 2017;23(1):94-105.
8. Kapoor R, Ho PR, Campbell N, Chang I, Deykin A, Forrestal F, 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):405-15.
9. Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. N Engl J Med. 2017;376(3):209-20.
10. Lamers I, Maris A, Severijns D, Dielkens W, Geurts S, Van Wijmeersch B, et al. Upper Limb Rehabilitation in People With Multiple Sclerosis: A Systematic Review. Neurorehabil Neural Repair. 2016;30(8):773-93.
11. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for Multiple Sclerosis. Mult Scler. 2017;23(5):711-20.
12. Kempen JC, de Groot V, Knol DL, Polman CH, Lankhorst GJ, Beckerman H. Community walking can be assessed using a 10-metre timed walk test. Mult Scler. 2011;17(8):980-90.
13. Mokkink LB, Knol DL, van der Linden FH, Sonder JM, D'Hooghe M, Uitdehaag BMJ. The Arm Function in Multiple Sclerosis Questionnaire (AMSQ): development and validation of a new tool using IRT methods. Disabil Rehabil. 2015;37(26):2445-51.
14. van Leeuwen LM, Mokkink LB, Kamm CP, de Groot V, van den Berg P, Ostelo R, et al. Measurement properties of the Arm Function in Multiple Sclerosis Questionnaire (AMSQ): a study based on Classical Test Theory. Disabil Rehabil. 2017;39(20):2097-104.
15. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7(5):541-6.
16. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA. 2014;312(13):1342-3.

17. Goldman MD, Motl RW, Scagnelli J, Pula JH, Sosnoff JJ, Cadavid D. Clinically meaningful performance benchmarks in MS: timed 25-foot walk and the real world. Neurology. 2013;81(21):1856-63.
18. Meyer-Moock S, Feng YS, Maeurer M, Dippel FW, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with Multiple Sclerosis. BMC Neurol. 2014;14:58.
19. D'Souza M, Yaldizli O, John R, Vogt DR, Papadopoulou A, Lucassen E, et al. Neurostatus e-Scoring improves consistency of Expanded Disability Status Scale assessments: A proof of concept study. Mult Scler. 2017;23(4):597-603.
20. Burggraaff J, Dorn J, D'Souza M, Morrison C, Kamm CP, Kontschieder P, et al. Video-Based Pairwise Comparison: Enabling the Development of Automated Rating of Motor Dysfunction in Multiple Sclerosis. Arch Phys Med Rehabil. 2020;101(2):234-41.
21. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res. 2019;21(4):e12286.
22. https://hackernoon.com/latent-space-visualization-deep-learning-bits-2-bd09a46920df [
23. Tur C, Moccia M, Barkhof F, Chataway J, Sastre-Garriga J, Thompson AJ, et al. Assessing treatment outcomes in Multiple Sclerosis trials and in the clinical setting. Nat Rev Neurol. 2018;14(2):75-93.
24. Freedman MS, Selchen D, Arnold DL, Prat A, Banwell B, Yeung M, et al. Treatment optimization in MS: Canadian MS Working Group updated recommendations. Can J Neurol Sci. 2013;40(3):307-23.
25. Sormani MP, Rio J, Tintore M, Signori A, Li D, Cornelisse P, et al. Scoring treatment response in patients with relapsing Multiple Sclerosis. Mult Scler. 2013;19(5):605-12.
26. Green R, Cutter G, Friendly M, Kister I. Which symptoms contribute the most to patients' perception of health in Multiple Sclerosis? Mult Scler J Exp Transl Clin. 2017;3(3):2055217317728301.
27. Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in Multiple Sclerosis. BMJ. 1997;314(7094):1580-3.
28. Kremer IEH, Evers S, Jongen PJ, Hiligsmann M. Comparison of preferences of healthcare professionals and MS patients for attributes of disease-modifying drugs: A best-worst scaling. Health Expect. 2018;21(1):171-80.
29. Salter A, Fox RJ, Tyry T, Cutter G, Marrie RA. New applications for independent activities of daily living in measuring disability in Multiple Sclerosis. Mult Scler. 2020:1352458519898591.
30. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther. 2014;36(5):648-62.
31. Goldman MD, Marrie RA, Cohen JA. Evaluation of the six-minute walk in Multiple Sclerosis subjects and healthy controls. Mult Scler. 2008;14(3):383-90.
32. Learmonth YC, Dlugonski DD, Pilutti LA, Sandroff BM, Motl RW. The reliability, precision and clinically meaningful change of walking assessments in Multiple Sclerosis. Mult Scler. 2013;19(13):1784-91.

33. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for Multiple Sclerosis. Mult Scler. 2017;23(5):704-10.
34. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). Neurology. 2003;60(1):31-6.
35. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for Multiple Sclerosis. Mult Scler. 2017;23(5):721-33.
36. Balcer LJ, Baier ML, Cohen JA, Kooijmans MF, Sandrock AW, Nano-Schiavi ML, et al. Contrast letter acuity as a visual component for the Multiple Sclerosis Functional Composite. Neurology. 2003;61(10):1367-73.
37. van Munster CE, Jonkman LE, Weinstein HC, Uitdehaag BM, Geurts JJ. Gray matter damage in Multiple Sclerosis: Impact on clinical symptoms. Neuroscience. 2015;303:446-61.
38. Ontaneda D, Cohen JA, Amato MP. Clinical outcome measures for progressive MS trials. Mult Scler. 2017;23(12):1627-35.
39. Chataway J. Inadequate outcome measures are the biggest impediment to successful clinical trials in progressive MS - YES. Mult Scler. 2017;23(4):505-6.
40. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing Multiple Sclerosis. N Engl J Med. 2006;354(9):899-910.
41. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal Multiple Sclerosis cohort. JAMA Neurol. 2015;72(2):152-8.
42. Brichetto G, Pedulla L, Podda J, Tacchino A. Beyond center-based testing: Understanding and improving functioning with wearable technology in MS. Mult Scler. 2019;25(10):1402-11.
43. Capra R, Mattioli F. Tele-health in neurology: an indispensable tool in the management of the SARS-CoV-2 epidemic. J Neurol. 2020.
44. Graves JS, Montalban X. Biosensors to monitor MS activity. Mult Scler. 2020;26(5):605-8.
45. https://floodlightopen.com/en-US [
46. Rhodes JK, Schindler D, Rao SM, Venegas F, Bruzik ET, Gabel W, et al. Multiple Sclerosis Performance Test: Technical Development and Usability. Adv Ther. 2019;36(7):1741-55.
47. Yu F, Bilberg A, Stenager E. Wireless medical sensor measurements of fatigue in patients with Multiple Sclerosis. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:3763-7.
48. D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in Multiple Sclerosis. Mult Scler J Exp Transl Clin. 2018;4(3):2055217318792399.
49. D'Souza M, Van Munster CEP, Dorn JF, Dorier A, Kamm CP, Steinheimer S, et al. Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study. J Med Internet Res. 2020;22(5):e16669.
50. Kontschieder P, Dorn JF, Morrison C, Corish R, Zikic D, Sellen A, et al. Quantifying progression of Multiple Sclerosis via classification of depth videos. Med Image Comput Comput Assist Interv. 2014;17(Pt 2):429-37.

51. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors. 2015;2(1):e11.
52. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in Multiple Sclerosis. Mult Scler. 2019;25(12):1673-81.
53. Peeters LM, van Munster CE, Van Wijmeersch B, Bruyndonckx R, Lamers I, Hellings N, et al. Multidisciplinary data infrastructures in Multiple Sclerosis: Why they are needed and can be done! Mult Scler. 2019;25(4):500-9.
54. Gourraud PA, Henry RG, Cree BA, Crane JC, Lizee A, Olson MP, et al. Precision medicine in chronic disease management: The Multiple Sclerosis BioScreen. Ann Neurol. 2014;76(5):633-42.
55. Katzan I, Speck M, Dopler C, Urchek J, Bielawski K, Dunphy C, et al. The Knowledge Program: an innovative, comprehensive electronic data capture system and warehouse. AMIA Annu Symp Proc. 2011;2011:683-92.
56. Ziemssen T, Kempcke R, Eulitz M, Grossmann L, Suhrbier A, Thomas K, et al. Multiple Sclerosis documentation system (MSDS): moving from documentation to management of MS patients. J Neural Transm (Vienna). 2013;120 Suppl 1:S61-6.
57. Salter A, Stahmann A, Ellenberger D, Fneish F, Rodgers WJ, Middleton R, et al. Data harmonization for collaborative research among MS registries: A case study in employment. Mult Scler. 2020:1352458520910499.



APPENDICES

NEDERLANDSE SAMENVATTING

Invaliditeit bij multiple sclerosis
"*Verbeteren van de klinische beoordeling*"

Achtergrond

Multiple Sclerose (MS) is een ernstige neurologische ziekte waarbij zowel ontstekingsprocessen (i.e. neuro-inflammatie) als progressieve zenuwschade (i.e. neuro-degeneratie) een rol spelen, en een gevolg is van een interactie tussen verschillende oorzakelijke (ofwel etiologische) factoren in de omgeving en genetica. Het is een relatief zeldzame ziekte die zich meestal openbaart bij jonge mensen, wat meer bij vrouwen dan bij mannen. Het klinisch beeld is heterogeen en gekarakteriseerd door het optreden van neurologische uitvalsverschijnselen (zoals verlammingsverschijnselen en blindheid van een oog) die ontstaan in dagen tot enkele weken, om vervolgens na enige tijd weer af te nemen, al dan niet met permanente restverschijnselen. Een periode met neurologische symptomen wordt een *schub* genoemd. Er is een grote variatie in wanneer en hoe vaak een schub optreedt, en hoe ernstig deze is. Een aanzienlijk deel van de patiënten ontwikkeld in de loop der jaren een meer geleidelijk verslechterend klinisch beeld met toenemende uitvalsverschijnselen, al dan niet in combinatie met optreden van schubs.

In dit klinisch beeld ligt ook een voorname uitdaging in de dagelijkse klinische praktijk en het wetenschappelijk onderzoek. Het maakt het stellen van de diagnose en opvolgen van het klinisch verloop niet rechttoe rechtaan, en vergt een zekere mate van opleiding en expertise. Omdat de behandelmogelijkheden de laatste drie decennia enorme sprongen vooruit hebben gemaakt, zijn juist deze facetten belangrijker geworden. Wanneer kan de diagnose met voldoende zekerheid gesteld worden om zo vroeg mogelijk te starten met behandeling? Is de ziekte voldoende onder controle, of moet er gestart worden met een meer effectieve (maar vaak ook risicovollere) behandeling? Alhoewel de aanvullende technieken, zoals *Magnetic Resonance Imaging* (MRI) en het meten van eiwitten in het bloed (genaamd biomarkers), zeer waardevol (kunnen) zijn, is hun plaats in de dagelijkse klinische praktijk soms nog onzeker, kostbaarder en tijdrovender dan klinische beoordeling van patiënten.

Doel van dit proefschrift

In dit proefschrift worden enkele studies gepresenteerd met als doel de klinische beoordeling van MS patiënten te verbeteren. **Hoofdstuk 2** in **Deel I** creëert daarvoor eerst het raamwerk door een overzicht te presenteren van klinische en paraklinische uitkomstmaten.

Resultaten

Deel II presenteert vier studies die betrekking hebben op het beoordelen van arm-handfunctie (AHF) en mobiliteit. Van oudsher lag de focus van het beoordelen van invaliditeit op loopstoornissen. Gelukkig wordt nu erkent dat ook andere functionele domeinen, zoals AHF en cognitie, vaak zijn aangedaan en wordt het belang ingezien van medebeoordeling hiervan om een volledig beeld te krijgen van de impact van MS op patiënten. Het is echter niet duidelijk wanneer stoornissen in deze domeinen zich voordoen, en hoe hun interactie is. Om hier meer duidelijkheid in te krijgen worden in **hoofdstuk 3** verschillende aspecten van AHF onderzocht bij 247 patiënten met verschillende mate van loopfunctiestoornissen. Bij meer dan 80% werd een zekere mate van AHF stoornis gevonden, zelfs als de loopfunctie nog normaal was. Alle maten voor AHF waren meer gestoord indien de loopfunctie ook meer gestoord was. Opmerkelijk was dat de meeste mensen nog slechts een geringe AHF stoornis hadden, ook bij diegenen met een ernstige loopstoornis. Samengevat laten deze bevindingen zien dat de beoordeling van invaliditeit van MS patiënten kan worden verbeterd wanneer AHF en loopfunctie onafhankelijk van elkaar worden beoordeeld.

