2015•2016
FACULTEIT GENEESKUNDE EN LEVENSWETENSCHAPPEN
master in de revalidatiewetenschappen en de kinesitherapie

Masterproef

Reliability of upper limb outcome measures in people with multiple sclerosis: a European RIMS study

Promotor :
dr. Ilse LAMERS

Copromotor :
dr. Ilse BAERT

Thomas Derboven

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

2015•2016
FACULTEIT GENEESKUNDE EN LEVENSWETENSCHAPPEN
master in de revalidatiewetenschappen en de kinesitherapie

Masterproef

Reliability of upper limb outcome measures in people with multiple sclerosis: a European RIMS study

Promotor :
dr. Ilse LAMERS

Copromotor :
dr. Ilse BAERT

Thomas Derboven

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie

Masterproef deel 2

Thomas Derboven, Bsc.; Ilse Lamers, Dr.; Ilse Baert, Dr.

RELIABILITY OF UPPER LIMB OUTCOME MEASURES IN PEOPLE WITH MULTIPLE SCLEROSIS: A EUROPEAN RIMS STUDY

I. Acknowledgement

After two years, this manuscript is the final result of my master thesis. With a different subject in my first and second master year, it has been especially varied and interesting. As a student in the field of internal rehabilitation, this master thesis in the neurological field has broadened my knowledge into the neurological field of multiple sclerosis.

First of all, I would like to thank my promotor Dr. Ilse Lamers. Her positivism, knowledge and experience always encouraged me to improve this manuscript. Thank you for all of your help, enthusiasm and advice. Furthermore, thanks to Dr. Ilse Baert and Lieven Vanbuel for their help with the data retrieved from the participants, and to my fellow student Tina Hendrickx, with whom I have worked together during the first part of this master thesis.

Finally thanks to my parents, sister, family, friends and girlfriend Yana for all of their support in any possible way.

Geel, 18/08/2016 TD

II. Research context

The study described in this master thesis was part of a multicenter study which investigated the psychometric properties of upper limb outcome measures in the neurological disease of multiple sclerosis. Multiple sclerosis (MS) is a progressive auto-immune disorder affecting the central nerve system, which in 75% of these patients reduces unilateral or bilateral manual dexterity (Johansson et al., 2007; Ytterberg, Johansson, Andersson, Widen Holmqvist, & von Koch, 2008). The multicenter study was set up within the European Rehabilitation In Multiple Sclerosis (RIMS) network for best practice and research (www.eurims.org), more specifically within the special interest group (SIG) on occupation. Eleven research centers across Europe and two centers in the United States participated in this multicenter study. Only at the Charles University in Prague, Czech Republic data was collected twice prior to rehabilitation and used to investigate reliability of upper limb outcome measures. The research protocol was designed and coordinated at Hasselt University, REVAL under supervision of Prof. Dr. P. Feys, Dr. I. Baert and Dr. I. Lamers. Data collection was performed by therapists at Charles University, while data analysis and interpreting of its results was performed by the master thesis student.

Ytterberg, C., Johansson, S., Andersson, M., Widen Holmqvist, L., & von Koch, L. (2008). Variations in functioning and disability in multiple sclerosis. A two-year prospective study. *J Neurol, 255*(7), 967-973. doi:10.1007/s00415-008-0767-0

Johansson, S., Ytterberg, C., Claesson, I. M., Lindberg, J., Hillert, J., Andersson, M., . . . von Koch, L. (2007). High concurrent presence of disability in multiple sclerosis. Associations with perceived health. *J Neurol, 254*(6), 767-773. doi:10.1007/s00415-006-0431-5

III. Article

Reliability of upper limb outcome measures in people with multiple sclerosis: a European RIMS study

1. Abstract

Background; No core set of appropriate outcome measures is available yet to evaluate upper limb function in people with MS (PwMS). Unfortunately, until now the required knowledge on psychometric properties of these outcome measures is lacking in PwMS.

Objectives; To investigate the reliability of outcome measures in PwMS.

Methods; Upper limb function of sixteen subjects at the Charles University located in Prague, Czech Republic was assessed prior to rehabilitation at two measuring points, within a time interval of 5 to 9 days. On the body functions and structures level, maximum isometric pinch, key, trippod grip and hand strength was evaluated using a dynamometer. Numeric rating scale (NRS) for spasticity and visual analogue scales (VAS) of muscle weakness, sensory impairment, coordination and fatigability were used to evaluate the perceived presence of impairments. Other tests on the body functions and structures level were the plate tapping test and trunk impairment scale - Norwegian version (TIS-NV). On the activity level the nine hole peg test (NHPT), box and block test (BBT), coin rotation test (CRT), action research arm test (ARAT), manual ability measure - 36 (MAM-36), ABILHAND and motor activity log (MAL) were conducted to assess upper limb capacity and performance. Intraclass correlation coefficients (ICC) were used to investigate test-retest, interrater and intrarater reliability, while internal consistency was evaluated using Cronbach's alpha.

Results; In test-retest situations pinch strength, hand grip strength, MAM-36, ABILHAND and MAL proved to be reliable with ICC values ranging from 0.59 to 0.99, while spasticity numeric rating scale and VAS scales tended to be more variable with lower ICC values. Inter -and intrarater reliability of NHPT, BBT, CRT, ARAT, TIS-NV and plate tapping test all resulted in high ICC values above 0.86. Cronbach's alpha values above 0.95 for MAM-36 and ABILHAND indicate good internal consistency of these outcome measures.

Conclusion; Except for NRS and VAS-scales all outcome measures were found to be reliable measures for PwMS, both for test-retest reliability, inter -and intrarater reliability and internal consistency.

2. Introduction

Multiple sclerosis (MS) is a progressive auto-immune disorder affecting the central nerve system resulting in de-myelination and degeneration, in which patients develop physical and cognitive impairments that correspond with the affected nerve fibers (ICD-10). About 75% of these patients show reduced unilateral or bilateral manual dexterity (Johansson et al., 2007; Ytterberg, Johansson, Andersson, Widen Holmqvist, & von Koch, 2008), influencing the performance of activities of daily living (ADL) and thereby decreasing quality of life and independence (Lamers et al., 2015; Yozbatiran, Baskurt, Baskurt, Ozakbas, & Idiman, 2006). Considering this critical role of the upper limb function in PwMS, special attention should go towards outcome measures that assess progression of upper limb dysfunction. To assess upper limb impairment more completely and to understand the impact of upper limb dysfunction on ADL, it is important to use outcome measures on the different levels of the International Classification of Functioning (ICF). (Lemmens, Timmermans, Janssen-Potten, Smeets, & Seelen, 2012; Metcalf, Adams, Burridge, Yule, & Chappell, 2007; Velstra, Ballert, & Cieza, 2011). In contrast to other neurological diseases, there is no core set of appropriate outcome measures to evaluate the upper limb function in MS available yet. In order to develop this core set, knowledge on the psychometric properties is necessary. Unfortunately, unlike in other neurological conditions, little is known about the psychometric properties of outcome measures in MS. One systematic review provided an overview of applied upper limb outcome measures and their psychometric properties in MS and concluded there was a lack of studies investigating the psychometric properties of upper limb outcome measures in MS (Lamers, Kelchtermans, Baert, & Feys, 2014).

Therefore the multicenter study aimed to investigate the psychometric properties (reliability, validity, responsiveness and floor/ceiling effects) of new and frequently used upper limb outcome measures. This master thesis will only report on the reliability of upper limb outcome measures in MS. Reliability is the extent of which a measurement is consistent and free of error, or reproducibility of a measurement under given conditions (Portney L., 2009). There are different types of reliability. Test-retest reliability assesses whether an instrument can measure a variable with consistency, measurements are gathered from a single rater who uses the same methods or instruments and the same testing conditions. Inter-rater reliability -as a part of rater reliability- assesses the variation between two or more raters who measure the same group of subjects. Intra-rater reliability however, refers to the stability of measurements recorded by one individual across different trials. Alternate forms reliability checks the interchangeability of different versions of a single measuring instrument, this is most frequently used in paper-and-pencil tests. The last type of reliability is internal consistency, which reflects the extent to which different items on e.g. a questionnaire measure different aspects of the same characteristic and nothing else (Portney L., 2009). This manuscript will be limited to test-retest reliability, rater reliability and internal consistency. In summary, the primary aim of this study is to investigate the test-retest reliability, rater reliability and internal consistency of new and frequently used upper limb outcome measures in PwMS.

