UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

PhD Thesis

Kirsten Lykke Knak, PT, MSc

Outcome Measures and Physical Activity in Myotonic Dystrophy Type 1

Supervisor: John Vissing, Professor, DMSci

Submitted on: 23 July 2021

This thesis has been submitted to the Graduate School of the Faculty of Health and Medical Sciences, University of Copenhagen

Outcome Measures and Physical Activity in Myotonic Dystrophy Type 1

PhD Thesis, 23 July 2021

Kirsten Lykke Knak, PT, MSc

Copenhagen Neuromuscular Center
Department of Neurology
Rigshospitalet
Denmark

Supervisor

John Vissing, MD, Professor, DMSci, Copenhagen Neuromuscular Center, Department of Neurology, Rigshospitalet, University of Copenhagen, Denmark

Co-supervisor

Nanna Witting, MD, PhD, Copenhagen Neuromuscular Center, Department of Neurology, Rigshospitalet, Denmark

Table of contents

1. PREFACE

In the process of my PhD, I have received help and support from different sites for which I am grateful for.

- Thanks to my principal supervisor John Vissing for substantial valuable guidance in the whole PhD process.
- Thanks to my secondary supervisor Nanna Witting for valuable guidance in the PhD process.
- Thanks to Henning Andersen for the collaboration and for my external research stays at his research site at Aarhus University Hospital.
- Thanks to Dorthe Sønderborg Welch for practical help related to may stays at Aarhus University Hospital.
- Thanks to Ulla Werlauff and Bente Kristensen from RCFM for informal guidance on people with myotonic dystrophy type 1.
- Thanks to Aisha Munawar Sheikh for the substantial help with testing of the participants with myotonic dystrophy type 1, for taking the photos of the test positions in this thesis, for being stand-in on my maternity leave and for your company.
- Thanks to Linda Kahr Andersen for being my co-PhD fellow and for the collaboration regarding the statistical courses in SAS.
- Thanks to medical students for the help to test healthy controls.
- Thanks to Thomas Krag for his help with Adobe Illustrator to improve figure quality.
- Thanks to Pia Hynne for preserving the overview of grants, especially regarding my maternity leave.

- Thanks for the grants to my PhD-project provided by:
    - Rigshospitalet Research Foundation
    - Axel Muusfeldt Foundation
    - Family Hede Nielsen Foundation

- Thanks to Jette Sørensen for proofreading of my PhD thesis.
- Thanks to my sister, Anne, for the residence and company during my external research stays in Aarhus.
- Thanks to my husband, Thomas, for transforming my figure drafts into impressive figures and for support and patience.
- Thanks to my daughter, Sara, for taking napes during my maternity leave so I could publish research articles.

- Thanks to all study participants who volunteered to participate in the studies of this thesis.

2. SUMMARIES

2.1 English

One of the cardinal signs in people with myotonic dystrophy type 1 (PwDM1) is muscle weakness which, among other things, affects physical function and balance. There is no cure for this disease, but to be prepared for upcoming clinical trials, psychometric properties of outcome measures within muscle strength, balance and functional mobility needs to be established. Moreover, it is likely that the muscle wasting in PwDM1 decreases the physical activity level. Because of the well-known benefits of physical activity in the general population, knowledge about the physical activity level and predictors of physical activity in these people is warranted. Intra-rater reliability, criterion- and construct validity, responsiveness at 1-year follow-up and feasibility of muscle strength-, balance-, and functional mobility measurements in non-congenital PwDM1 were investigated. Lower limb muscle strength was evaluated by Hand-Held Dynamometry (HHD) and stationary dynamometry. Balance was assessed by the dynamic balance tests: step test and Timed-Up-and-Go (TUG). Static balance was evaluated by modified Clinical Test of Sensory Integration and Balance (m-CTSIB), feet-together stance, tandem stance, and one-leg-stance eyes open and -closed. Functional mobility was measured with Sit-To-Stand (STS) and 10-meter Walk Test (10mWT). The physical activity level in PwDM1 and healthy controls were registered objectively by an accelerometer and subjectively by the International Physical Activity Questionnaire (IPAQ).