Vervolgens wordt in **hoofdstuk 4** onderzocht wat de relatieve waarde is van verschillende klinische maten voor AHF en mobiliteit ten opzichte van elkaar. In het statistisch model kon het grootste deel van AHF en mobiliteit beoordeeld worden met deze klinische maten. De maten die algemene dagelijkse levensverrichtingen (ADL) weerspiegelen, zoals drinken uit een beker of opstaan uit een stoel, droegen het meeste bij in deze modellen. Het beoordelen van ADL functies kan dus het meten van invaliditeit verbeteren. De ADL maten kunnen tevens snel en gemakkelijk worden afgenomen.

De maten die in **hoofdstuk 3 en 4** zijn gebruikt meten de fysieke prestatie van bewegingen. Het geeft echter geen inzicht in het perspectief van de patiënt op de onderzochte functie. Een *patient-reported outcome measure* (PROM) kan hier uitkomst bieden. Dit zijn vragenlijsten over een bepaald functioneel domein, klacht of andere specifiek facet. In **hoofdstuk 5** werd een waarde voor minimaal klinisch relevant verschil bepaald van de *Arm Function in Multiple Sclerosis Questionnaire* (AMSQ). Dit is een vragenlijst voor een verscheidenheid aan ADL taken voor arm-handfunctie.

De bevindingen die in **hoofdstuk 3 tot en met 5** werden gedaan, zijn gebruikt in de studie in **hoofdstuk 6** waarin de effectiviteit van fampridine getest is in een cohort MS patiënten uit de dagelijkse klinische praktijk. Fampridine is een middel dat gebruikt wordt om de loopfunctie te verbeteren van MS patiënten, maar heeft ook gunstige effecten op andere motorische functies. Onderzoek in een cohort zoals deze heeft een aantal voordelen. Ten eerste geeft het inzicht in het verschil tussen de beoordeling door de clinicus van het effect op motorisch functioneren enerzijds, en de formele vergoedingsregels om

effectiviteit te beoordelen anderzijds. Ten tweede laat subgroep analyse zien welke patiënten het meeste voordeel laten zien. Tenslotte geef het inzicht in het behandelkeuze proces in de klinische praktijk.

De overgrote meerderheid continueerde de behandeling. Het grootste effect op loopsnelheid was te zien bij patiënten met een relatief slechte loopfunctie. Bij patiënten met een relatief gespaarde loopfunctie werd het grootste effect gevonden op een PROM voor loopfunctie, en continueerde ook hier de meeste patiënten. Bij een subgroep van patiënten met tevens een gestoorde arm-handfunctie, werd bij 40% een klinisch relevante verbetering gevonden met de AMSQ. In deze groep werd ook vaker een verbetering op de PROM voor loopfunctie gezien, en werd de behandeling vaker gecontinueerd. Dit kan betekenen dat een mogelijk placebo effect invloed had op het besluit van de clinicus om door te gaan met de behandeling. Ook wordt hiermee het verschil geïllustreerd tussen wetenschappelijk onderzoek en klinische praktijk, en de uitdagingen waarvoor clinici komen te staan. Deze bevindingen hebben bijgedragen aan de aanpassing van de vergoedingsregels van fampridine in Nederland.

Deel III presenteert drie studies waarin de toegevoegde waarde van video-ondersteunde beoordeling van motorisch functioneren is onderzocht. Het gebruik hiervan kan de klinische evaluatie van patiënten op verschillende manieren verbeteren. Ten eerste kunnen bepaalde (psychomotore) eigenschappen van uitkomstmaten worden verbeterd, zoals de inter- en intra-beoordelaar variabiliteit. Ten tweede kunnen de digitale data gebruikt worden voor het maken van *Machine Learning Algorithms* (MLA). Dit kan bijdragen aan de ontwikkelingen van technieken voor bijvoorbeeld het automatisch beoordeling van bewegingen. Tenslotte, kan het motorisch functioneren van patiënten op een voor de patiënt geschikt locatie en tijd worden beoordeeld, en naar de behandelaar of onderzoeker worden verzonden. Zeker in tijden van een wereldwijde crisis zoals de COVID-19 pandemie, kan dit soort technieken waardevol zijn.

Hoofdstuk 7 toont aan dat de variabiliteit van beoordeling van motorische functies verbetert als er gebruik wordt gemaakt van referentie video's. Dit zijn voorbeeld video's van patiënten die een beweging uitvoeren, waaraan een beoordeling volgens het scoringssysteem is gegeven (bijvoorbeeld normaal, geringe stoornis, matig ernstige stoornis, ernstige stoornis, niet mogelijk).

De detectie van verandering in AHF en mobiliteit kan verbeterd worden met video-ondersteunde beoordeling, als meerdere bewegingen (inclusief ADL taken) worden beoordeeld en gecombineerd. **Hoofdstuk 8** laat dit zien in een groep patiënten die behandeld zijn met fampridine. Er werden enkele aanvullende patiënten gevonden met een klinisch relevante verbetering, die geen verbetering toonden op twee standaard maten voor AHF en mobiliteit. Deze studie laat het belang zien van het gebruik van meerdere maten om

ziektebeloop en behandeleffect te meten.

De data van de video-ondersteunde beoordeling van motorisch functioneren kunnen voor verschillende doelen gebruikt worden. Echter zal dan moeten worden samengewerkt met IT specialisten en niet-medisch personeel, wat een potentieel risico is voor behoud van dataprivacy. Het gebruik van auto-encoders kan hiervoor een oplossing zijn. Auto-encoders zijn een digitale techniek die visuele informatie kunnen comprimeren tot informatie die wel gebruikt kan worden voor ontwikkeling van MLA, maar niet te interpreteren is door het menselijk brein. Vervolgens kan de data weer gedecomprimeerd worden tot een begrijpelijk beeld. **Hoofdstuk 9** laat zien dat de gedecomprimeerde data klinisch relevante videobeelden bevat, en daarmee het gebruik van auto-encoders een potentieel waardevolle methode is om dataprivacy te waarborgen.