3. Methods

3.1. Participants

3.1.1. Recruitment

Participants were recruited within several rehabilitation centers across Europe within the Rehabilitation in Multiple Sclerosis (RIMS) network. Data used for this reliability study was only collected in the Charles University located in Prague, Czech Republic. The assessment of participants took place from September 2014 until March 2015.

3.1.2. Selection (in-exclusion criteria)

Participants were included when following criteria were met: 1) diagnose of MS using the McDonald criteria (Polman et al., 2011), 2) receiving treatment which aims to maintain or improve the upper limb function, 3) no relapse within the last six months prior to the study, 4) no changes in disease modifying medication and no corticoid-therapy within the last month prior to the study and 5) received at least ten sessions of upper limb rehabilitation (in- or outpatient rehabilitation), with a maximal duration of 3 months. Participants were excluded if; 1) they were not able to understand and execute simple instructions or 2) have other medical conditions interfering with upper limb function (e.g. fractures, stroke, pregnancy). The study was approved by the ethical committees of UZ Leuven S56575(B322201421636; 31/07/2014) and the local ethical committee of the participating center. A written informed consent of all included participants was obtained before enrolment in the study.

3.1.3. Descriptive measures

A range of different descriptive data were recorded at baseline: age, gender, height, weight, disease duration (since diagnosis), type of MS, drug use. Cognitive function, more particular information processing speed, was evaluated using the symbol digit modalities test (SDMT). In this test rows of nine symbols are arranged pseudo-randomly, while participants are asked to say the number corresponding with each symbol (Smith, 1982). Severity of MS-related cognitive and motor fatigue was measured using the fatigue scale for motor and cognitive functions (FSMC) consisting of 20 questions with a five point likert-scale, a minimum score of 20 expressing no fatigue at all and a maximum score of 100 being the severest grade of fatigue (Penner et al., 2009). To assess the participants disability level, the golden standard expanded disability status scale (EDSS) is used. It is scored in half point scores ranging from 0.0 (normal) to 10.0 (dead from MS) (Kurtzke, 1983).

3.2. Procedure

The data collection for this reliability study took place at the Charles University in Prague, Czech Republic. Descriptive and experimental measures were conducted prior to the rehabilitation program. Various upper limb outcome measures were measured during two sessions with 5 to 9 days in between the sessions prior to rehabilitation. Both arms were included in testing. To minimize measurement biases between assessors some precautions were made. A standardized instruction booklet was provided to the assessors, containing details on test procedures such as order of testing, resting periods, level of encouragement, use of assistive devices and the possibility to divide testing into two visits to avoid fatigue of participants or fit the time schedule of assessors. For correct assessment and scoring of tests, uniform scoring sheets were used and a training session was organized at the University of Hasselt.

3.3. Experimental measures

Experimental outcome measures, clinician reported as well as patient reported, are described below following the different ICF levels and are shown in figure 1.

3.3.1. Clinician reported outcome measures

Body functions and structures level

On this level, the maximum isometric hand grip strength and isometric pinch strength were measured using average values of three trials. Pinch strength was assessed on a pinch gauge dynamometer or the E-link hand evaluation kit (Biometrc Ltd) performing a key, tri-pod or tip-tip grip, while hand grip strength was assessed using the JAMAR hand grip dynamometer (Chen, Kasven, Karpatkin, & Sylvester, 2007; Guclu-Gunduz, Citaker, Nazliel, & Irkec, 2012). The plate tapping test was used to assess the speed and coordination of the upper limb. Participants move one upper limb back and forth between two discs as quickly as possible, while the other upper limb remains in the middle. Time needed to complete 25 cycles (50 taps) is recorded. The last test on this level was the trunk impairment scale - Norwegian version. It was originally developed to measure motor impairment of the trunk by Verheyden et al. as the trunk impairment scale (Verheyden et al., 2004). It has been translated in Norwegian and modified to the trunk impairment scale - Norwegian version to a 6-item hierarchically organized ordinal rating scale (Gjelsvik et al., 2012). Its total score ranged from 0 to 16, where higher scores indicated less trunk impairment.

Capacity measures on activity level

The nine hole peg test (NHPT) assesses unilateral fine manual dexterity, 9 pegs are placed and removed one at a time from a board as quickly as possible with a time limit of 300 seconds. Time needed to perform the NHPT is recorded and pegs/s were calculated (Cutter et al., 1999).

The box and block test (BBT) was used to assess unilateral gross manual dexterity, the goal is to transfer as many blocks, one at a time from one box to another, during one minute (Platz et al., 2005).

Manual dexterity and in-hand manipulation were measured using the coin rotation task (CRT), where a coin is rotated as fast as possible between the thumb, index, and middle finger. Time needed to perform 20 half turns (180 degrees) is measured for both hands, or the number of turns in one minute is recorded if it is not possible to perform 20 half turns in one minute. The number of turns per second were calculated and used in the data analysis (Kamm et al., 2012).

3.3.2. Patient reported outcome measures

Body functions and structures level

Visual analogue Scales (VAS) were used to evaluate the perceived muscle weakness, coordination, sensory impairment and fatigability. Respondents specify their level of agreement to a statement by indicating a position along a continuous line between two end-points (0-10). For perceived severity of spasticity, the Spasticity 0-10 numeric rating scale was used. Spasticity was defined to the participants as the muscle stiffness they were experiencing, with anchors of 0 indicating 'no spasticity' and 10 meaning 'worst possible spasticity' (Farrar, Troxel, Stott, Duncombe, & Jensen, 2008).

Perceived performance measures on activity level

Measures on the activity level were the MAM-36, ABILHAND and the MAL. The Manual Ability Measurement - 36 (MAM-36) is a questionnaire on perceived performance of 36 unilateral and bilateral ADL tasks rated using a 4-point scale (Chen & Bode, 2010). Perceived performance of bilateral ADL tasks during the last 3 months is evaluated using the ABILHAND. 23 bilateral ADL tasks are rated using a 3-point ordinal scale. (Barrett, Cano, Zajicek, & Hobart, 2013). Quantity and quality of upper limb use in daily life was rated by using a modified version of the upper extremity motor activity log (MAL). Upper limb use during ADL tasks was scored by participants using a 6-point ordinal scale (0-5 points) reflecting the 'amount of use' and the 'quality of use', where the sum score of both scales resulted in a 'total score' (0-10) (Lamers, Kerkhofs, et al., 2013; Lamers, Timmermans, et al., 2013; Mark et al., 2008).

3.4. Data-analysis

Test-retest and both types of rater reliability were analyzed using the intraclass correlation coefficient (ICC) and standard error of measurement (SEM). For test-retest and interrater reliability a 2,k ICC model was used while a 3,k ICC model was used for intrarater reliability. ICC values below 0.50 indicated poor, from 0.50 to 0.75 suggested moderate, from 0.75 to 0.9 indicate good and above 0.90 was considered excellent reliability (Portney L., 2009). SEM was calculated as SD * $\sqrt{1-icc}$, and reflects the standard deviation of the measurement errors. For internal consistency Cronbach's alpha was used, this was calculated for MAM-36, ABILHAND and MAL. Values approaching 0.90 were considered reliable (Portney L., 2009). To present a visual image of reliability, Bland-Altman plots were generated. Next to the measures of reliability, minimal detectable change (MDC), which was calculated as SEM * 1.96 * $\sqrt{2}$, was reported. To test the normality of the results, a Shapiro-Wilk test was executed. IBM SPSS was used to perform the analysis, significance level was set at p 0.05.

