Year: 2018

Recommended outcome measures for inpatient rehabilitation of multiple sclerosis are not appropriate for the patients with substantially impaired mobility

Barin, Laura ; Vaney, Claude ; Puhan, Milo Alan ; von Wyl, Viktor

Abstract: BACKGROUND In multiple sclerosis (MS) rehabilitation, most currently used outcome measures were validated in patients with a relapsing remitting MS and mild to moderate impairments. We aimed to assess whether these measures were also adequate in more impaired patients, frequently encountered in those with progressive MS (PMS). METHODS Outcome measurements were extracted from medical records of 229 patients with PMS undergoing 3 weeks of routine inpatient rehabilitation between 2011 and 2015. We assessed the acceptability of Nine-Hole Peg Test (9HPT), Timed 25-Foot Walk (T25FW), 2-Minute Walk Test (2MWT), Rivermead Mobility Index (RMI) and the Functional Independence Measure (FIM) by analysing their statistical distributions, concurrent validity by comparing Spearman correlations with pre-specified hypotheses, and responsiveness across impairment status by calculating standardized response means. RESULTS Our concurrent validity hypotheses were mainly satisfied. However, all outcome measures had skewed distributions, showed low variability, and thus were inadequately discriminative. Moreover, 9HPT was never responsive across the impairment states, whereas the T25FW was responsive for mildly impaired patients, and the 2MWT for mild to moderate MS, respectively. Generic multi-items measures such as RMI and FIM-motor were adequately responsive for all severity levels. CONCLUSIONS Currently used outcome measures are inadequate for patients with impaired mobility, and there is a dire need of specifically designed outcome measures for routine care that are less burdensome and short-term responsive.

DOI: https://doi.org/10.1016/j.msard.2018.04.001

Originally published at:
Barin, Laura; Vaney, Claude; Puhan, Milo Alan; von Wyl, Viktor (2018). Recommended outcome measures for inpatient rehabilitation of multiple sclerosis are not appropriate for the patients with substantially impaired mobility. Multiple Sclerosis and Related Disorders, 22:108-114.
DOI: https://doi.org/10.1016/j.msard.2018.04.001

Recommended outcome measures for inpatient rehabilitation of multiple sclerosis are not appropriate for the patients with substantially impaired mobility.

Laura Barin[a], Claude Vaney[b], Milo Alan Puhan[a], Viktor von Wyl[a]

[a]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zürich, Switzerland

[b]Berner Klinik Montana, Impasse Palace Bellevue, 3963 Crans-Montana, Switzerland

Corresponding author:

Laura Barin

Epidemiology, Biostatistics and Prevention Institute

University of Zurich

Hirschengraben 84

8001 Zurich

Switzerland

Tel. +41 44 634 4609

Email: laura.barin@uzh.ch

Other authors' contacts:

Claude Vaney, Email: claude.vaney@gmail.com

Milo Alan Puhan, Email: miloalan.puhan@uzh.ch

Viktor von Wyl, Email: viktor.vonwyl@uzh.ch

Abstract

Background: In multiple sclerosis (MS) rehabilitation, most currently used outcome measures were validated in patients with a relapsing remitting MS and mild to moderate impairments. We aimed to assess whether these measures were also adequate in more impaired patients, frequently encountered in those with progressive MS (PMS).

Methods: Outcome measurements were extracted from medical records of 229 patients with PMS undergoing 3 weeks of routine inpatient rehabilitation between 2011 and 2015. We assessed the acceptability of Nine-Hole Peg Test (9HPT), Timed 25-Foot Walk (T25FW), 2-Minute Walk Test (2MWT), Rivermead Mobility Index (RMI) and the Functional Independence Measure (FIM) by analysing their statistical distributions, concurrent validity by comparing Spearman correlations with pre-specified hypotheses, and responsiveness across impairment status by calculating standardized response means.

Results: Our concurrent validity hypotheses were mainly satisfied. However, all outcome measures had skewed distributions, showed low variability, and thus were inadequately discriminative. Moreover, 9HPT was never responsive across the impairment states, whereas the T25FW was responsive for mildly impaired patients, and the 2MWT for mild to moderate MS, respectively. Generic multi-items measures such as RMI and FIM-motor were adequately responsive for all severity levels.

Conclusions: Currently used outcome measures are inadequate for patients with impaired mobility, and there is a dire need of specifically designed outcome measures for routine care that are less burdensome and short-term responsive.

Key-words: Neurological Rehabilitation, Progressive Multiple Sclerosis, Outcome Assessment (Health Care), performance-based measure, responsiveness, validity

Abbreviations: 5STS = 5 repetition Sit-to-Stand test, CI = confidence interval, FIM = Functional Independence Measure, FIST = Function in Sitting Test, MSEDGE = Multiple Sclerosis Evaluation Database to Guide Effectiveness, MSOAC = Multiple Sclerosis Outcome Assessments Consortium, MSTF = Multiple Sclerosis Task Force (MSTF), OM = Outcome Measure, PMS = Progressive Multiple Sclerosis, RMI = Rivermead Mobility Index, SaGAS10 = Short and Graphic Ability Scale-10, SD = Standard Deviation, SRM = Standardized Response Mean.

1. Introduction

Multiple sclerosis (MS) is a chronic disease of the Central Nervous System, characterized by a variety of manifestations in different functional systems. Due to the limited options for immunomodulatory treatment (Elovaara et al., 2018), rehabilitation plays an essential role in reducing the multi-faceted MS-caused impairments in persons with progressive MS (PMS). For objective and reliable evaluation of rehabilitation interventions, the use of standardized outcome measures (OMs) is crucial (Jette DU, Halbert J, 2009). Generally, a comprehensive core set of OMs should include feasible, acceptable, valid, and responsive OMs. They should also cover all the relevant International Classification of Functioning, Disability and Health (ICF) domains (Stucki, 2016), namely activity and participation (Khan et al., 2007).

In response to the multitude of proposed OMs in MS rehabilitation, the American Physical Therapy Association Neurology Section nominated the MS Task Force (MSTF) to review and assess the psychometric data and clinical utility of the most common OMs in the field. Subsequently, the MSTF issued setting and MS-severity specific recommendations for 63 OMs (henceforth MSEDGE guidelines) (Potter et al., 2014). Moreover, the MS Outcome Assessments Consortium (MSOAC) pursued a similar goal, but with a focus on clinical trials, and aims „to accelerate the development of therapies for MS“ (Larocca et al., 2017).