Toekomstperspectieven

In **deel IV** worden bovenstaande studies uitgebreid bediscussieerd en toekomstperspectieven geschetst. Wat bepaald de geschiktheid van een klinische uitkomstmaat voor de klinische praktijk en wetenschappelijk onderzoek in MS? Het maken van de juiste keus is uiterst belangrijk voor het maken van klinische beslissingen en bepalen van effectiviteit in onderzoek naar nieuwe medicatie. Deze keuze heeft dus invloed op de prognose van een patiënt en welk medicijn uiteindelijk wordt toegelaten op de markt.

Een voorstel wordt gedaan voor een minimale set van klinische uitkomstmaten, waarmee de meest relevante functionele neurologische gebieden bij MS worden beoordeeld, namelijk: loopfunctie, AHF, cognitief functioneren en de visus. Medebeoordeling van het perspectief van de MS patiënten met betrekking tot de invloed die de ziekte heeft op deze gebieden is relevant in aanvulling op prestatiegerichte uitkomstmaten (zoals zo snel mogelijk een afstand van *25 foot* afleggen). Deze meetmethode kan in combinatie met gangbare (bijvoorbeeld MRI) en toekomstige (bijvoorbeeld biomarkers in bloed) technieken gedaan worden. De beschikbaarheid en het juist gebruiken van klinische en paraklinische maten wordt steeds belangrijker, omdat het behandelspectrum van MS patiënten groter wordt en de lat van de behandeldoelen steeds hoger wordt gelegd.

Om een holistisch beeld van een patiënt te krijgen, zou er een paradigmaverandering moeten komen van een systeem waarin niet de dokter, maar de patiënt centraal staat. In zo een systeem participeert de patiënt zelf actief aan het proces door gebruik te maken van vragenlijsten (PROM), moderne technieken zoals biosensor technieken en andere *on-demand* data levering aan het team. Hiermee kan naast de informatie uit reguliere ziekenhuisbezoeken de behandeling gericht worden op de specifieke behoeften van de patiënt.

Biosensor technieken kunnen het meten van MS patiënten verbeteren, en hebben de potentie om MLA en *artificial intelligence* (AI) te ontwikkelen. Er worden verschillende innovatieve biosensor technieken besproken, zoals accelerometers met of zonder gyroscopen, *wireless radio reflection signals* en video-opnames van bewegingen. De komende decennia zullen deze technieken in toenemende mate de klinische praktijk en wetenschappelijk onderzoek doordringen.

Al deze klinisch en paraklinische data van MS patiënten uit de dagelijkse praktijk en het wetenschappelijk onderzoek worden primair gebruikt voor patiëntenzorg en het beantwoorden van wetenschappelijke vragen. Echter zouden deze welverdiende data ook voor andere doelen gebruikt kunnen worden als het wordt samengevoegd in grote multidisciplinaire data-infrastructuren. Een belangrijk argument voor de ontwikkeling van deze data-infrastructuren is de complexiteit van MS waarbij geen universele maat bestaat die de ziekte volledig in beeld brengt. Er zijn reeds verschillende initiatieven die dit doel nastreven, zoals het Nederlands MS kwaliteitsregister.

De klinische beoordeling van MS patiënten is een boeiend veld dat gebaseerd is op historische grondslagen en gegroeid door hedendaagse technologische innovatie. Verdere verbetering en integratie in multidisciplinaire data-infrastructuren zal uiteindelijk leiden tot een beter begrip van en controle over MS, wat uiteindelijk de kwaliteit van leven van onze patiënten verbetert.

AUTHOR AFFILIATIONS

J. Boisvert, Novartis Pharma AG, Switzerland

J. Burggraaff, Department of Neurology, Amsterdam Neuroscience, Amsterdam MS Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands

F. Dahlke, Novartis Pharma AG, Basel, Switzerland

M. Diederich, Department of Neurology, departments of Medicine, Biomedicine and Clinical Research and Research Center for Clinical Neuroimmunology and Neuroscience, University Hospital Basel and University of Basel, Basel, Switzerland

J. Dorn, Novartis Pharma AG, Basel, Switzerland

M. D'Souza, Department of Neurology, departments of Medicine, Biomedicine and Clinical Research and Research Center for Clinical Neuroimmunology and Neuroscience, University Hospital Basel and University of Basel, Basel, Switzerland. CMIO office, Dept. D&ICT, University Hospital Basel, Basel, Switzerland

M. Johnson, Microsoft Research, Cambridge, United Kingdom

N.F. Kalkers, Department of Neurology, OLVG, Amsterdam, The Netherlands

C.P. Kamm, Department of Neurology and Neurorehabilitation, Neurology and Neurorehabilitation Centre, Lucerne Cantonal Hospital, Lucerne, Switzerland. Department of Neurology, University Hospital Bern and University of Bern, Bern, Switzerland

L. Kappos, Department of Neurology, departments of Medicine, Biomedicine and Clinical Research and Research Center for Clinical Neuroimmunology and Neuroscience, University Hospital Basel and University of Basel, Basel, Switzerland

L. Kaya, Department of Neurology, Amsterdam Neuroscience, Amsterdam MS Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands

J. Killestein, Department of Neurology, Amsterdam Neuroscience, Amsterdam MS Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands

K. Kravalis, Department of Neurology, departments of Medicine, Biomedicine and Clinical Research and Research Center for Clinical Neuroimmunology and Neuroscience, University Hospital Basel and University of Basel, Basel, Switzerland

K. Lam, Department of Neurology, Amsterdam Neuroscience, Amsterdam MS Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands

C. Morrison, Microsoft Research, Cambridge, United Kingdom

C.E.P. van Munster, Department of Neurology, Amphia, Breda, the Netherlands

M. Obura, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, The Netherlands

A. Sellen, Microsoft Research, Cambridge, United Kingdom

S. Steinheimer, Department of Neurology, University Hospital Bern and University of Bern, Bern, Switzerland

B.M.J. Uitdehaag, Department of Neurology, Amsterdam Neuroscience, Amsterdam MS Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands

L. Walsh, Novartis Pharma AG, Basel, Switzerland

Y. Zaykov, Microsoft Research, Cambridge, United Kingdom

LIST OF PUBLICATIONS

First author

Caspar E.P. van Munster, Jessica Burggraaf, Saskia Steinheimer, Christian P. Kamm, Marcus d'Souza, Jonas Dorn, Jacques Boisvert, Manuela Diederich, Kristina Kravalis, Frank Dahlke, Ludwig Kappos, Bernard M.J. Uitdehaag. Measuring various aspects of upper extremity function independent from ambulation in Multiple Sclerosis enhances disability assessment. *Submitted for publication*

Caspar E.P. van Munster, Ka-Hoo Lam, Marcus D'Souza, Saskia Steinheimer, Christian Kamm, Jessica Burggraaff, Matthew Johnson, Yordan Zaykov, Jonas Dorn, Frank Dahlke, Ludwig Kappos, Joep Killestein, Bernard M.J. Uitdehaag. Video-assisted rating improves the detection of change in motor function in multiple sclerosis. *Submitted for publication*

Caspar E.P. van Munster, Marcus d'Souza, J. Dorn, A. Dorier, C.P. Kamm, S. Steinheimer, F. Dahlke, B.M.J. Uitdehaag, L. Kappos, M. Johnson. Autoencoder: a new method for keeping data privacy when analysing videos of patients with motor dysfunction – a proof of concept study. *Journal of Medical Internet Research. 2020; Mar 19*

Caspar E.P. van Munster, Levent Kaya, Ka-Hoo Lam, Nynke Kalkers, Joep Killestein, Bernard M.J. Uitdehaag. Responder rates to fampridine differ between clinical subgroups of MS patients and patient reported outcome influences treatment decision making. *Multiple Sclerosis and Related Disorders. 2019; Nov 5*

Caspar E.P. van Munster, Levent Kaya, Morgan Obura, Nynke Kalkers, Bernard Uitdehaag. Minimal clinically important difference of improvement on the Arm Function in Multiple Sclerosis Questionnaire (AMSQ). *Multiple Sclerosis Journal.* 2019; Jan 18

Caspar E.P. van Munster, Marcus d'Souza, Saskia Steinheimer, Christian P. Kamm, Jessica Burggraaf, Jonas Dorn, Jacques Boisvert, Manuela Diederich, Kristina Kravalis, Frank Dahlke, Ludwig Kappos, Bernard M.J. Uitdehaag. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. *Multiple Sclerosis Journal.* 2018; Aug 31

Caspar E.P. van Munster, Wouter J.C. van Ballegoij, Jutta M. Schroeder – Tanka, Renske M. van den Berg – Vos. Een ernstig herseninfarct na carotismassage. *Nederlands Tijdschrift voor Geneeskunde*. 2017; 161(0): D826

Caspar E.P. van Munster, Bernard M.L. Uitdehaag. Outcome measures for treatment efficacy in clinical trials for multiple sclerosis. *CNS drugs*. 2017 Mar; 31(3): 217-236

Caspar E.P. van Munster, Jan van Gijn, Henry C. Weinstein. Response letter to "The plantar reflex – A study of observer agreement, sensitivity, and observer bias" by Araújo et al, Neurol Clin Pract 2015; 5:309-316. *Neurology Clinical Practice*. 2015 Nov 12

Caspar E.P. van Munster, L.E. Jonkman, Henry C. Weinstein, Bernard M.J. Uitdehaag, Jeroen J.G. Geurts. Gray matter damage in multiple sclerosis: impact on clinical symptoms. *Neuroscience.* 2015 Jul; 303(8): 446-461

Caspar E. P. van Munster, René van den Berg, Henry C. Weinstein. A Dural Fistula as a treatable cause of cognitive impairment. *Neurohospitalist*. 2014 Apr; 4(2): 111-2.

Caspar E.P. van Munster, Jan van Gijn, Bernard M.J. Uitdehaag, Henry C. Weinstein. The additional diagnostic value of stroking different sites of the foot to provoke an up going toe of the plantar reflex and the influence of experience. *Journal of Neurology*. 2012 Nov; 259(11): 2424-8

Caspar E.P. van Munster, Mikael von und zu Fraunberg, Gabriel J.E. Rinkel, Jaakko Rinne, Timo Koivisto, and Antti Ronkainen. Differences in aneurysm and patient characteristics between cohorts of Finnish and Dutch patients with subarachnoid hemorrhage: time trends between 1986 to 2005. *Stroke*. 2009 Dec; 39(12): 3166-71

Second author

Liesbet M. Peeters, Caspar E.P. van Munster, Bart Van Wijmeersch, Robin Bruyndonckx, Ilse Lamers, Niels Hellings, Veronica Popescu, Christoph Thalheim, Peter Feys. Multidisciplinary data infrastructures in multiple sclerosis: why they are needed & can be done! *Multiple Sclerosis Journal*. 2018; Nov 1

Senior author

Isabelle Schilders, Caspar E.P. van Munster. Snel progressieve dementie: ABRAcadabra! Accepted for publication in *Nederlands Tijdschrift voor Neurologie en Neurochirurgie*. 2020

Floris de Kleermaeker, Bernard M.J. Uitdehaag, Bob W. van Oosten, Caspar E.P. van Munster. Paroxysmale hypothermie bij multiple sclerose. *Nederlands Tijdschrift voor Geneeskunde.* 2019; Jan 14: 163

Jasper Kliffen, Marc Engelen, Giorgos Karas, Caspar E.P. van Munster. Neuroradiozoölogie. *Nederlands Tijdschrift voor Neurologie en Neurochirurgie*. 2015 Dec; 116(4): 224-229