The HHD is sufficiently valid for single individuals, reliable for flexor muscles and responsive for both some of the distal and proximal muscle groups in a heterogeneous group of PwDM1. The dynamic balance tests are valuable reliable and valid tools in PwDM1, and the dynamic TUG is also responsive. The static balance tests, except the m-CTSIB, are not recommended in a heterogeneous cohort of PwDM1 due to ceiling- or floor effects. The 10mWT and the STS are recommended for reliable and valid outcomes in PwDM1, but the 10mWT is most reliable and is also responsive based on a subjective anchor.

The physical activity level in PwDM1 is predicted by educational level and lower than in healthy people. Promotion of physical activity is, therefore, warranted and especially in PwDM1 who possesses lower degrees of education.

2.2 Dansk

Et af karakteristikaene hos personer med dystrophia myotonica type 1 (PmDM1) er muskelsvaghed, hvilket blandt andet påvirker muskelfunktion og balance. Der findes ingen behandling til denne sygdom, men der er behov for at kende de psykometriske egenskaber af målemetoder inden for muskelstyrke, balance og funktion som forberedelse til kommende kliniske studier. Desuden er det sandsynligt, at muskelsvækkelsen hos PmDM1 mindsker det fysiske aktivitetsniveau. Grundet de velkendte gavnlige effekter af fysisk aktivitet i den generelle befolkning, er det relevant at få kortlagt det fysiske aktivitetsniveau og prædiktorer for fysisk aktivitet hos PmDM1.

Intra-tester reliabilitet, kriterie- og konstruktionsvaliditet, sensitivitet efter 1 år og gennemførlighed af målemetoder inden for muskelstyrke, balance og funktion hos ikke-kongenite PmDM1 blev undersøgt. Muskelstyrken i underekstremiteten blev undersøgt ved hjælp af et håndholdt dynamometer (HHD) og et stationært dynamometer. Balancen blev undersøgt med en trin-test, rejse-sig-op-og-gå test (TUG), modificeret klinisk test af sensorisk integration og balance (m-CTSIB), samlede fødder, tandem stand og et-ben-stand med øjnene åbne og lukkede. Funktionsevnen blev målt ved hjælp af en rejse-sætte-sig test (STS) og en 10-meter gangtest (10mWT). Det fysiske aktivitetsniveau hos PmDM1 og raske kontroller blev målt objektivt med en accelerationsmåler og subjektivt med det internationale spørgeskema vedrørende fysiske aktivitet.
HHD er tilstrækkeligt validt for enkelte individer, pålideligt for fleksormuskler og sensitiv for både distale og proksimale muskelgrupper hos PmDM1. De dynamiske balancetests er værdifulde pålidelige og valide testmetoder hos PmDM1, og den dynamiske TUG er også sensitiv. De statistiske balancetests, undtagen m-CTSIB, anbefales ikke til en heterogen gruppe af PmDM1 grundet loft- og gulveffekter. 10mWT og STS anbefales som pålidelige og valide målemetoder til PmDM1, men 10mWT var mest pålidelig og tillige sensitiv baseret på et subjektivt anker. PmDM1 er mindre fysiske aktive sammenlignet med raske personer og i forhold til WHO's anbefalinger for fysisk aktivitet. Uddannelse er den eneste prædiktor for fysisk aktivitet hos PmDM1. Derfor bør PmDM1 blive adresseret, og særligt de lavest uddannede, med henblik på at øge det fysiske aktivitetsniveau hos denne målgruppe.