In routine inpatient rehabilitation settings, patient characteristics tend to differ very substantially from the general MS population in outpatient clinics and clinical studies, mainly due to a much larger proportion of patients with a PMS form. PMS forms are also underrepresented in most rehabilitation outcome validation studies, which were mainly performed in patients with relapsing

remitting MS (Coleman et al., 2012; Fischer et al., 1999; Gijbels et al., 2012; Kaufman et al., 2000; Kragt et al., 2006; Rabadi and Vincent, 2013; Rosti-Otajärvi et al., 2007). The patients with PMS have longer disease duration and higher levels of disability that might affect the feasibility as well as the psychometric properties of the OMs used in rehabilitation. Precisely to take the effect of disability into account, there have also stratified the patients by disability level (Baert et al., 2014). Therefore, the aim of this study was to evaluate the psychometric properties of the most commonly used OMs in routine rehabilitation for patients with PMS in a routine inpatient rehabilitation programme.

2. Methods

2.1 Participants

This study included patients with a PMS form, who underwent an inpatient programme in 2011-2015 at the Neurorehabilitation Center *Berner Klinik Montana.* The use of these data was approved by the Ethics Committee Zurich (BASEC 2017-00077), who also issued a waiver for the retrospective retrieval of informed consent. The *Berner Klinik Montana* is a neurorehabilitation clinic in South-western Switzerland and part of RIMS, the European network for best practice and research in MS Rehabilitation. Its MS inpatient rehabilitation programme cares for approximately 250 patients per year, which corresponds to approximately 35% of all MS neurorehabilitation stays in Switzerland. The majority of patients (70%) have a progressive form of MS, either primary or secondary. This high percentage can be explained by the reimbursement policies of Swiss insurances.

MS inpatient multidisciplinary rehabilitation consists of a 3-week programme with individualized

objectives, consisting of physiotherapy and occupational therapy, complemented with speech therapy, nutrition counselling, psychology, neuropsychology, and hippotherapy as needed. The therapy sessions are at least 30 minutes long, with a total daily duration of at least 2-3 hours, for at least 5 days a week (Khan et al., 2007).

2.2 The measurements

At the *Berner Klinik Montana*, the OMs assessed, both at baseline and at discharge, are: Functional Independence Measure (FIM), Rivermead Mobility Index (RMI), the 2-Minute Walk Test (2MWT), and the Timed 25-Foot Walk test (T25FW) assessed by the physiotherapists, and the Nine-Hole Peg Test (9HPT) measured by the occupational therapists. The Short and Graphic Ability Scale-10 (SaGAS10) is then computed from T25FW and 9HPT (Barin et al., 2016). The neurologists assess the Expanded Disability Status Scale (EDSS) only at discharge.

The measurements are performed within 48 hours from admission and at discharge of the patient. All OMs used cover the ICF domain of activity, while FIM and EDSS additionally assess the body structures and function domain. Moreover, FIM also covers the participation domain. All OMs are performance-based and, with the exception of the patient-reported RMI, assessed by health care professionals. According to MSEDGE guidelines, T25FW and 9HPT are considered highly recommended, while RIM and FIM are recommended, EDSS is not recommended, and 2MWT is classified as "unable to recommend" due to lack of evidence at the time of the review. SaGAS10 was not assessed in MSEDGE guidelines.

2.3 Validation status of the measurements

The advantages, but also shortcomings, of these OMs are well documented for the general MS population, but have remained largely unexplored in the context of PMS. EDSS is the gold standard disability measure in MS (Amato and Portaccio, 2007). However, it has an unequal mean stay-time at each progression stage, it is prone to subjectivity, and is poorly responsive in the short-term (Amato and Portaccio, 2007). The RMI scores 15 mobility tasks with 1 (able to perform) or 0. Despite being frequently used for neurological conditions, it has not been validated for MS patients (Franchignoni et al., 2003).

FIM includes 13 motor and five cognitive items that evaluate the independence level in daily activities (Skinner and Turner-Stokes, 2006). Its good reliability and concurrent validity were demonstrated for MS (Brosseau and Wolfson, 1994), but responsiveness data are limited as illustrated by the lack of a Minimal Clinically Important Difference. FIM has been systematically collected in the Berner Klinik since 2015 and data were available for 67 patients in our population.

The 2MWT was recently suggested as a substitute for the gold standard 6-Minute Walk Test (Baert et al., 2014; Gijbels et al., 2012). The T25FW at highest safe speed is part of the MS Functional Composite (MSFC), and has excellent properties in ambulatory MS patients (Fischer et al., 1999; Kaufman et al., 2000). In the *Berner Klinik Montana* the T25FW test is performed once with a flying start (Gijbels et al., 2012), and non-ambulatory patients are scored with a time value of 180 s (Hoogervorst et al. 2002).

9HPT assesses fine manual dexterity and is also a component of the MSFC (Fischer et al., 1999), whose properties are well-studied (Kragt et al., 2006). The 9HPT is executed once per hand and, in case of failure to complete, 300 s were assigned.(Hoogervorst et al., 2002) For both 9HPT and T25FW the standard procedure would require repeating the tasks twice, but, due to time constraints and more substantial impairments, it is frequently infeasible in stationary rehabilitation settings.

SaGAS10 is a revised version of SaGAS (Vaney et al., 2004) and combines logarithmic

transformations of right- and left-hand 9HPT, as well as T25FW results. It correlates moderately with RMI and EDSS, highly with 2MWT and is moderately responsive (Barin et al., 2016).
For study inclusion, we required the patients to have a PMS form, and the complete assessments of RMI, T25FW, both 9HPTs, and EDSS at discharge. We excluded multiple visits from the same patients, including only the first visit, provided the above-mentioned criteria were fulfilled. We comprehensively checked the data quality (range checks, missing data), by crosschecking suspicious values against patient charts.

2.4 Statistical analysis

We used descriptive statistics to describe the demographic information and baseline assessments. We stratified the results by severity level of MS: mild MS when $0 \leq EDSS \leq 4$, moderate MS when $4.5 \leq EDSS \leq 6.5$, and severe MS when $EDSS \geq 7$ (Gijbels et al., 2010).