Co-author

M. Peters, R. Michels, K. Schiffers, P. Bouma, G. Hengstman, R. Hupperts, Caspar E.P. van Munster, E. Zeinstra, G. De Voer, M. Krol. A Delphi panel on treatment of high disease activity relapsing remitting multiple sclerosis in the Netherlands. Accepted for publication in *Journal of Comparative Effectiveness Research. 2020*

Marcus D`Souza, Saskia Steinheimer, Jonas Dorn, C. Morrison, Jacques Boisvert, Kristina Kravalis, Jessica Burggraaff, Caspar E.P. van Munster, Manuela Diederich, A. Sellen, Christian P. Kamm, Frank Dahlke, Bernard M.J. Uitdehaag, Ludwig Kappos. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. *Multiple Sclerosis Journal: Experimental, Translational and Clinical.* 2018; Aug 9; 4(3)

Iris Kleerekooper, Zoé van Kempen, Cyra Leurs, Iris Dekker, Theo Rispens, Birgit Witte, Caspar E.P. van Munster, Brigit Alette de Jong, Bob W van Oosten, Bernard Uitdehaag, Mike Peter Wattjes, Joep Killestein. Disease activity following pregnancy-related discontinuation of natalizumab in multiple sclerosis. *Neuroimmunology & Neuroinflammation*. 2017; Dec 5; 5(1): e424

Mark Willis, Owen Pearson, Zsolt Illes, Tobias Sejbaek, Christian Nielsen, Martin Duddy, Kate Petheram, Caspar E.P. van Munster, Joep Killestein, Clas Malmestrom, Emma Tallantyre, Neil Robertson. Alemtuzumab following fingolimod for multiple sclerosis. *Neurology, Neuroimmunology and Neuroinflammation.* 2017 Jan; 10; 4(2): e320

DANKWOORD

Begin met het einde voor ogen en doe de belangrijke dingen eerst. Deze eigenschappen hebben mij geholpen bij het maken van dit boekje. Wat dit betreft zou ik het dankwoord eigenlijk al vóór het eerste hoofdstuk moeten plaatsen. Dankzij velen heb ik mijn focus op het einddoel en de echt relevante aspecten kunnen vasthouden. Dat was mij alleen nooit gelukt. In dit belangrijke laatste deel van mijn boekje wil ik deze personen bedanken en benoemen wat zij voor mij hebben betekend.

First of all, prof. dr. B.M.J. Uitdehaag. Beste Bernard, jij staat aan de basis van mijn neurologische loopbaan en loopt er als een rode draad doorheen. We spraken elkaar voor het eerst samen met Chris Polman in 2009 op een door mijn vader gearrangeerde bijeenkomst. Daarna heb je meerdere keren een belangrijke rol gespeeld op voor mij cruciale momenten. Je hielp mij kiezen tussen een promotieplek of starten met de opleiding, faciliteerde een wetenschappelijke stage, en als belangrijkste creëerde je voor mij een functie als fellow in het MS centrum. Ik wilde graag promoveren, maar dit raadde je mij af. Je legde uit dat een jonge neuroloog die eerst 4 jaar uit kliniek zou treden voor onderzoek, waarschijnlijk uiteindelijk een minder grote kans heeft op een mooie klinische neurologenplek. Je hielp mij daarmee het einde voor ogen te zien. Desalniettemin heb je volgens mij altijd in de gaten gehouden of het niet toch op een promotie zou kunnen uitkomen. Je creëerde kansen en gaf mij keuzes die achteraf gezien de broodkruimels waren die leidden naar dit boekje. Deze stijl, de rust die je kan geven als het even niet lukt (er is altijd een oplossing) en de onvoorwaardelijke vriendelijkheid bewonder ik enorm en zijn voor mij een voorbeeld. Bernard, ik wil je bedanken voor wat je voor mij betekent hebt de afgelopen 10 jaar.

Prof. dr. J. Killestein, beste Joep, alleen jij had mijn copromotor kunnen zijn. Zonder dat je het misschien wist ben jij doorslaggevend geweest om het onderzoek door te zetten tot een promotie. Op een gewone donderdag in de bunker, die zich polikliniek laat noemen, vroeg je mij of ik al dat onderzoek niet zou moeten bundelen. Ik vertelde je dat ik me er eigenlijk al bij had neergelegd dat dit niet meer zou lukken. In plaats dit gewoon als persoonlijke tegenslag voor mij te laten, baalde je zelf juist dat je dit niet had voorzien. Op één op andere manier bleef dit in mijn hoofd zitten, en hielp het mij het doel weer te zien. De afgelopen jaren zijn er meer van dit soort momenten geweest waarbij je op een totaal onverwachte, originele manier reageerde en mij op weg hielp. Ik wil je hiervoor bedanken, en voor al die keren dat ik even bij je binnen kon hobbelen met een vraag of één of andere flauwe grap.

Prof. dr. H.C. Weinstein, beste Henry, als opleider heb je natuurlijk veel voor mij betekend. Je hebt voor een groot deel bijgedragen aan hoe ik nu ben als neuroloog, maar het is meer dan dat alleen. Sinds mijn komst in Amsterdam had ik altijd het gevoel dat je achter me stond. Je leerde mij dingen van een

andere kant te zien en wat echt belangrijk was voor mijn opleiding. Ook meer recent stond je voor me klaar en hielp je me weer op weg. Wat je me leerde heeft mij als persoon verbeterd, zowel op professioneel als persoonlijk vlak. Ooit wilden we samen onderzoek doen naar het neurologisch onderzoek. Alhoewel dat niet helemaal gelukt is, ligt er hier nu wel een boekje over klinisch onderzoek en zit je in de leescommissie. De cirkel is rond. Dank je voor alles.

De andere leden van de leescommissie, prof. dr. V. de Groot, prof. dr. L.H. Visser, Dr. N.F. Kalkers en Dr. M.M. Schoonheim, wil ik hartelijk danken het zitting nemen in de leescommissie en de kritische beoordeling van mijn proefschrift. Beste Nynke, jij ook veel dank voor de fijne samenwerking als co-auteur en mede-clinicus.