3. LIST OF PAPERS

I

Intra-rater Reliability and Validity of Outcome Measures in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Andersen H, Witting N, and Vissing J

Neurology 2020; 94: e2508-e2520. doi:10.1212/WNL.0000000000009625

II

Responsiveness of Outcome Measures in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Witting N, and Vissing J

Annals of Clinical and Translational Neurology 2020; 7(8): 1382-1391

III

Physical Activity in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Witting N, and Vissing J

Journal of Neurology 2020; 267: 1679-1686

4. ABBREVIATIONS

PwDM1	People with Myotonic Dystrophy type 1
DM1	Myotonic Dystrophy type 1
HC	Healthy Controls
PA	Physical Activity
HHD	Hand-Held Dynamometry
m-CTSIB	modified Clinical Test of Sensory Integration and Balance
STS	Sit-To-Stand test
10mWT	10-meter Walk Test
TUG	Timed-Up-and-Go test
GRS	Global Rating Scale
MIRS	Muscular Impairment Rating Scale
AES-S	Apathy Evaluation Scale (Self-rated)
FSS-7	Fatigue Severity Scale (7 items)
IPAQ	International Physical Activity Questionnaire
Nm	Newton-meter
AOF	Ankle-Foot Orthosis
MCID	Minimal Clinically Important Difference
MDD	Minimal Detectable Difference
SEM	Standard Error of Measurement
ICC	Intraclass Correlation Coefficient
ROC	Receiver Operating Characteristic curve
AUC	Area Under the Curve

5. OBJECTIVES

5.1 Overall objective

The overall objective of this thesis was to study the psychometric properties of a broad spectrum of clinically relevant outcome measures in people with DM1 (PwDM1), and to investigate the physical activity level and predictors of physical activity level in PwDM1. The outcome measure scope is in preparation for interventional trials to evaluate disease-modifying drugs and symptomatic therapies in PwDM1. The physical activity scope serves as a guideline of the actual physical activity level and predictors for this, which helps the clinician to enhance the physical activity level in PwDM1. All, in order to help improve and sustain physical function in PwDM1.

5.2 Specific objectives

Study I:	To study criterion- and construct validity of muscle strength, balance and functional mobility measurements in PwDM1
Study I:	To study intra-rater reliability of muscle strength, balance and functional mobility measurements in PwDM1
Study II:	To study responsiveness of muscle strength, balance and functional mobility measurements in PwDM1
Study III:	To study physical activity level and predictors of physical activity level in PwDM1

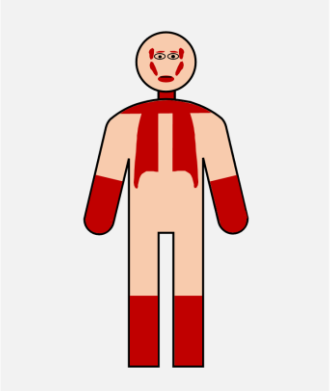
6. INTRODUCTION

6.1 Myotonic Dystrophy type 1

Myotonic Dystrophy type 1 (DM1) is an autosomal, dominantly inherited neuromuscular disease caused by an unstable expansion of the cytosine thymine guanine (CTG) repeat in the dystrophia myotonica protein kinase (DMPK) gene on chromosome 19q13.3 (1).

DM1 is the most common myopathy in adults (2) with a prevalence of 10:100,000 in Europe (3). It is a heterogeneous disease with a phenotypical variability ranging from almost asymptomatic to severely affected (4). In general, the DM1 disease evolves slowly and encompasses muscle affection with myotonia and weakness most pronounced distally in the limbs (5), but the disease also shows extra-muscular, multi-organic dysfunctions in the central nervous system, endocrine systems, respiratory function, eyes (cataract) and heart (6) (Figure 1). Cognitive impairment (7), disease unawareness (8), apathy (9) and fatigue (10) are features of DM1 and should be considered for outcome measures and interventional trials.

Figure 1: Clinical manifestations of muscle weakness in PwDM1



6.2 Studies I + II: Outcome measures

There are currently no curative treatments available for DM1 (11,12), but as disease-modifying clinical trials are in the planning (13), there is a need for valid, reliable and responsive endpoints and biomarkers for disease progression in PwDM1. The establishment of evidence-based outcome measures before execution of interventional trials helps prevent inadequate experimental design and unreliable conclusions, which postpones human application and increases the costs. The target of Studies I-II in this thesis in relation to the overall research process is visualized in Figure 2.