2.4.1 Acceptability

To assess the acceptability (Fitzpatrick et al., 1998) of the OMs, we investigated the range, mean, standard deviation (SD), skewness, and ceiling and floor effects, all at baseline. Ideally, an OM's theoretical range should be fully covered by the observed data; while the empirical distribution should be approximately normal (Trochim and Donnelly, 2006), with a skewness index not exceeding the absolute value of 2. Furthermore, the fraction of measurements with the lowest or highest score should be limited because longitudinal changes may not be detected otherwise. We considered ceiling (resp. floor) effects as excellent when the percentage of patients obtaining the best (resp. worst) score was ≤5%, adequate when within 5-20%, and poor when ≥20% (Andresen, 2000; Lim et al., 2008). For interval measures, we applied the following thresholds: ≥11 s for

9HPT, ≤240 m for 2MWT, ≥2 s for T25FW. As 9HPT and T25FW are stopped respectively after 5 and 3 min if not completed, the ceilings were set at 13 s, 2 s, 235 m, the floors at 250 s, 150 s, 5 m, for 9HPT, T25FW, and 2MWT respectively. Hence a minimum change of 20% in 9HPT or T25FW and of 5 m in 2MWT between rehabilitation entry and discharge could be detected (Baert et al., 2014; Kragt et al., 2006) .

2.4.2 Concurrent validity

To investigate the cross-sectional concurrent validity, i.e. the concept that OMs truly measure what they are supposed to, we followed the approach of McDowell and Jerkinson (McDowell and Jenkinson, 1996).

We first hypothesized expected ranges of correlation coefficients for each pair of OMs (Appendix, tables A1, A2, A4), based on their constructs and literature (Brosseau and Wolfson, 1994; Fischer et al., 1999; Franchignoni et al., 2003; Gijbels et al., 2012).

We then calculated Spearman correlation coefficients for each pair of OMs at baseline and checked whether they fell within the expected range (and thus indicating good concurrent validity). In addition, the strength of the observed correlation was classified as weak (0-0.3), moderate (0.3-0.5), strong (0.5-0.8), and very strong (0.8-1).

Since all the OMs are associated with impairment level (EDSS) and disease duration, we checked whether, after accounting for these two variables, each pair of OMs were still associated. For this analysis we performed partial Spearman correlation on each pair of OMs (table A3) and checked if they lay within the predefined correlation ranges (table A2).

We also assessed the longitudinal concurrent validity. We calculated for each pair of OMs Spearman correlation of the differences between baseline and discharge values (tables A5), and

then checked if they lay within the predefined correlation ranges (table A4).

2.4.3 Relative Responsiveness

For an OM to detect whether a patient's status has changed and how, it must assure that it will itself change accordingly. This property is called relative responsiveness when more OMs are simultaneously assessed, and it was investigated by computing the standardized response mean (SRM) (Puhan et al., 2007), with 95% confidence intervals obtained under the assumption of normality. We deemed the responsiveness poor if SRM<0.5, adequate when 0.5-0.8, excellent if ≥0.8 (Cohen, 1977). Pairwise statistical significance of the SRM differences was assessed for all possible OM combinations with paired t-test (Puhan et al., 2007).

All analyses were performed using R v3.3.1 (R foundation, Vienna, Austria).

3. Results

3.1 Participants

There were 920 recorded visits from 467 patients with PMS. Of these visits, 337 fulfilled the inclusion criteria, leading to 229 individual patients being included in this study (Figure 1). Tables 1 and 2 illustrate, respectively, their demographic characteristics and baseline assessments stratified by severity level.

Figure 1. Sample selection flowchart.