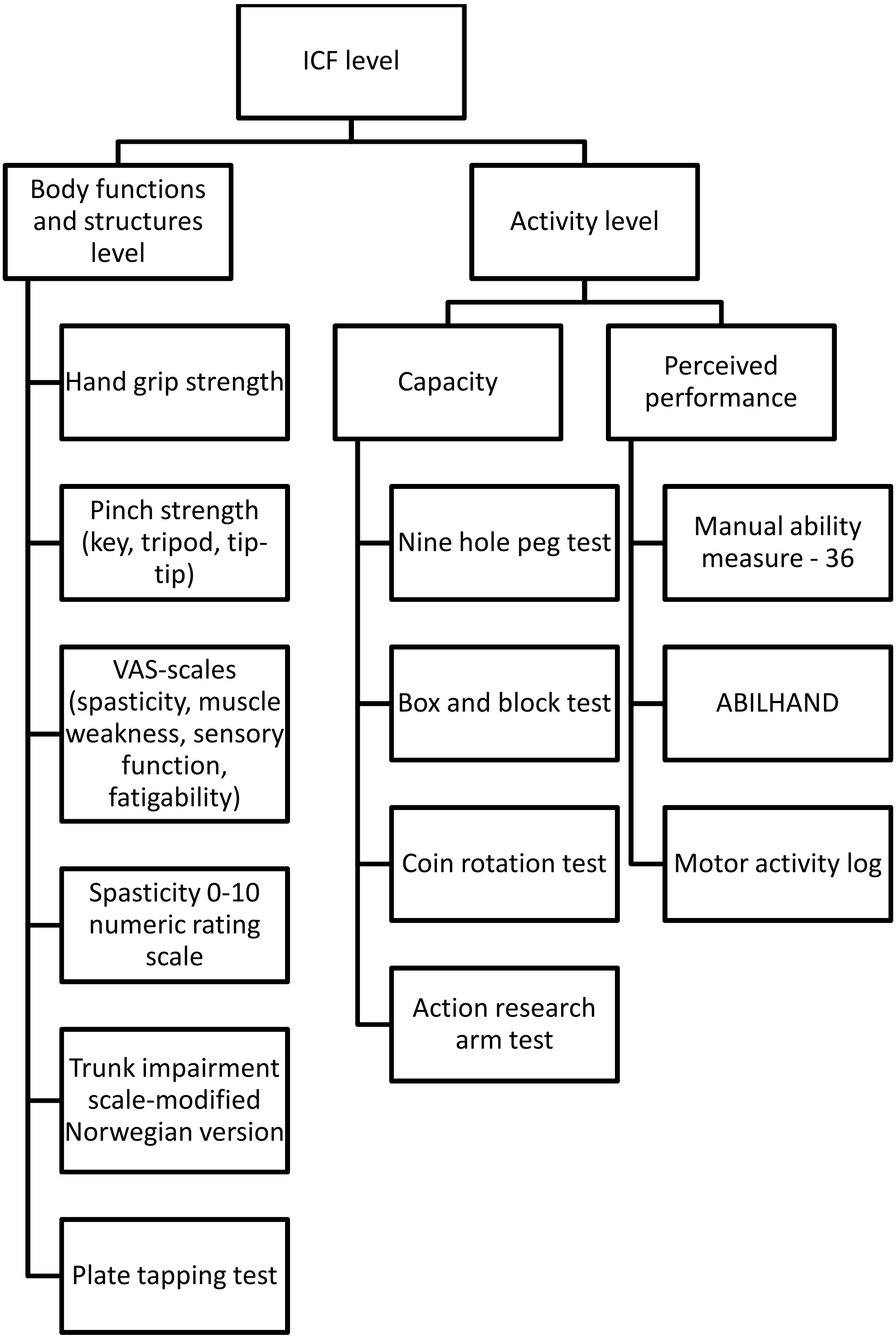


Figure 1 Outcome measures

4. Results

4.1. Patient and outcome characteristics

Sixteen subjects were included in this reliability study. Outcomes of all sixteen subjects were used to assess test-retest reliability, while intra rater reliability was assessed on ten subjects, and inter rater reliability on six subjects. Details on patient characteristics are showed in table 1. Subjects suffered from different types of MS (6.25% involving primary progressive or relapsing progressive MS, 31.25% secondary progressive MS and 62.5% relapsing remitting MS), disease severity was expressed by a mean EDSS score of 4.47, indicating significant disability where activities of daily living are affected by the disease but the person is still able to walk.

Table 1 Descriptive measures

Variable	n (%), mean ± SD
Gender	
Male	4 (25%)
Female	12 (75%)
Age in years	45.31 ± 10.69
Type of MS	
PP/RPMS	1 (6.25%)
SP	5 (31.25%)
RR	10 (62.5%)
EDSS score	4.47 ± 2.24
Walking aid	
None	8 (50%)
Cane, crutch (unilateral)	2 (12.5%)
Cane, crutch (bilateral)	1 (6.25%)
Rollator	3 (18.75%)
Wheelchair	2 (12.5%)
Hand dominance	
Right	14 (87.5%)
Left	2 (12.5%)
FSMC	
Total	63.94 ± 16.63
Motor	33.63 ± 8.36
Cognitive	30.31 ± 9.12
SDMT	44.13 ± 13.56

PP: primary progressive, RP: relapsing progressive, SP: secondary progressive, RR: relapsing remitting, EDSS: expanded disability status scale, FSMC: fatigue scale for motor and cognitive functions, SDMT: symbol digit modalities test

Descriptive statistics of UL outcome measures are reported in table 2, structured by observation, hand dominance and ICF level, where the activity level was divided in capacity and perceived performance measures. Pinch strength varied from 3.0 to 6.61 kg, hand grip strength varied from 22.53 to 25.75 kg. The spasticity 0-10 numeric rating scale and all VAS scales tended to vary between the first and second assessment and between subjects, with VAS scale assessing fatigue yielding the highest score with a mean of 3.38 on the first and 2.66 on de second assessment. Trunk stability of participants as measured by the trunk impairment scale resulted in mean scores of 13.5 and 11.69. On the activity level, the box and block test with a range of 48.06 to 48.75 blocks/min and the nine hole peg test with 0.33 to 0.37 pegs/s recorded fine motor skill level. Participants scored well on the action research arm test with a median score of 54 to 55. In the questionnaires of the performance scales, high scores were found on the motor activity log, indicating low quality and quantity of movement. Scores on both questionnaires evaluating activities of daily living (ADL), ABILHAND and manual ability measure 36 indicated PwMS having difficulties performing ADL tasks.

4.2. Reliability

The different values of reliability are reported in tables 3 through 5 and will be discussed below following the different ICF levels, Bland-Altman plots are found in the appendix.

Body functions and structures

ICC values for test-retest reliability for pinch and hand strength ranged from 0.59 for key-pinch of the non-dominant hand to 0.99 for hand grip strength of the dominant and non-dominant hand, with SEM and MDC ranging from 0.5 to 1.54 and 1.3 to 4.26, respectively. For the spasticity 0-10 numeric rating scale low reliability was detected by ICC values of 0.53 for the dominant, and 0.42 for the non dominant hand. In VAS-scales there was greater variability between the ICC's with a poor ICC for the VAS fatigue to good reliability for the VAS muscle weakness, SEM ranged from 0.9 to 1.79 and MDC from 2.5 to 4.96. Rater reliability analysis revealed excellent ICC values for both inter -and intrarater reliability of the plate tapping test and TIS-NV with SEM and MDC values of 0.97 to 3.08 and 2.7 to 8.53 for the plate tapping test, while these values ranged from 0.77 to 1.19 and 2.15 to 3.3 for the trunk impairment scale.