Figure 2: Path from research to clinical practice

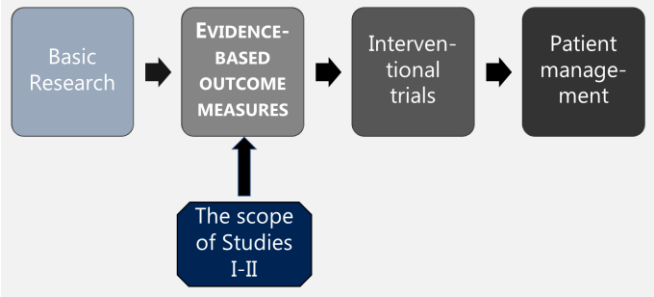
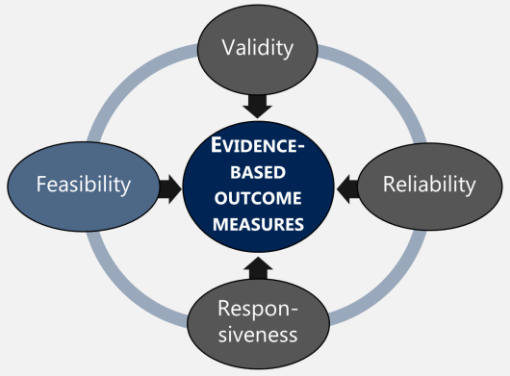
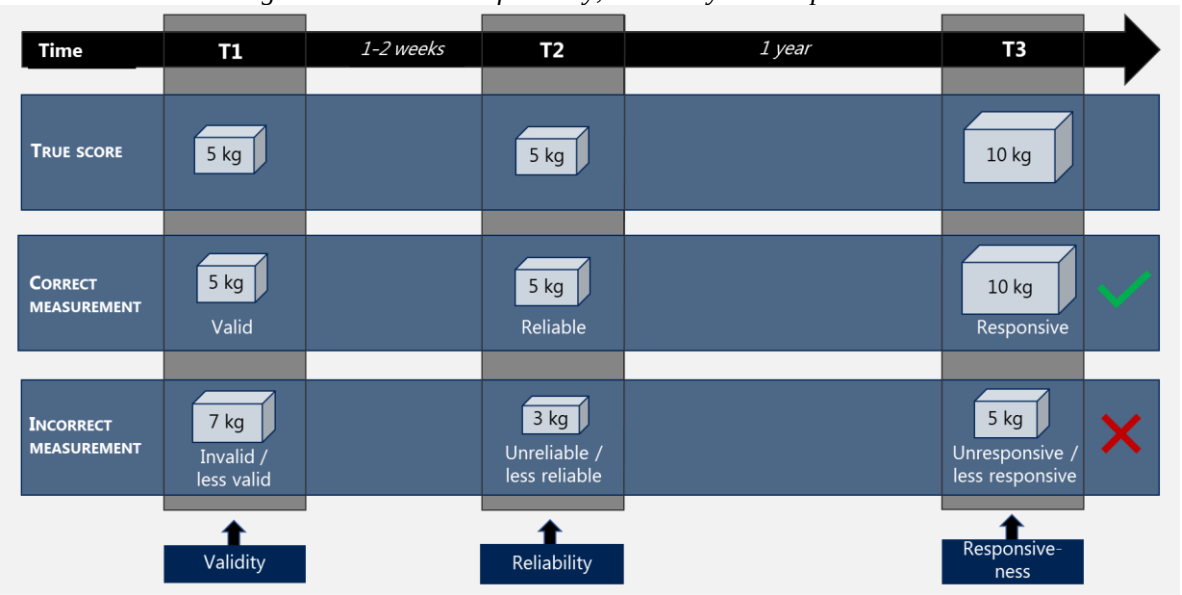


Figure 3: Evidence-based components



Evidence of outcome measures is established by the following psychometric properties of measurements: (I) reliability (II) validity (III) responsiveness to detect change, and (IV) feasibility (Figure 3). Reliability refers to a tool's quality of measuring consistently and faultlessly (14), which is illustrated in Figure 4 (column T2).