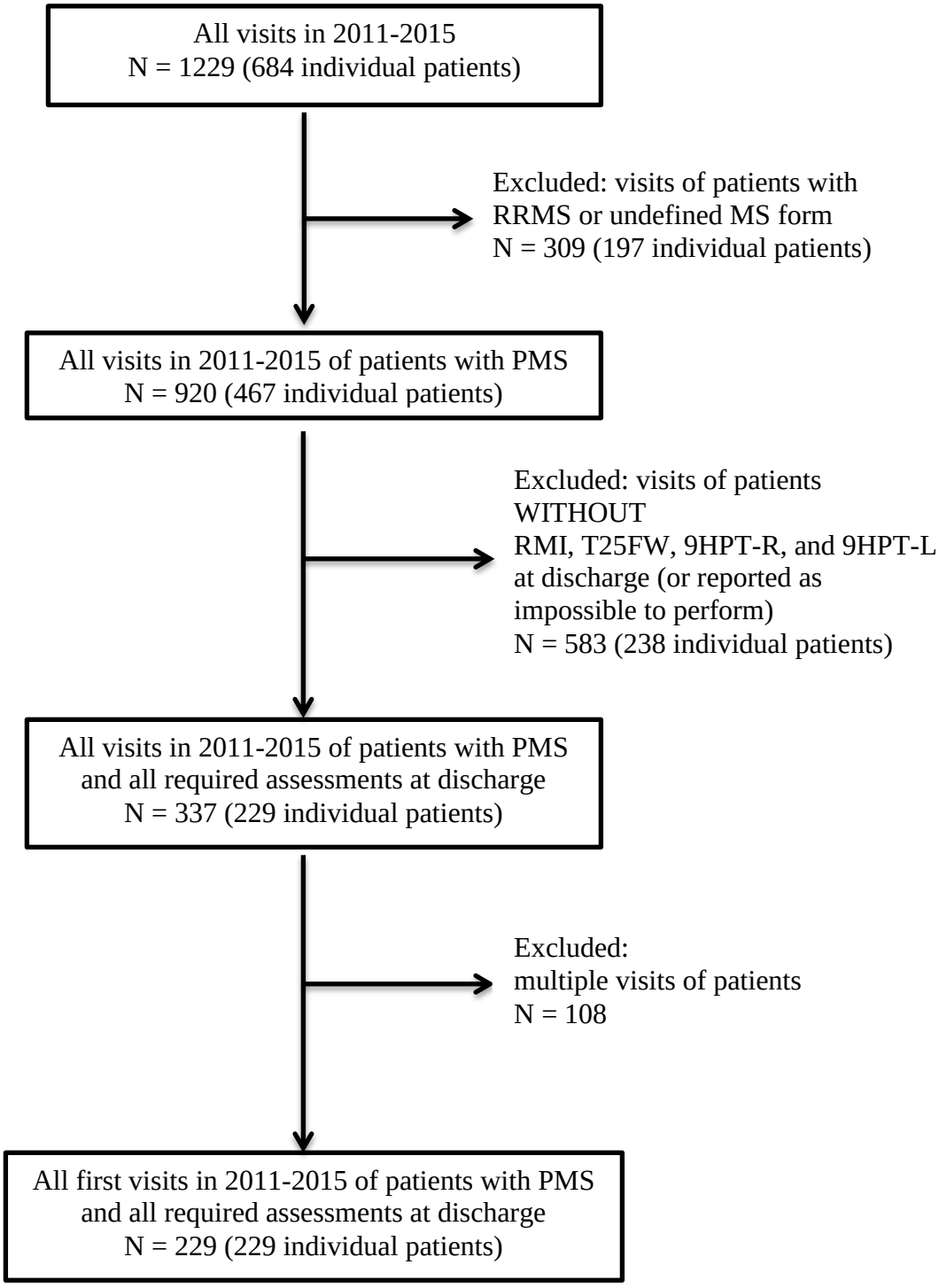


Table 1. Demographic and MS specific characteristics.

		MS subgroups		
Variable	Total MS	Mild MS	Moderate MS	Severe MS
Size (n)	229	24	132	73
Age	57.1 ± 11.9	51.3 ± 14.6	58.0 ± 11.4	57.6 ± 11.3
Gender F, n (%)	148 (64.6%)	18 (75.0%)	85 (64.4%)	45 (61.6%)
Years since diagnosis	18.6 ± 9.7	16.6 ± 13.2	18.1 ± 9.3	20.3 ± 9.0
MS form:				
SP, n (%)	197 (86.0%)	21 (87.5%)	112 (84.8%)	64 (87.7%)
PP, n (%)	27 (11.8%)	3 (12.5%)	17 (12.9%)	7 (9.6%)
P, n (%)	5 (2.2%)	0 (0%)	3 (2.3%)	2 (2.7%)
Length of stay in days	21.7 ± 7.8	21.7 ± 4.0	20.8 ± 4.6	23.4 ± 12.1
EDSS	6.2 ± 1.3	3.5 ± 0.5	6 ± 0.6	7.5 ± 0.5
Residence:				
German-speaking Switzerland, n (%)	121 (52.8%)	13 (54.2%)	67 (50.8%)	41 (56.1%)
French-speaking Switzerland, n (%)	107 (46.7%)	11 (45.8%)	65 (49.2%)	31 (42.5%)

Results are shown as mean ± standard deviation. SP: secondary progressive; PP: primary progressive; P: progressive MS form, not distinguishable between primary and secondary; EDSS: Kurtzke's Expanded Disability Status Scale; Mild MS: 0-4 EDSS; Moderate MS: 4.5-6.5 EDSS; Severe MS: 7-10 EDSS.

Table 2. **Baseline clinical assessments: mobility and cognitive measures.**

Variable	Total MS	MS subgroups		
		Mild MS	Moderate MS	Severe MS
Size (n)	229	24	132	73
RMI	9.0 (4.0-12.0)	13.0 (12.0-14.0)	10.0 (8.0-12.0)	2.5 (1.0-5.0)
FIM-mot*	75.0 (69.5-80.0)	82.5 (79.8-85.3)	76.0 (73.5-79.0)	70.0 (53.5-78.0)
FIM-cog*	30.0 (27.0-33.0)	32.5 (32.0-33.5)	31.0 (27.8-33.0)	29.0 (24.5-31.0)
2MWT (m)	44.0 (0-83.0)	130.0 (90.0-159.0)	60.0 (37.0-87.0)	0 (0-0)
T25FW (s)	18.0 (10.2-180.0)	7.0 (5.0-11.0)	13.9 (9.0-21.0)	180 (180.0-180.0)
9HPT-R (s)	33.0 (26.0-46.0)	24.5 (23.0-28.8)	29.0 (25.0-38.0)	46.0 (37.0-79.0)
9HPT-L (s)	36 (28.0-48.8)	28 (25.3-37.8)	31 (27.0-41.0)	46 (38.0-73.5)
SaGAS10	2.5 (1.2-4.3)	5.5 (3.7-6.2)	3.4 (2.2-4.5)	1.2 (0.8-1.6)

Results are shown as median (interquartile range).

Mild MS: 0-4 EDSS; Moderate MS: 4.5-6.5 EDSS; Severe MS: 7-10 EDSS; RMI: Rivermead Mobility Index; FIM-mot: Functional Independence Measure motor subscale; FIM-cog: Functional Independence Measure cognitive subscale; 2MWT: 2-Minute Walk Test; T25FW: Timed 25-Foot Walk; 9HPT-R: Nine-Hole Peg Test-Right hand; 9HPT-L: Nine-Hole Peg Test-Left hand; SaGAS10: Short and Graphic Ability Score 10.

* FIM available for n=4; 40; and 23 patients with mild; moderate; and severe MS; respectively.

3.2 Acceptability

In table 3 the distributional characteristics of each OM are illustrated. Observed data covered theoretical ranges well for RMI, 2MWT, and the 9HPTs, whereas other OMs only partially covered their full range, likely owing to our study focus on more severely disabled patients. Except for RMI scores, all other OMs displayed highly skewed distributions with means shifted towards worse

scores (e.g. the 2MWT with a mean of 52 m and a midpoint of 120 m). In terms of floor and ceiling effects, the EDSS performed well, whereas FIM-mot, SaGAS10, and 9HPT-R had only adequate floor and ceiling effects. While walk tests generally performed well with respect to ceiling effects, they performed poorly on floor effects, owing to the inclusion of non-ambulatory patients.

Table 3. Outcome measures distribution features.