Table 2. Experimental measures

Variable		Observation 1	observation 2
Body functions and structures level			
Pinch strength (kg)			
Key	D	6.28 ± 3.28	6.61 ± 1.77
	ND	6.34 ± 2.94	6.21 ± 1.81
Tripod	D	4.68 ± 2.66	4.57 ± 1.53
	ND	4.18 ± 2.80	4.1 ± 2.08
Tip to tip	D	3.38 ± 1.89	3.67 ± 1.45
	ND	3.0 ± 1.77	3.31 ± 1.45
Hand grip strength (kg)	D	22.53 ± 13.37	25.75 ± 12.76
	ND	25.27 ± 13.84	24.42 ± 11.80
NRS (0-10)	D	2.94 ± 2.35	2.69 ± 2.09
	ND	2.38 ± 2.83	2.5 ± 2.25
VAS scales (0-10)			
Muscle weakness	D	3.38 ± 2.68	2.98 ± 2.87
	ND	1.9 ± 2.03	2.59 ± 2.70
Sensitivity	D	2.47 ± 1.96	2.39 ± 2.57
	ND	3.24 ± 2.43	2.28 ± 2.50
Coordination	D	3.13 ± 2.99	2.46 ± 2.75
	ND	2.42 ± 2.40	2.38 ± 2.93
Fatigue		3.83 ± 2.29	2.66 ± 2.41
TIS (0-16)		12.13 ± 4.40	11.69 ± 4.87
Plate tapping test (s)	D	24.16 ± 9.99	24.40 ± 11.65
	ND	23.69 ± 10.60	24.97 ± 12.15
Activity level			
Capacity			
NHPT (pegs/s)	D	0.33 ± 0.12	0.38 ± 0.11
	ND	0.33 ± 0.14	0.35 ± 0.12
BBT (blocks/min)	D	48.13 ± 14.54	48.06 ± 19.26
	ND	48.67 ± 14.90	48.75 ± 19.30
CRT (s)	D	0.93 ± 0.50	0.93 ± 0.41
	ND	0.79 ± 0.40	0.85 ± 0.34
ARAT (0-57)	D	38.5 ± 26.61	38.7 ± 26.73
	ND	39 ± 26.93	39 ± 26.94
Performance			
MAM-36 (0-100)		64.63 ± 15.38	67.03 ± 15.93
ABILHAND (logits)		3.31 ± 3.44	3.53 ± 3.11
MAL			
Amount (0-5)	D	4.31 ± 1.28	4.28 ± 1.29
	ND	3.75 ± 1.53	3.69 ± 1.65
How (0-5)	D	3.97 ± 1.12	4.06 ± 1.08
	ND	3.63 ± 1.44	3.69 ± 1.52
Total (0-10)	D	8.28 ± 2.26	8.34 ± 2.26
	ND	7.41 ± 2.76	7.38 ± 2.84

Mean ± SD

NRS: Spasticity 0-10 numeric rating scale, VAS: Visual analogue scale, TIS-NV: Trunk impairment scale - Norwegian version, NHPT: Nine hole peg test, BBT: Box and block test, CRT: Coin rotation test, ARAT: Action research arm test, MAM-36: Manual ability measure - 36, MAL: Motor activity log

Activity

Rater reliability of the capacity measures NHPT, BBT, CRT and ARAT proved to be excellent. ICC's ranged from 0.86 to 1. SEM and MDC values can be found in table 4. Test-retest and internal consistency reliability for both questionnaires –MAM-36, ABILHAND and MAL– yielded excellent ICC values, all above 0.82. SEM's ranged from 0.16 to 9.94 while MDC's ranged from 0.45 to 10.93. Internal consistency as measured by Cronbach's alpha proved that MAM-36 and ABILHAND were internally consistent with values above 0.90, while the MAL failed to reach this preset value.

Table 2. Test-retest reliability (N=16)

Variable		ICC (95%CI)	SEM	MDC
Pinch strength				
Key	D	0.73 (-0.29 - 0.94)*	1.29	3.57
	ND	0.59 (-0.96 - 0.91)	1.41	3.9
Tripod	D	0.91 (0.61 - 0.98)**	0.7	1.95
	ND	0.76 (0.07 - 0.94)*	0.99	2.74
Tip to tip	D	0.93 (0.69 - 0.98)**	0.47	1.3
	ND	0.87 (0.48 - 0.97)**	0.5	1.4
Hand grip strength	D	0.99 (0.95 - 1)**	1.39	3.86
	ND	0.99 (0.94 - 1)**	1.54	4.26
NRS	D	0.53 (-0.33 - 0.84)	1.6	4.42
	ND	0.42 (-0.69 - 0.8)	1.69	4.69
VAS scales				
Muscle	D	0.81 (0.46 - 0.94)**	1.19	3.3
	ND	0.83 (0.53 - 0.94)**	0.9	2.5
Sensitivity	D	0.64 (-0.07 - 0.88)*	1.42	3.93
	ND	0.59 (-0.06 - 0.85)*	1.52	4.21
Coordination	D	0.71 (0.17 - 0.9)*	1.65	4.57
	ND	0.71 (0.15 - 0.9)*	1.21	3.36
Fatigue		0.44 (-0.42 - 0.79)	1.79	4.96
MAM-36		0.97 (0.90 - 0.99)**	1,76	7,66
ABILHAND		0.97 (0.91 - 0.99)**	0.59	1,65
MAL				
How	D	0.93 (0.79 - 0.87)**	0.35	0.96
	ND	0.91 (0.75 - 0.97**	0.39	1.07
Amount	D	0.99 (0.97 - 1)**	0.16	0.45
	ND	0.82 (0.47 - 0.94)**	0.6	1.66
Total	D	0.97 (0.92 - 0.99)**	0.44	1.23
	ND	0.95 (0.86 - 0.98)**	0.56	1.56

* p<0.05

** p<0.01

NRS: Spasticity 0-10 numeric rating scale, VAS: Visual analogue scale, MAM-36: Manual ability measure, MAL: Motor activity log

Table 3. Rater reliability

Variable		Intrarater reliability			Interrater reliability		
		ICC (95%CI)	SEM	MDC	ICC (95%CI)	SEM	MDC
NHPT	D	0.99 (0.93 - 1)**	0.02	0	0.97 (0.53 - 1)**	0.02	0.06
	ND	0.94 (0.75 - 0.99)**	0.02	0.1	0.95 (0.63 - 0.99)**	0.03	0.08
BBT	D	0.96 (0.65 - 0.99)**	3.38	9.4	0.98 (0.87 - 1)**	2.65	7.35
	ND	0.95 (0.77 - 0.99)**	3.45	9.6	0.97 (0.79 - 1)**	3.11	8.61
CRT	D	0.98 (0.2 - 1)**	0.07	0.2	0.92 (0.46 - 0.99)**	0.1	0.29
	ND	0.96 (0.83 - 0.99)**	0.07	0.2	0.86 (0.17 - 0.98)*	0.09	0.26
ARAT	D	0.99 (0.99 - 1)**	0	0			
	ND	0.99 (0.99 - 1)**	0	0			
Plate tapping	D	0.97 (0.89 - 0.99)**	2.09	5.8	0.94 (0.47 - 0.99)**	2.57	7.13
test	ND	0.99 (0.96 - 1)**	0.97	2.7	0.92 (-0.09 - 0.99) **	3.08	8.53
TIS		0.94 (0.78 - 0.99)**	1.19	3.3	0.96 (0.71 - 0.99)**	0.77	2.15

* p<0.05

** p<0.01

NHPT: Nine hole peg test, BBT: Box and block test, CRT: Coin rotation test, ARAT: Action research arm test, TIS-NV: Trunk impairment scale - Norwegian version

Table 5. Internal consistency

Variable		Cronbach's alpha
MAM-36		0.97
ABILHAND		0.96
MAL	D	0.87
	ND	0.76

MAM-36: Manual ability measure, MAL: Motor activity log

5. Discussion

This study aimed to investigate the reliability of different upper limb outcome measures in people with MS, the results showed good reliability with exception of VAS and NRS scores