Uiteraard wil ik ook de patiënten bedanken die hebben deelgenomen in de Assess MS studie en de Fampyra-poli. Dankzij jullie bereidheid langer in het ziekenhuis te blijven en gekke bewegingen voor een camera te doen, heb ik dit boekje kunnen schrijven. We zijn samen een stapje verder gekomen in het doorgronden van de ziekte MS.

The Assess MS project gave me the opportunity to do scientific research in an international multi-center study. I would like to thank the Assess MS team for this fantastic experience. The clinicians: Christian Kamm, Kristina Kravalis, Ludwig Kappos, Manuela Diederich, Marcus D'Souza and Saskia Steinheimer. The Novartis team: Frank Dahlke, Jacques Boisvert, Jonas Dorn and Lorcan Walsh. The Microsoft team: Abigail Sellen, Cecily Morrison, Matthew Johnson and Yordan Zaykov. Jessica Burggraaff, jou wil ik hier speciaal noemen. Het is ons allebei gelukt! Ik weet nog goed toen we allebei aan de wetenschap begonnen te ploeteren in 2013. We hebben elkaar denk ik goed bij onze frustraties kunnen helpen door er vooral ook de grap van in te zien. Dank je wel daarvoor.

Beste Brigit de Jong en Bob van Oosten, uiteraard wil ik ook jullie bedanken. Hoewel jullie niet direct hebben bijgedragen aan dit boekje, hebben jullie mij geleerd "hoe MS werkt", voor zover we dat natuurlijk werkelijk weten. Bob, ik vond het fantastisch dat ik altijd even bij je langs kon komen voor een ingewikkeld neurologisch probleem. Bizar hoe in een paar minuten iets ingewikkelds opeens overzichtelijk en hanteerbaar kan worden. Daar pluk ik nog de vruchten van. Brigit, dank voor alle hulp die je me gegeven hebt. Hiermee bedoel ik niet alleen op het gebied van MS, maar ook de dingen van alle dag. Ik bewonder jouw balans tussen werk en privé; dat is voor mij een voorbeeld.

Ook wil ik de onderzoekers van het MS Centrum Amsterdam bedanken voor de fijne samenwerking: Cyra Leurs, Danko Coric, Djoeke Doesburg, Floor Loonstra, Ilse Nauta, Iris Dekker, Jenny Nij Bijvank, Lisanne Balk, Marloes

Hagens en Martijn Wijburg. Dank voor jullie enorme inzet als Fampyra-ridders. Ka-Hoo Lam, jou wil ik in het bijzonder danken voor de fijne samenwerking als co-auteur. Ik bewonder je integere houding en doorzettingsvermogen, en kijk uit naar de resultaten van jouw onderzoek. Levent Kaya en Morgan Obura, jullie ook veel dank voor de soepele samenwerking als co-auteur.

Dan mijn kamergenoten Luuk van Rooij en Floris de Kleermaeker. Jullie zijn fantastische kerels en ik heb enorm geluk gehad om met jullie samen te werken. Vooral ook heel erg bedankt voor het weglachen van alle frustraties en ellende die we tegen kwamen. Ik vind het mooi dat we elkaar blijven tegenkomen in de MS wereld. En dan.... Zoé van Kempen, ouwe dikzak, wat een geschenk dat we na onze opleiding nog 2,5 jaar roomies konden zijn! We hebben elkaar beter leren kennen en hebben elkaar verder kunnen helpen. Maar ik wil je natuurlijk vooral bedanken voor de mafkees die je bent om wie ik zo ontzettend veel heb moeten gelachen.

De dames van het secretariaat neurologie van het VUmc mogen niet ontbreken in mijn dankwoord. Beste Brenda, Claudia, Elles, Karin en Lianne, dank jullie voor de ondersteuning en gezellige praatjes die we hadden. Het voelde altijd een beetje als thuiskomen bij jullie.

Verder wil ik Gianina Eissens en Danielle Savelkoel van het clinical trial bureau bedanken voor de robuuste ondersteuning en vriendelijkheid die jullie gegeven hebben. Ook Jana, Laura, Loek en Sara van de data unit wil ik danken voor het harde werk aan de eentjes en nulletjes.

Alle neurologen van VUmc wil ik bedanken voor de fijne samenwerking in de kliniek en polikliniek. Dank voor de back-up die ik soms nodig had als beginnend neuroloog.

De periode voor mijn fellowship tijdens mijn opleiding heeft in zekere zin ook bijgedragen aan dit boekje. Ik wil alle neurologen van het SLAZ hiervoor bedanken. Ik heb van jullie geleerd goed klinische te redenen. Dat is de basis van neurologie. Hiermee is een klinisch proefschrift als de mijne beter geworden. Minstens zo belangrijk uit deze periode zijn de arts-assistenten. Dank jullie allemaal voor het team dat we waren en vooral voor de lol die we hebben gehad. Juerd Wijntjes, jou in het bijzonder wil ik bedanken voor de vriendschap die in deze periode is ontstaan. Dank hiervoor, en dank je dat je met Marieke dicht bij ons bent komen wonen in de Provence.

Ook wil ik mijn collega neurologen, assistenten, verpleegkundigen en de medewerkers van onze afdelingen van Amphia bedanken voor de fijne werkplek die ik nu heb.

Ik vraag me altijd af waarom een jaarclub in een dankwoord van een proefschrift wordt genoemd. Die hebben daar toch niets aan bijgedragen?

Echter begon het me bij het schrijven van dit boekje te dagen: zonder goed ontspannen kan je niet hard werken. Voor dat eerste betekenen jullie veel voor me en brengen jullie de zaak in balans.

Bij de genoemde balans hoort ook mijn familie, de van Munsters en de Wijnens. Dank jullie voor onze sterke familieband waar ik steun en lol kan vinden. Nina, jou zou ik hier in het bijzonder willen bedanken voor je kritische blik op de inhoud en het Engels van de introductie en discussie van dit boekje. Het is er veel beter van geworden.

Tim van Veen, je bent mijn oudste vriend, en alleen daarom al wil ik je te bedanken. Onze vriendschap is goud. Hoewel we elkaar minder zien dan 25 jaar geleden, is het mooie van dit soort vriendschappen dat je bij elke ontmoeting gewoon weer verder gaat met het gesprek. Daarmee ben je een rode draad door mijn leven die vast ook ergens door dit boekje heen getrokken is. Dank je voor je vriendschap.