	Coverage			Distribution			Floor/ceiling effects	
Variable	N (over 229)	Scale range	Actual range	Mean	SD	Skewness	Floor effects (worst score)	Ceiling effects (best score)
EDSS	229	0-10	2.5-8.5	6.2	1.3	-0.7	0%	0%
RMI	224	0-15	0-15	8.0	4.5	-0.4	5.8%	1.8%
FIM-mot	67	13-91	23-91	72.0	13.4	-2.0	0%	1.5%
FIM-cog	67	5-35	11-35	29.1	5.2	-1.5	0%	10%
2MWT (m)	209	0-240	0-225	52.0	50.4	0.9	28.7%	0%
T25FW (s)	216	2-180	3.5-180	62.2	73.4	0.9	26.9%	0%
9HPT-R (s)	214	11-300	15-303	49.2	53.4	3.7	3.3%	0%
9HPT-L (s)	214	11-300	12-300	54.9	63.6	4.4	5.1%	0.9%
SAGAS10	201	0-10	0-6.8	2.8	1.8	0.5	3.9%	0%

The cells highlighted represent the fields where each OM is not adequate.

SD: Standard Deviation; EDSS: Kurtzke's Expanded Disability Status Scale; RMI: Rivermead Mobility Index; FIM-mot: Functional Independence Measure motor subscale; FIM-cog: Functional Independence Measure cognitive subscale; 2MWT: 2-Minute Walk Test; T25FW: Timed 25-Foot Walk; 9HPT-R: Nine-Hole Peg Test-Right hand; 9HPT-L: Nine-Hole Peg Test-Left hand; SaGAS10: Short and Graphic Ability Score 10.

3.3 Concurrent validity

Table 4 illustrates cross-sectional correlation coefficients and whether their absolute values lie within our pre-specified ranges (table A1 of the Appendix). 9HPT-L is not reported because of its similarity to 9HPT-R. Both in the total sample and in the severe subgroup the hand dexterity tests correlate strongly (ρ=0.63 and ρ=0.70, respectively).

Table 4. Cross-sectional concurrent validity: Spearman correlation.

Sample	Outcome measure	SaGAS10	RMI	FIM-mot	FIM-cog	2MWT	T25FW	9HPT-R
Total MS	EDSS	**-0.75**	**-0.84**	**-0.45**	**-0.43**	**-0.84**	0.82	**0.49**
	SaGAS10	1	0.79	0.50	**0.48**	**0.88**	**-0.69**	-0.70
	RMI		1	0.56	**0.43**	0.87	-0.84	-0.55
	FIM-mot			1	0.65	**0.54**	**-0.48**	-0.45
	FIM-cog				1	**0.52**	**-0.49**	**-0.34**
	2MWT					1	-0.96	-0.55
	T25FW						1	0.54
Severe MS	EDSS	-0.19	-0.55	**-0.36**	**-0.37**	-0.55	**0.59**	0.17
	SaGAS10	1	0.45	**0.35**	-0.05	0.31	-0.37	-0.81
	RMI		1	0.54	**0.32**	**0.66**	-0.68	-0.36
	FIM-mot			1	0.63	0.45	-0.27	-0.42
	FIM-cog				1	0.41	-0.22	-0.11
	2MWT					1	-0.95	-0.23
	T25FW						1	0.17

The correlations written in bold were within the hypothesized ranges.

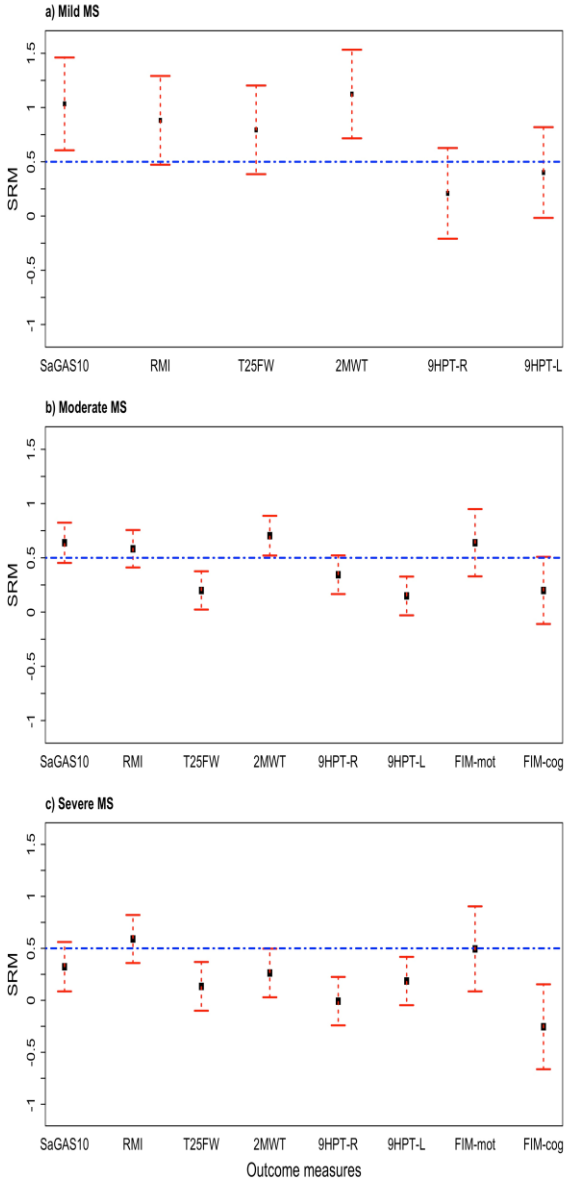
EDSS: Kurtzke's Expanded Disability Status Scale; SaGAS10: Short and Graphic Ability Score 10; RMI: Rivermead Mobility Index; FIM-mot: Functional Independence Measure motor subscale; FIM-cog: Functional Independence Measure cognitive subscale; 2MWT: 2-Minute Walk Test; T25FW: Timed 25-Foot Walk; 9HPT-R: Nine-Hole Peg Test-Right hand; 9HPT-L: Nine-Hole Peg Test-Left hand; Severe MS: 7-10 EDSS. 9HPT-L is not reported, because of its similarity to 9HPT-R.