5.1. Reliability

Body functions and structures

On the body functions and structures level, good reliability of pinch- and hand grip strength was shown, except for key pinch strength of the non dominant hand. In contrast, NRS and VAS-scales scores seemed to be more variable and therefore tended to be less reliable. For test-retest reliability, excellent ICC values for hand grip strength in PwMS were found in this study which confirmed the results of Paltamaa et al., where maximum isometric grip strength proved to be reliable with ICC values of 0.98 for both left and right hand in a sample of 19 ambulatory PwMS (Paltamaa, West, Sarasoja, Wikstrom, & Malkia, 2005). Poor ICC values on the spasticity 0-10 numeric rating scale in this study were contradicted by previous research of Farrar et al., who reported ICC values of 0.83 indicating good reliability (Farrar et al., 2008). In this study, a much greater sample of 189 PwMS was studied. The average NRS value in the study of Farrar was 5.49, while we recorded scores between 2.5 and 2.94. The difference in spasticity level, together with the differences in sample size between the two studies could be an explanation for the differences in reliability coefficients. This was the first study to investigate the reliability on pinch-strength and VAS-scales in neurologic conditions. Pinch strength yielded good results. According to ICC values, key pinch strength was found to be moderately reliable where other pinch strength seemed to have good reliability. The lower ICC's found for key pinch grip could possibly be explained by the more complex finger position needed to perform the key grip. It may be possible that PwMS had difficulties performing this test, this could possibly be improved by practicing this grip on an actual key before performing the test, or adapt the test material to resemble an actual key. It seems that VAS scales tended to be less reliable, which could probably be explained by the variability of characteristics being measured by VAS-scales. Rater reliability proved to be good for the plate tapping test and TIS-NV, as did the latter's internal consistency, which was confirmed by Gjelsvik at al. who designed the Norwegian version of the trunk impairment scale and reported a ICC value of 0.77 for interrater reliability (Gjelsvik et al., 2012). It must be noted that the psychometric properties of the TIS-NV were evaluated on stroke patients. However, psychometric properties for PwMS have been researched for the original trunk impairment scale, which was found to have good interrater reliability indicated by an ICC value of 0.97 (Verheyden et al., 2006).

Activity

On the activity level, all of the assessed outcome measures showed good reliability and internal consistency, except for the MAL for which lower Cronbach's alpha values were recorded.

Reliability of upper limb impairment questionnaires is not well researched. In charcot marie tooth disease excellent test-retest reliability was established for the MAM-36 by Poole et al. who reported an ICC value of 0.96 (Poole, Huffman, Hunter, Mares, & Siegel, 2015), and ABILHAND seemed to be consistent in 468 neurologist confirmed PwMS according to a Cronbach's alpha of 0.98 (Barrett et al., 2013). Both of these findings are in line with our current study results. Lower Cronbach's alpha values of the MAL could possibly be explained by the little amount (2) of test items. Psychometric properties of capacity measures are more extensively researched, excellent rater reliability found here for NHPT, BBT, CRT and ARAT were in line with previous research. For the NHPT, Solari et al. found an ICC value of 0.98 for interrater reliability in 32 PwMS with a mean NHPT score of 27.90 seconds (Solari, Radice, Manneschi, Motti, & Montanari, 2005). Paltaama et al. examined the interrater reliability of the box and block test in 19 ambulatory PwMS, recording ICC values of 0.93 for the right hand and 0.94 for the left hand (Paltamaa et al., 2005). This study did not differentiate between dominant and non dominant hand, nor was hand dominance mentioned in the patient characteristics. The last capacity measure on the activity level was the ARAT. Unfortunately, interrater reliability could not be calculated in this study. No previous research is available on intrarater reliability, in contrary to interrater reliability on which Nijland et al. reported an ICC value of 0.92 for interrater reliability (Nijland et al., 2010). Previous literature confirms the excellent reliability coefficients found in this study on capacity measures like the NHPT, BBT, CRT and ARAT.

5.2. Limitations

Although this study was part of a bigger multicenter study on the psychometric properties of upper limb outcome measures in MS, only one center agreed to assess participants twice before rehabilitation started, making a study on reliability possible. In this center, sixteen subjects were included, of which only data of six subjects could be used to assess interrater reliability. On the other hand, several upper limb outcome measures were assessed simultaneously allowing a good comparison of reliability between diverse outcomes.

5.3. Conclusions

This study is the first to assess reliability of several upper limb outcome measures simultaneously in MS. Based on the findings of this study, we can conclude that most of the outcome measures were reliable. Pinch strength, hand grip strength, MAM-36, ABILHAND and MAL had good test-retest reliability, while the VAS scales seemed to be more variable. Data analysis for inter -and intrarater reliability NHPT, BBT, CRT, ARAT, TIS-NV and plate tapping test all resulted in great ICC values. Good Cronbach's alpha values for MAM-36, ABILHAND indicate good internal consistency of these outcome measures, although MAL showed to be less consistent. Preliminary results of this study suggests that most of the frequently used upper limb outcome measures can be used in PwMS to give a reliable image of upper limb function, although further research assessing more subjects is necessary to confirm the good results we concluded.

6. Reference list

Reference list

Barrett, L. E., Cano, S. J., Zajicek, J. P., & Hobart, J. C. (2013). Can the ABILHAND handle manual ability in MS? *Mult Scler, 19*(6), 806-815. doi:10.1177/1352458512462919

Chen, C. C., & Bode, R. K. (2010). Psychometric validation of the Manual Ability Measure-36 (MAM-36) in patients with neurologic and musculoskeletal disorders. *Arch Phys Med Rehabil, 91*(3), 414-420. doi:10.1016/j.apmr.2009.11.012

Chen, C. C., Kasven, N., Karpatkin, H. I., & Sylvester, A. (2007). Hand strength and perceived manual ability among patients with multiple sclerosis. *Arch Phys Med Rehabil, 88*(6), 794-797. doi:10.1016/j.apmr.2007.03.010

Cutter, G. R., Baier, M. L., Rudick, R. A., Cookfair, D. L., Fischer, J. S., Petkau, J., . . . Willoughby, E. (1999). Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain, 122 (Pt 5)*, 871-882.

Farrar, J. T., Troxel, A. B., Stott, C., Duncombe, P., & Jensen, M. P. (2008). Validity, reliability, and clinical importance of change in a 0-10 numeric rating scale measure of spasticity: a post hoc analysis of a randomized, double-blind, placebo-controlled trial. *Clin Ther, 30*(5), 974-985. doi:10.1016/j.clinthera.2008.05.011

Gjelsvik, B., Breivik, K., Verheyden, G., Smedal, T., Hofstad, H., & Strand, L. I. (2012). The Trunk Impairment Scale - modified to ordinal scales in the Norwegian version. *Disabil Rehabil, 34*(16), 1385-1395. doi:10.3109/09638288.2011.645113

Guclu-Gunduz, A., Citaker, S., Nazliel, B., & Irkec, C. (2012). Upper extremity function and its relation with hand sensation and upper extremity strength in patients with multiple sclerosis. *NeuroRehabilitation, 30*(4), 369-374. doi:10.3233/nre-2012-0768

Johansson, S., Ytterberg, C., Claesson, I. M., Lindberg, J., Hillert, J., Andersson, M., . . . von Koch, L. (2007). High concurrent presence of disability in multiple sclerosis.

Associations with perceived health. *J Neurol, 254*(6), 767-773. doi:10.1007/s00415-006-0431-5

Kamm, C. P., Heldner, M. R., Vanbellingen, T., Mattle, H. P., Muri, R., & Bohlhalter, S. (2012). Limb apraxia in multiple sclerosis: prevalence and impact on manual dexterity and activities of daily living. *Arch Phys Med Rehabil, 93*(6), 1081-1085. doi:10.1016/j.apmr.2012.01.008

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology, 33*(11), 1444-1452. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6685237

Lamers, I., Cattaneo, D., Chen, C. C., Bertoni, R., Van Wijmeersch, B., & Feys, P. (2015). 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, 95*(1), 65-75. doi:10.2522/ptj.20130588

Lamers, I., Kelchtermans, S., Baert, I., & Feys, P. (2014). Upper limb assessment in multiple sclerosis: a systematic review of outcome measures and their psychometric properties. *Arch Phys Med Rehabil, 95*(6), 1184-1200. doi:10.1016/j.apmr.2014.02.023

Lamers, I., Kerkhofs, L., Raats, J., Kos, D., Van Wijmeersch, B., & Feys, P. (2013). Perceived and actual arm performance in multiple sclerosis: relationship with clinical tests according to hand dominance. *Mult Scler, 19*(10), 1341-1348. doi:10.1177/1352458513475832