Wendy Koster, mijn lieve paranimf. Het cirkeltje is weer rond. Ik dank je voor de oude vriendschap die wij hebben, al bijna 20 jaar. Sinds onze avonturen bij Sensation Black in 2001 ben jij mijn studiemaatje. Niet om te studeren, maar om te borrelen en te ouwehoeren. Ik was in een andere tijd jouw paranimf, jij nu de mijne. Veel logischer dan dat had het niet kunnen zijn.

Lieve Eduard, Valérie, Rik en Freek. Dank jullie voor de goede zorgen voor mijn zussen en broer. Jullie liefde voor hun stelt mij als broer gerust.

Lieve Daphne, Harmen, Edith, Iris en Dominique. Ik ben trots op jullie en het unieke gezin dat wij samen vormen. Met elk van jullie heb ik een speciale band. We hebben geleerd onvoorwaardelijk er voor elkaar te zijn, niet alleen in moeilijke tijden, maar vooral ook om er een feestje van te maken. Lieve Daph, fantastisch dat je mijn paranimf bent. Met mijn grote zus naast me moe(s)t het wel goed komen. Dank je wel daarvoor.

Het is gelukt pap! Dat heb ik niet in de laatste plaats aan jou te danken. Jij hebt altijd gelooft dat dit eens zou gaan lukken, terwijl ik de moed meerdere keren had opgegeven. Hierbij heb je me steeds weer gesteund en gezegd dat ik moest volhouden. Dank je wel daarvoor. De laatste jaren ben ik je steeds beter gaan begrijpen, misschien ook omdat onze wegen steeds meer op elkaar zijn gaan lijken. Daarin ben je een voorbeeld voor mij hoe ik als neuroloog, maar ook als mens wil zijn. Dank voor al je liefde en steun.

Eigenlijk zou ik je in één adem willen noemen met papa, maar je weet hoe graag hij het zo wilt hebben ;-) Lieve mam, door jou ben ik geworden wie ik ben. De fijne tijd thuis met al mijn zussen en broer is de basis hiervan. Jij hebt mij de belangrijkste kernwaarden geleerd. Dit heeft mij geholpen bij het maken van dit boekje, en helpt mij elke dag. Ik wil je bedanken voor alles.

Lieve Pépé en Mémé, ik geloof niet in sprookjes, maar ben ervan overtuigd dat jullie ergens mee kijken naar wat we hier uitspoken. Deze gedachte geeft mij een gerust gevoel. Pépé, dit boekje draag ik op aan u. Ik heb veel aan u gedacht toen ik ermee bezig was. Hoe zou Pépé dit doen? Wat zij hij ervan vinden? Kan ik dit niet doen? Mijn herinneringen aan u hebben mij geholpen dit boekje te schrijven. U had het beremooi gevonden om erbij te zijn. We zien elkaar ooit weer.

Lieve Inca, Nola en Aura. Mijn dametjes die van elke dag een feest maken. Nu dit boekje af is ga ik niet opeens meer tijd aan jullie besteden. Dat hoeft namelijk helemaal niet: jullie hebben altijd al mijn hoogste prioriteit. Ik ben enorm trots op jullie. Niet te snel groot worden hè?

Allerliefste liefde van me, lieve Maya. Aan jou heb ik alles te danken, zo simpel is dat. Jij maakt het voor mij mogelijk dat ik kan doen wat ik doe. Je creëert daarmee een balans in mijn leven die ik alleen niet vinden kan. De afgelopen 10 jaar waren de beste tot nu toe, laten we hetzelfde doen voor de rest van ons leven, en daarna. Dank je wel voor alles wat je me geeft. Ik hou van je.

ABOUT THE AUTHOR

Caspar Erik Pieter was born on 20 April 1983 in Nijmegen as the first son of Erik, son of Petrus (i.e. Pépé). He grew up together with his brother Harmen, and four sisters Daphne, Edith, Iris and Dominique. After finishing secondary school (atheneum) at the Sint-Janslyceum in 's-Hertogenbosch in 2001, he started medical school at the University of Utrecht. Already early during his studies he developed a special interest for neurology and scientific research. Caspar completed scientific internships at the department of Neurology in the University Medical Centre Utrecht and the department of Neurosurgery of Kuopio University Hospital in Finland, respectively. Furthermore, he broadened his horizon during internships in Tuele Hospital in Muheza (Tanzania) and the Sint Elisabeth hospital (Curacao). In December 2007 Caspar became a medical doctor, and started his first job a month later on the department of neurology in the University Medical Centre Utrecht. After a year he decided to go backpacking for some time in Southeast Asia. Back in the Netherlands, he started his second job at the department of neurology in Onze Lieve Vrouwe Gasthuis (formerly known as Sint Lucas Andreas hospital) in Amsterdam, which led to his residency of neurology in 2009. During this period, he won two awards and participated as board member in several committees. In 2013, his interest in multiple sclerosis (MS) developed during a research traineeship at the department of Anatomy and Neurosciences in the VU Medical Center in Amsterdam under supervision of prof.dr. J.J.G. Geurts. A year later, Caspar decided again to go on an adventure abroad, that time on his classic Honda SS50 moped. After this bumpy ride he became a neurologist in 2015. To deepen his knowledge regarding MS, he started working as a clinical and research fellow at the VU Medical Center. Under supervision of prof.dr. B.M.J. Uitdehaag he participated in the international multi-center study *Assess MS*, which contributed to several chapters of this thesis. His clinical work contained setting up a specialized outpatient clinic for treating MS patients with fampridine. Data collected from these patients were analyzed by Caspar, and used in multiple publications included in his thesis. Eventually, prof.dr. J. Killestein became actively involved as co-promotor. Caspar is board member of the Dutch MS task force of the Dutch Association of neurologists, and member of the scientific board of the Dutch national MS registry. He participates in several medical advisory boards of MS. Since 2018 he works as a neurologist – MS specialist in Amphia, and lives in Breda with his wife Maya and three daughters Inca, Nola and Aura.