With some exceptions (RMI, T25FW and 9HPT-R), the a priori hypotheses of construct correlations were satisfied in the total sample. However, T25FW exhibited unexpectedly strong correlations with RMI and 2MWT. For instance, the correlation of T25FW and 2MWT was hypothesized to lie between -0.8 and -0.5 but instead was -0.92. RMI correlates strongly instead of moderately (as expected) with FIM-mot and 9HPT-R, and very strongly instead of strongly (expectation) with the walk tests. Surprisingly, 9HPT-R exhibited moderate to strong correlations with all other OMs, which do not assess upper-body mobility. It is known that 9HPT and the walking tests depend on the MS severity level (Fischer et al., 1999), so we performed partial correlation analysis adjusted for EDSS level and disease duration. We found all partial correlations were within the hypothesized ranges (table A2-A3 in the Appendix).
Looking at differences between baseline and discharge measurements, most OMs also showed correlation coefficients in the hypothesized ranges (tables A4-A5 in the Appendix).

3.4 Relative responsiveness

In figure 2 the SRM values with 95% confidence interval (CI) are illustrated, stratified by severity level. Table A6 shows the paired t-tests per severity level of the SRM differences for each pair of OMs.

Figure 2. Standardized Response Mean with 95% CI.
(single column fitting image, colour only online)



The dotted horizontal line at 0.5 represents the threshold for the SRM to be adequate.
Mild MS: 0-4 EDSS; Moderate MS: 4.5-6.5 EDSS; Severe MS: 7-10 EDSS; SaGAS10: Short and Graphic Ability Score 10; RMI: Rivermead Mobility Index; T25FW: Timed 25-Foot Walk; 2MWT: 2-Minute Walk Test; 9HPT-R: Nine-Hole Peg Test-Right hand; 9HPT-L: Nine-Hole Peg Test-Left hand; FIM-mot: Functional Independence Measure motor subscale; FIM-cog: Functional Independence Measure cognitive subscale.

The SRM behaves quite differently according to severity level. In the mild group, RMI, SaGAS10 and the walk tests have an excellent responsiveness, but they are all comparable (as indicated by the overlapping confidence intervals in figure 2). By comparison, the 9HPTs have a poor-to-moderate responsiveness.

In the moderate disease severity group SaGAS10, RMI, 2MWT, and FIM-mot all have an equally moderate to excellent responsiveness, whereas T25FW, 9HPTs and FIM-cog are only poorly responsive.

In the severe subgroup, all the walk-related tests further lose their responsiveness. The 9HPTs remain poorly responsive, whereas RMI and FIM-mot achieve a moderate responsiveness.

4. Discussion

In this study we retrospectively analyzed OMs for 229 patients with PMS who underwent inpatient rehabilitation in an established, specialized clinic. Our analysis revealed that the gold standard OMs in MS rehabilitation such as 9HPT and T25FW (Potter et al., 2014) fail to be adequate in a group of persons with PMS from routine care with more severe mobility impairments than usually seen in other very selected study populations, and when the follow up time is less than a month. First, there were substantial percentages of persons who were unable to perform some tasks. Second, in terms of responsiveness the OMs performed particularly poorly among persons with moderate and severe MS. By contrast, up to an EDSS of 6.5 the 2MWT demonstrated a better responsiveness, but it was not feasible for most persons with severe MS, resulting in a floor effect. Additionally, 2MWT had a very skewed distribution and exhibited only a small within-sample variability. For persons with severe mobility impairments (EDSS $\geq$7) the only responsive measures were RMI and FIM-mot, which encompass more body functioning dimensions. However, RMI and

FIM-mot are generic measures and showed problematic tendencies in our analysis in terms of substantial floor effects and very skewed distributions.

PMS is increasingly the focus of international efforts such as the Progressive MS Alliance, that aims at accelerating the development and approval process of treatments for PMS types (Zaratin et al., 2016). Persons with more severe impairments (such as patients with PMS) are rarely represented in validation studies, thus questioning the application of standard OMs as endpoints in treatment evaluations for PMS.

To our knowledge, this is the first validation study to specifically look at OM performance in persons with PMS. Nonetheless, our study results are in line with clinical trials involving less impaired patients. For instance, Baert and colleagues showed in a multicentre study that long walking tests, such as 2MWT, are more responsive than short ones, (e.g. the T25FW) (Baert et al., 2014). Interestingly, our finding of a limited responsiveness of 9HPT is challenged by a more positive evaluation of 9HPT in a recent Ocrelizumab trial (Montalban et al., 2017) or the Methotrexate trial (Goodkin et al., 1995), but may be explained by the much longer follow-up period between measurements of at least 3 years instead of 3 weeks as in our study.

Recently, there were several publications on novel OMs for persons with severe MS. For example, Sung et al. published a validation study of the Function In Sitting Test (FIST), which is a measure of sitting balance that is specifically designed for non-ambulatory MS patients (Sung et al., 2016). In addition, the 5 repetition Sit-to-Stand test (5-STS) was shown to be valid for MS patients with EDSS <7, but its feasibility for non-ambulatory patients remains unclear (Møller et al., 2012).

4.1 Implications on practice

When considering only a moderate to severely impaired population with MS (as frequently observed in the group of persons with PMS), the OMs already validated for MS do not show the expected properties of acceptability and responsiveness.

Based on our results, we suggest the use of 2MWT, as well as FIM-mot and RMI up to moderate MS. Moreover, other potentially interesting OMs, namely the FIST and the 5-STS, could be well suited for severe MS (Møller et al., 2012; Sung et al., 2016). In addition, to better identify therapy responders, one could include in the core set MS-specific, patient-reported outcomes, such as MSIS-29 (Hobart et al., 2001) and MSQoL-54 (Vickrey et al., 1995). However, time restrictions may pose an obstacle for implementation in routine care, particularly if they need to be administered via interview due to disability severity.

4.2 Study limitations

The present study has evident limitations. First, it is a retrospective analysis of routine OM data. Additionally, requiring data completeness might have affected the sample representativeness. For this reason our inclusion criteria were kept to a necessary minimum.

Moreover, a preferable approach to prove responsiveness would be to employ anchor-based methods. Due to the lack of a suitable external clinical endpoint or a patient-rated improvement measure as an anchor, we took a distribution-based approach.

In addition, we did not distinguish between patients with primary and secondary progressive MS in our analysis. It is not clear whether considering them separately would have had an effect on the OMs' properties (Campbell et al., 2016).