Lamers, I., Timmermans, A. A., Kerkhofs, L., Severijns, D., Van Wijmeersch, B., & Feys, P. (2013). Self-reported use of the upper limbs related to clinical tests in persons with multiple sclerosis. *Disabil Rehabil, 35*(23), 2016-2020. doi:10.3109/09638288.2013.771703

Lemmens, R. J., Timmermans, A. A., Janssen-Potten, Y. J., Smeets, R. J., & Seelen, H. A. (2012). Valid and reliable instruments for arm-hand assessment at ICF activity level in persons with hemiplegia: a systematic review. *BMC Neurol, 12*, 21. doi:10.1186/1471-2377-12-21

Mark, V. W., Taub, E., Bashir, K., Uswatte, G., Delgado, A., Bowman, M. H., . . . Cutter, G. R. (2008). Constraint-Induced Movement therapy can improve hemiparetic progressive multiple sclerosis. Preliminary findings. *Mult Scler, 14*(7), 992-994. doi:10.1177/1352458508090223

Metcalf, C., Adams, J., Burridge, J., Yule, V., & Chappell, P. (2007). A review of clinical upper limb assessments within the framework of the WHO ICF. *Musculoskeletal Care, 5*(3), 160-173. doi:10.1002/msc.108

Nijland, R., van Wegen, E., Verbunt, J., van Wijk, R., van Kordelaar, J., & Kwakkel, G. (2010). A comparison of two validated tests for upper limb function after stroke: The Wolf Motor Function Test and the Action Research Arm Test. *J Rehabil Med, 42*(7), 694-696. doi:10.2340/16501977-0560

Paltamaa, J., West, H., Sarasoja, T., Wikstrom, J., & Malkia, E. (2005). Reliability of physical functioning measures in ambulatory subjects with MS. *Physiother Res Int, 10*(2), 93-109.

Platz, T., Pinkowski, C., van Wijck, F., Kim, I. H., di Bella, P., & Johnson, G. (2005). Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil, 19*(4), 404-411. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15929509

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., . . . Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol, 69*(2), 292-302. doi:10.1002/ana.22366

Poole, J. L., Huffman, M., Hunter, A., Mares, C., & Siegel, P. (2015). Reliability and validity of the Manual Ability Measure-36 in persons with Charcot-Marie-Tooth disease. *J Hand Ther, 28*(4), 364-367; quiz 368. doi:10.1016/j.jht.2015.04.003

Portney L., W. M. (2009). *Foundations of ClinicalRresearch Applications to Practice*.

Solari, A., Radice, D., Manneschi, L., Motti, L., & Montanari, E. (2005). The multiple sclerosis functional composite: different practice effects in the three test components. *J Neurol Sci, 228*(1), 71-74. doi:10.1016/j.jns.2004.09.033

Velstra, I. M., Ballert, C. S., & Cieza, A. (2011). A systematic literature review of outcome measures for upper extremity function using the international classification of functioning, disability, and health as reference. *Pm r, 3*(9), 846-860. doi:10.1016/j.pmrj.2011.03.014

Verheyden, G., Nieuwboer, A., Mertin, J., Preger, R., Kiekens, C., & De Weerdt, W. (2004). The Trunk Impairment Scale: a new tool to measure motor impairment of the trunk after stroke. *Clin Rehabil, 18*(3), 326-334.

Verheyden, G., Nuyens, G., Nieuwboer, A., Van Asch, P., Ketelaer, P., & De Weerdt, W. (2006). Reliability and validity of trunk assessment for people with multiple sclerosis. *Phys Ther, 86*(1), 66-76.

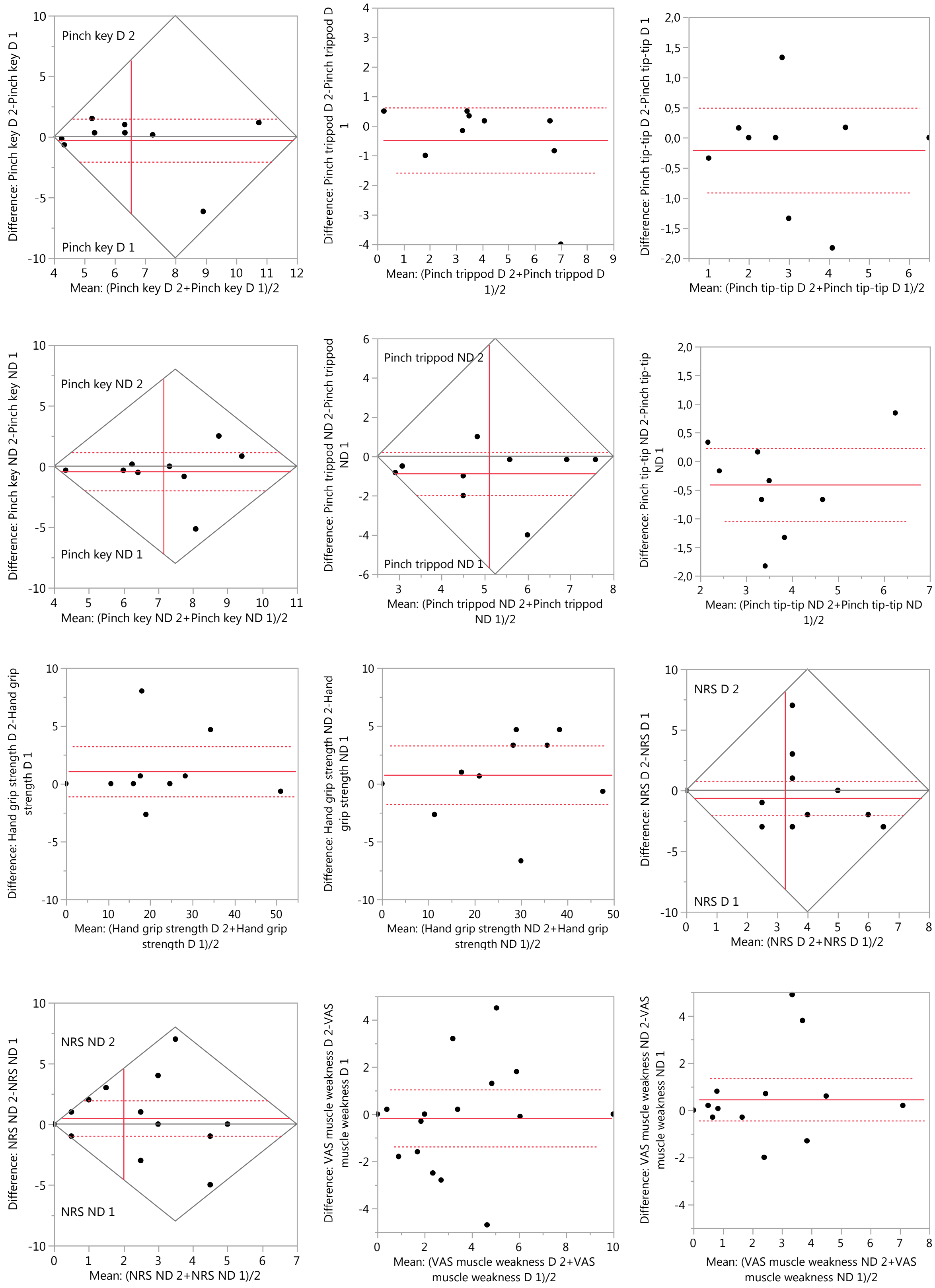
Yozbatiran, N., Baskurt, F., Baskurt, Z., Ozakbas, S., & Idiman, E. (2006). Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. *J Neurol Sci, 246*(1-2), 117-122. doi:10.1016/j.jns.2006.02.018