Finally, it would have been worth exploring the test-retest reliability for the OMs assessed, but no data was collected in that regard.

Nevertheless, this study represents the first attempt in PMS to simultaneously compare rehabilitation OMs. To our knowledge, our study is among the largest for validation studies in MS.

4.3 Conclusion

In summary, our results highlight further potential for improvement in routine neurorehabilitation assessments, particularly for patients with moderate to severe mobility impairment, which is frequently encountered among persons with a progressive form. Our findings stress the need for OMs that are specifically developed and validated for persons with more severe MS, as the majority of currently recommended measures exhibited inadequate measurements characteristics.

Conflict of Interest

The authors report no conflicts of interest.

Acknowledgements

This study was supported by an unconditional grant from the Swiss Multiple Sclerosis Society. The Society had no role in study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

We would like to thank Jenny Piket for checking the grammar and improving the clarity of this manuscript, and Haris Babačić for helping in reviewing the medical charts.

References

Amato, M.P., Portaccio, E., 2007. Clinical outcome measures in multiple sclerosis. J. Neurol. Sci. 259, 118–122. https://doi.org/10.1016/j.jns.2006.06.031

Andresen, E.M., 2000. Criteria for assessing the tools of disability outcomes research. Arch. Phys. Med. Rehabil. 81, S15–S20.

Baert, I., Freeman, J., Smedal, T., Dalgas, U., Romberg, A., Kalron, A., Conyers, H., Elorriaga, I., Gebara, B., Gumse, J., Heric, A., Jensen, E., Jones, K., Knuts, K., Maertens de Noordhout, B., Martic, A., Normann, B., Eijnde, B.O., Rasova, K., Santoyo Medina, C., Truyens, V., Wens, I., Feys, P., 2014. Responsiveness and clinically meaningful improvement, according to disability level, of five walking measures after rehabilitation in multiple sclerosis: a European multicenter study. Neurorehabil. Neural Repair 28, 621–631. https://doi.org/10.1177/1545968314521010

Barin, L., Rapillard, T., Von Wyl, V., Vaney, C., 2016. Validity and responsiveness of SaGAS 10 for rehabilitation research in multiple sclerosis (MS): an improved score for routine assessments, in: ECTRIMS Online Library. London, United Kingdom, 14-17 September 2016, p. 146185.

Brosseau, L., Wolfson, C., 1994. The inter-rater reliability and construct validity of the Functional Independence Measure for multiple sclerosis subjects. Clin. Rehabil. 8, 107–115. https://doi.org/10.1177/026921559400800203

Campbell, E., Coulter, E.H., Mattison, P.G., Miller, L., McFadyen, A., Paul, L., 2016.

Physiotherapy rehabilitation for people with progressive multiple sclerosis: A systematic review. Arch. Phys. Med. Rehabil. 97, 141–151. https://doi.org/10.1016/j.apmr.2015.07.022

Cohen, J., 1977. Statistical power analysis for the behavioural sciences., rev. ed. ed. Academic Press, New York.

Coleman, C.I., Sobieraj, D.M., Marinucci, L.N., 2012. Minimally important clinical difference of the Timed 25-Foot Walk Test: results from a randomized controlled trial in patients with multiple sclerosis. Curr. Med. Res. Opin. 28, 49–56. https://doi.org/10.1185/03007995.2011.639752

Elovaara, I., Giovannoni, G., Havrdová, E., Kesselring, J., Kobelt, G., Langdon, D., Morrow, S.A., Oreja-guevara, C., Schippling, S., Thalheim, C., Thompson, H., Vermersch, P., Aston, K., Bauer, B., Demory, C., Paz, M., Hlavacova, J., Nouvet-gire, J., Pepper, G., Pontaga, M., Rogan, E., Rogalski, C., Galen, P. Van, Ben-amor, A., 2018. Unmet needs, burden of treatment, and patient engagement in multiple sclerosis : A combined perspective from the MS in the 21st Century Steering Group. Mult. Scler. Relat. Disord. 19, 153–160. https://doi.org/10.1016/j.msard.2017.11.013

Fischer, J.S., Rudick, R.A., Cutter, G.R., Reingold, S.C., Ms, N., Clinical, S., 1999. The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment. Mult. Scler. 5, 244–250. https://doi.org/10.1177/135245859900500409

Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R., 1998. Evaluating patient-based outcome measures for use in clinical trials. Health Technol. Assess. (Rockv). 2, 19–

43.

Franchignoni, F., Tesio, L., Benevolo, E., Ottonello, M., 2003. Psychometric properties of the Rivermead Mobility Index in Italian stroke rehabilitation inpatients. Clin. Rehabil. 17, 273–282. https://doi.org/10.1191/0269215503cr608oa

Gijbels, D., Alders, G., Van Hoof, E., Charlier, C., Roelants, M., Broekmans, T., Eijnde, B.O. 'T, Feys, P., 2010. Predicting habitual walking performance in multiple sclerosis: relevance of capacity and self-report measures. Mult. Scler. 16, 618–626. https://doi.org/10.1177/1352458510361357

Gijbels, D., Dalgas, U., Romberg, A., de Groot, V., Bethoux, F., Vaney, C., Gebara, B., Medina, C.S., Maamâgi, H., Rasova, K., de Noordhout, B.M., Knuts, K., Feys, P., 2012. Which walking capacity tests to use in multiple sclerosis? A multicentre study providing the basis for a core set. Mult. Scler. 18, 364–71. https://doi.org/10.1177/1352458511420598

Goodkin, D., Rudick, R., VanderBrug Medendorp, S Daughtry, M., Schwetz, K., Fischer, J., Van Dyke, C., 1995. Low-dose (7.5 mg) oral methotrexate reduces the rate of progression in chronic progressive multiple sclerosis. Ann. Neurol. 37, 30–40.