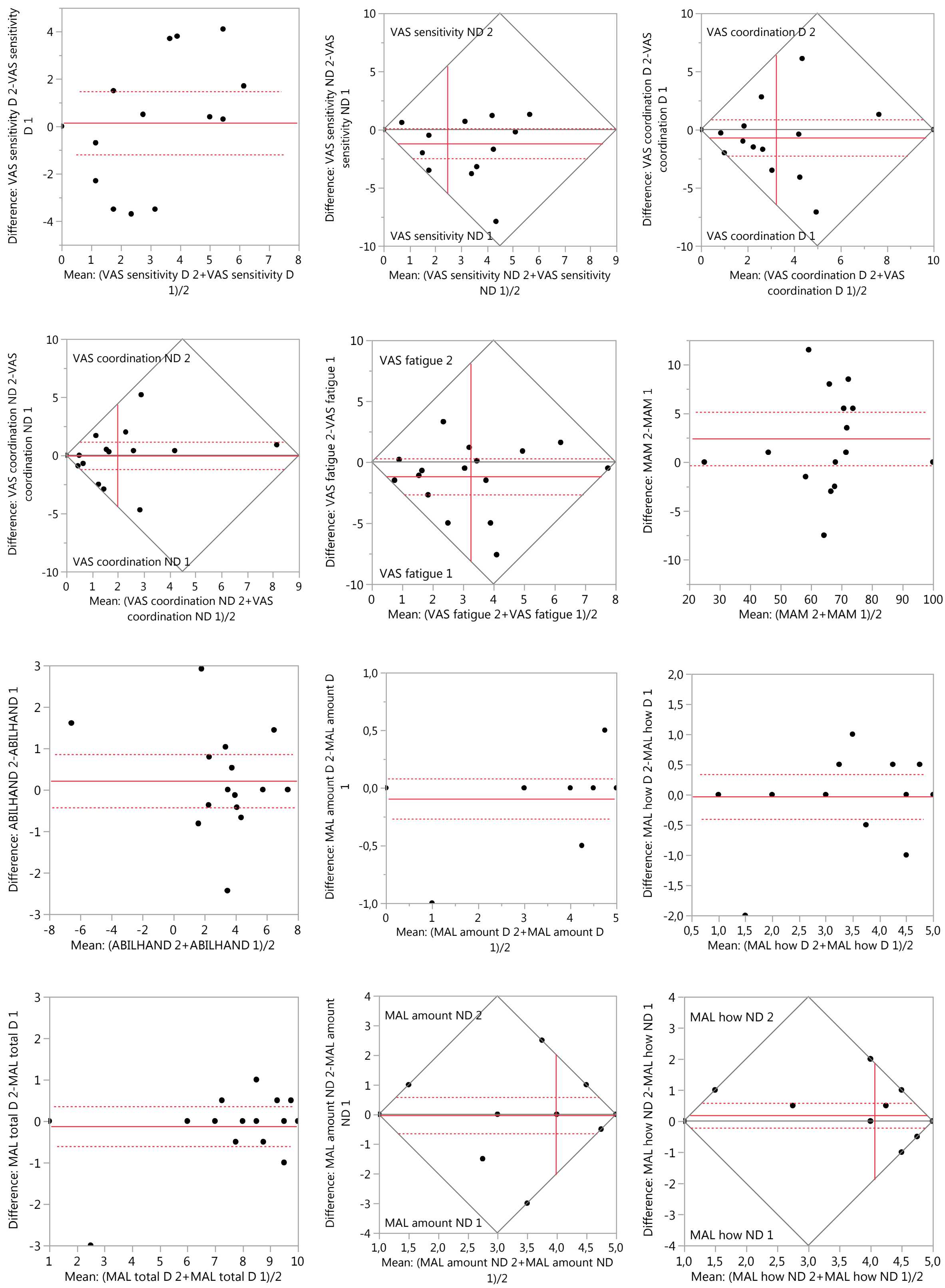
Ytterberg, C., Johansson, S., Andersson, M., Widen Holmqvist, L., & von Koch, L. (2008). Variations in functioning and disability in multiple sclerosis. A two-year prospective study. *J Neurol, 255*(7), 967-973. doi:10.1007/s00415-008-0767-0

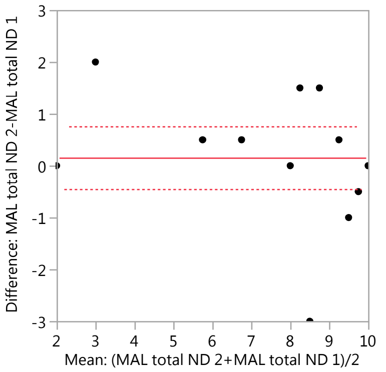
IV. Appendix

Appendix 1: Bland-Altman plots

Test-retest reliability

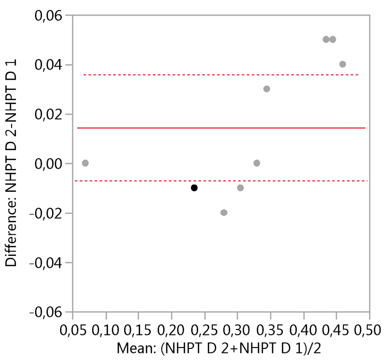


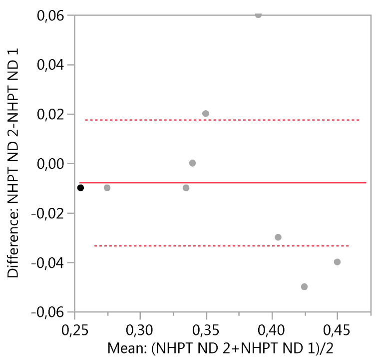

Difference: VAS sensitivity D 2-VAS sensitivity D 1
Mean: (VAS sensitivity D 2+VAS sensitivity D 1)/2
VAS sensitivity ND 2
VAS sensitivity ND 1
Difference: VAS sensitivity ND 2-VAS sensitivity ND 1
Mean: (VAS sensitivity ND 2+VAS sensitivity ND 1)/2
VAS coordination D 2
VAS coordination D 1
Difference: VAS coordination D 2-VAS coordination D 1
Mean: (VAS coordination D 2+VAS coordination D 1)/2
VAS coordination ND 2
VAS coordination ND 1
Difference: VAS coordination ND 2-VAS coordination ND 1
Mean: (VAS coordination ND 2+VAS coordination ND 1)/2
VAS fatigue 2
VAS fatigue 1
Difference: VAS fatigue 2-VAS fatigue 1
Mean: (VAS fatigue 2+VAS fatigue 1)/2
Difference: MAM 2-MAM 1
Mean: (MAM 2+MAM 1)/2
Difference: ABILHAND 2-ABILHAND 1
Mean: (ABILHAND 2+ABILHAND 1)/2
Difference: MAL amount D 2-MAL amount D 1
Mean: (MAL amount D 2+MAL amount D 1)/2
Difference: MAL how D 2-MAL how D 1
Mean: (MAL how D 2+MAL how D 1)/2
Difference: MAL total D 2-MAL total D 1
Mean: (MAL total D 2+MAL total D 1)/2
MAL amount ND 2
MAL amount ND 1
Difference: MAL amount ND 2-MAL amount ND 1
Mean: (MAL amount ND 2+MAL amount ND 1)/2
MAL how ND 2
MAL how ND 1
Difference: MAL how ND 2-MAL how ND 1
Mean: (MAL how ND 2+MAL how ND 1)/2

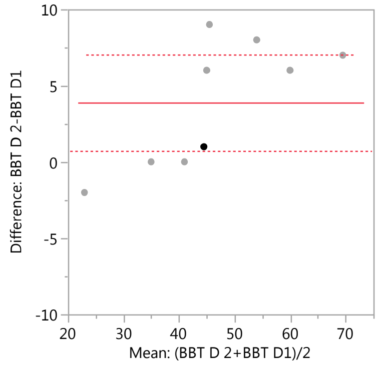


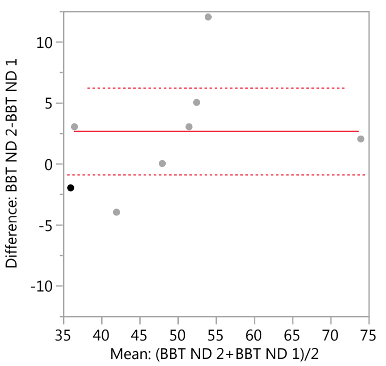
Rater reliability

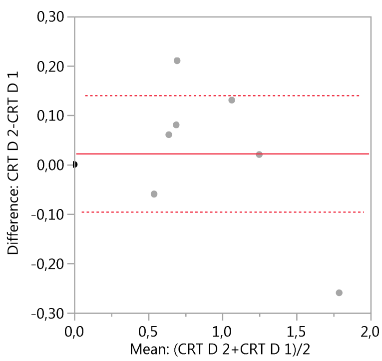
Intrarater reliability

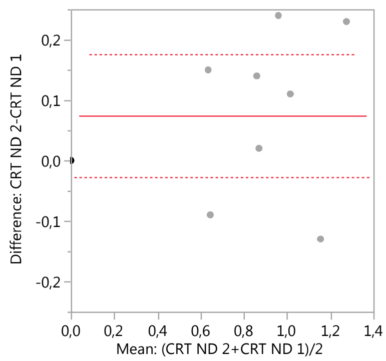


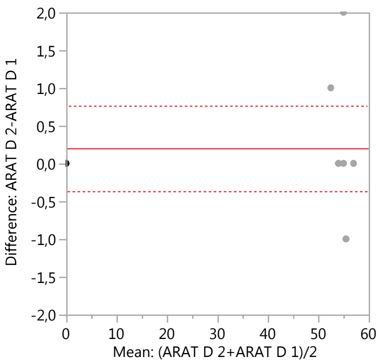


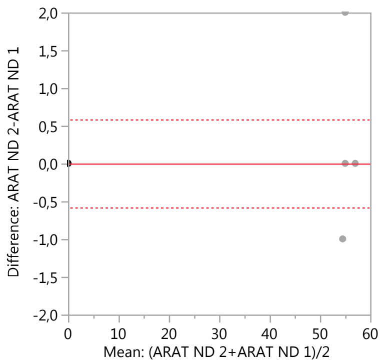


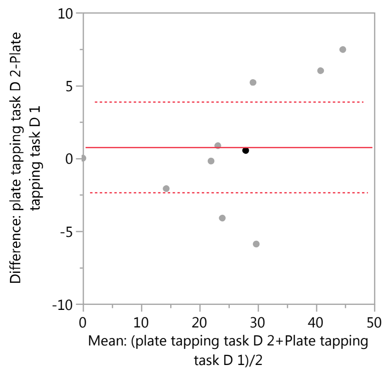


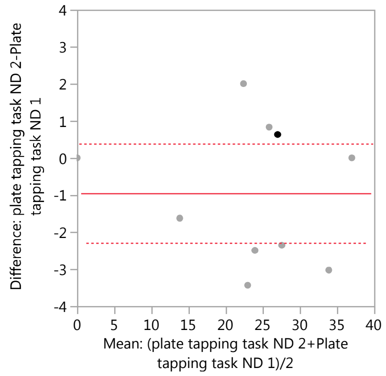


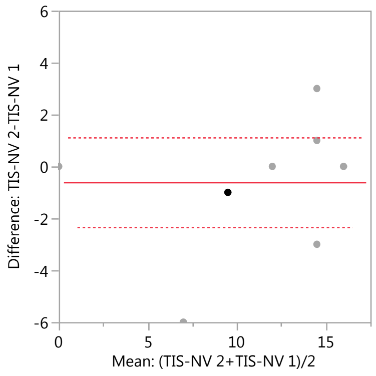




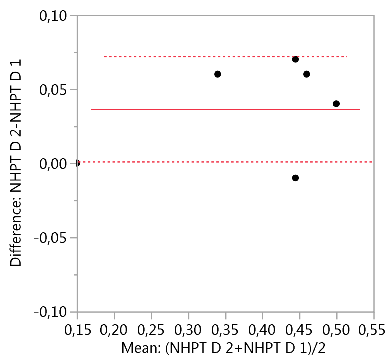


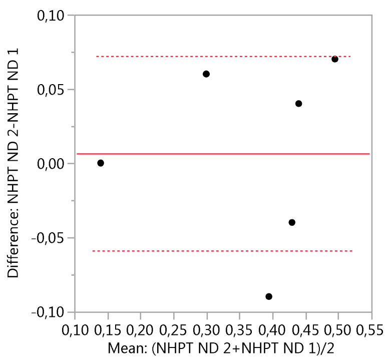


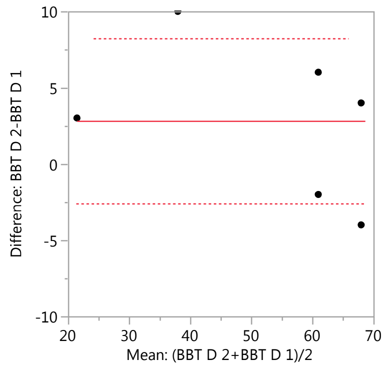


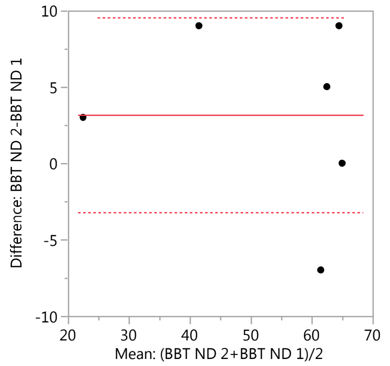


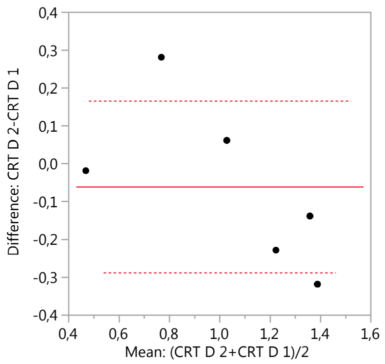
Interrater reliability

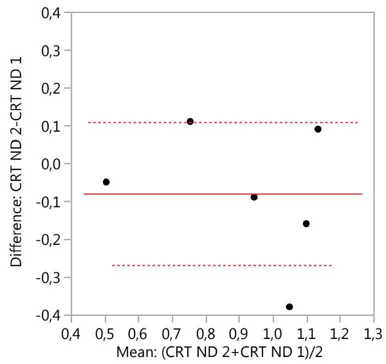


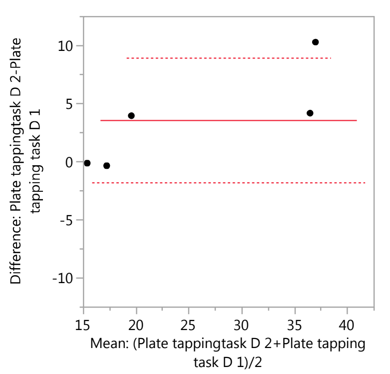


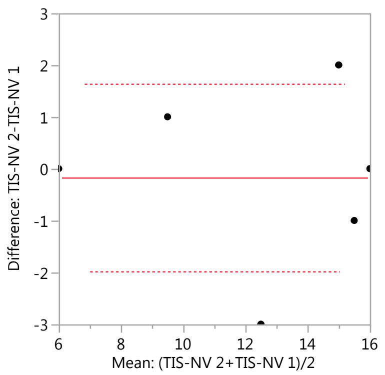


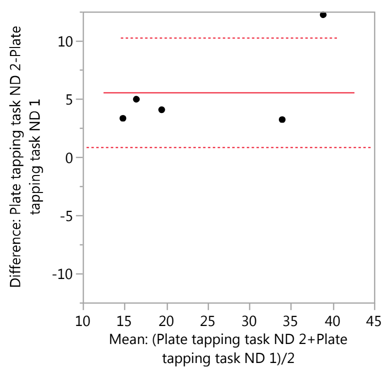












Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Reliability of upper limb outcome measures in people with multiple sclerosis: a European RIMS study

Richting: **master in de revalidatiewetenschappen en de kinesitherapie-revalidatiewetenschappen en kinesitherapie bij inwendige aandoeningen**
Jaar: **2016**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Derboven, Thomas