Hobart, J., Lamping, D., Fitzpatrick, R., Riazi, A., Thompson, A., 2001. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. Brain 124, 962–973. https://doi.org/10.1093/brain/124.5.962

Hoogervorst, E.L.J., Kalkers, N.F., Uitdehaag, B.M.J., Polman, C.H., 2002. A study validating changes in the multiple sclerosis functional composite. Arch. Neurol. 59, 113–116. https://doi.org/10.1001/archneur.59.1.113

Jette DU, Halbert J, I.C., 2009. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys. Ther. 89, 125–135. https://doi.org/10.2522/ptj.20080234

Kaufman, M., Moyer, D., Norton, J., 2000. The significant change for the Timed 25-foot Walk in the multiple sclerosis functional composite. Mult. Scler. 6, 286–90. https://doi.org/10.1177/135245850000600411

Khan, F., Turner-Stokes, L., Ng, L., Kilpatrick, T., Amatya, B., 2007. Multidisciplinary rehabilitation for adults with multiple sclerosis (Review). Cochrane Database Syst. Rev. Art. No.: CD006036. https://doi.org/10.1002/14651858.CD006036.pub2

Kragt, J., van der Linden, F., Nielsen, J., Uitdehaag, B., Polman, C., 2006. Clinical impact of 20% worsening on Timed 25-foot Walk and 9-hole Peg Test in multiple sclerosis. Mult. Scler. 12, 594–598. https://doi.org/10.1177/1352458506070768

Larocca, N.G., Hudson, L.D., Rudick, R., Amtmann, D., Balcer, L., Benedict, R., Bermel, R., Chang, I., Chiaravalloti, N.D., Chin, P., Cohen, J.A., Cutter, G.R., Davis, M.D., Deluca, J., Feys, P., Francis, G., Goldman, M.D., Hartley, E., Kapoor, R., Lublin, F., Lundstrom, G., Matthews, P.M., Mayo, N., Meibach, R., Miller, D.M., Motl, R.W., Mowry, E.M., Naismith, R., Neville, J., Panagoulias, J., Panzara, M., Phillips, G., Robbins, A., Sidovar, M.F., Smith, K.E., Sperling, B., Uitdehaag, B.M.J., 2017. The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability. Mult. Scler. J. 1–16. https://doi.org/10.1177/1352458517723718

Lim, L.L.Y., Seubsman, S.A., Sleigh, A., 2008. Thai SF-36 health survey: Tests of data

quality, scaling assumptions, reliability and validity in healthy men and women. Health Qual. Life Outcomes 6, 1–9. https://doi.org/10.1186/1477-7525-6-52

McDowell, I., Jenkinson, C., 1996. Development standards for health measures. J Hlth Serv Res Policy 1, 238–246.

Møller, A.B., Bibby, B.M., Skjerbæk, A.G., Jensen, E., Sørensen, H., Stenager, E., Dalgas, U., 2012. Validity and variability of the 5-repetition sit-to-stand test in patients with multiple sclerosis. Disabil. Rehabil. 34, 2251–2258. https://doi.org/10.3109/09638288.2012.683479

Montalban, X., Hauser, S.L., Kappos, L., Arnold, D.L., Bar-Or, A., Comi, G., de Seze, J., Giovannoni, G., Hartung, H.-P., Hemmer, B., Lublin, F., Rammohan, K.W., Selmaj, K., Traboulsee, A., Sauter, A., Masterman, D., Fontoura, P., Belachew, S., Garren, H., Mairon, N., Chin, P., Wolinsky, J.S., 2017. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. N. Engl. J. Med. 376, 209–220. https://doi.org/10.1056/NEJMoa1606468

Potter, K., Cohen, E.T., Allen, D.D., Bennett, S.E., Brandfass, K.G., Widener, G.L., Yorke, A.M., 2014. Outcome Measures for Individuals With Multiple Sclerosis: Recommendations From the American Physical Therapy Association Neurology Section Task Force. Phys. Ther. 94, 593–608. https://doi.org/10.2522/ptj.20130149

Puhan, M.A., Guyatt, G.H., Goldstein, R., Mador, J., McKim, D., Stahl, E., Griffith, L., Schünemann, H.J., 2007. Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. Respir. Med. 101,

308–316. https://doi.org/10.1016/j.rmed.2006.04.023

Rabadi, M.H., Vincent, A.S., 2013. Comparison of the Kurtkze Expanded Disability Status Scale and the Functional Independence Measure: measures of multiple sclerosis-related disability. Disabil. Rehabil. 35, 1877–1884. https://doi.org/10.3109/09638288.2013.766269

Rosti-Otajärvi, E., Hämäläinen, P., Koivisto, K., Hokkanen, L., 2007. The reliability of the MSFC and its components - Rosti-Otajärvi - 2007 - Acta Neurologica Scandinavi. Acta Neurol. Scand. 117, 421–427. https://doi.org/10.1111/j.1600-0404.2007.00972.x

Skinner, A., Turner-Stokes, L., 2006. The use of standardized outcome measures in rehabilitation centres in the UK. Clin. Rehabil. 20, 609–615. https://doi.org/10.1191/0269215506cr981oa

Stucki, G., 2016. Olle Höök lectureship 2015: The World Health Organization's paradigm shift and implementation of the international classification of functioning, disability and health in rehabilitation. J. Rehabil. Med. 48, 486–493. https://doi.org/10.2340/16501977-2109

Sung, J., Ousley, C.M., Shen, S., Isaacs, Z.J., Sosnoff, J.J., Rice, L.A., 2016. Reliability and validity of the function in sitting test in nonambulatory individuals with multiple sclerosis. Int. J. Rehabil. Res. 39, 308–312. https://doi.org/10.1097/MRR.0000000000000188

Trochim, W.M.K., Donnelly, J.P., 2006. The research methods knowledge base, 3rd ed. Atomic Dog, Cincinnati.

Vaney, C., Vaney, S., Wade, D.T., 2004. SaGAS, the Short and Graphic Ability Score: an

alternative scoring method for the motor components of the Multiple Sclerosis Functional Composite. Mult. Scler. 10, 55–60. https://doi.org/10.1191/1352458504ms1000oa

Vickrey, B., Hays, R., Harooni, R., Myers, L., Ellison, G., 1995. A health-related quality of life measure for multiple sclerosis. Qual. Life Res. 4, 187–206.

Zaratin, P., Comi, G., Coetzee, T., Ramsey, K., Smith, K., Thompson, A., Panzara, M., 2016. Alliance Industry Forum : Maximizing Collective Impact To Enable Drug Development. Trends Pharmacol. Sci. 37, 808–810. https://doi.org/10.1016/j.tips.2016.07.005