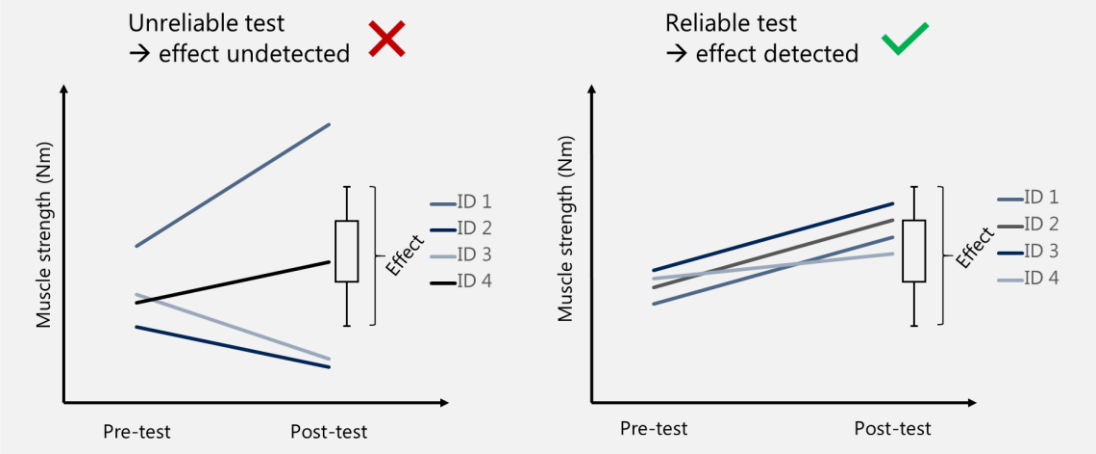
Figure 4: Illustration of validity, reliability and responsiveness



A high degree of reliability is important to have confidence in retest results and to capture real changes. In Figure 5, it is simplified how interventional effects can be masked using unreliable outcome measures in

contrast to the application of reliable outcome measures.

Figure 5: Illustration of reliability as to capturing interventional effects



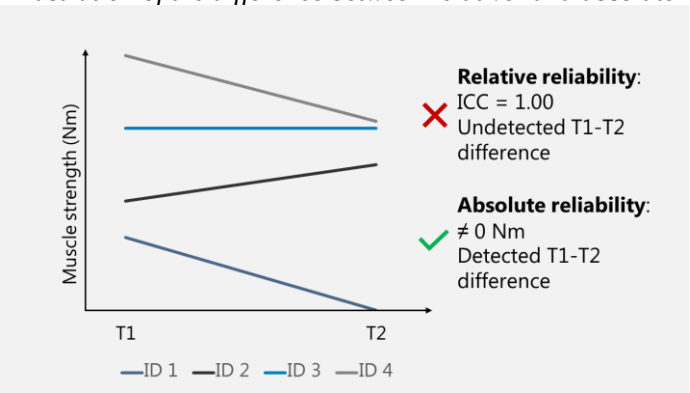
An observed score consists of true variance and error variance (15), and the error component is partitioned into systematic- and random error (15):

$$R = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_{se}^2 + \sigma_{re}^2}$$

R = reliability coefficient, σ_t^2 = true variance, σ_{se}^2 = systematic error variance, σ_{re}^2 = random error variance. Statistical significance testing addresses the systematic error component by investigating the presence of systematic bias in repeated measurements, but is unable to detect random variation (16). Thus, a non-statistically significant difference between repeated measurements rule out systematic measurement error with 95% confidence within a tool, but it does not signify that random error does not exist because individual changes above the mean change is still possible (16). Furthermore, reliability is not considered a dichotomous phenomenon, but rather a continuum (17). Random error is captured by relative reliability statistics and absolute reliability statistics. Intraclass Correlation Coefficient (ICC) is a relative reliability statistical analysis, which states the degree of a tool's reliability, but it has no unit (15). Interpretation and application of relative reliability should be made with caution due to its difficult interpretation, limited practical implications and sensibility towards sample variation bias where ICC is increased with sample heterogeneity (16). Since relative reliability analyzes if the rank of individuals' scores in a group changes from test to retest (15), scores may have changed between repeated measurements within subjects, suggesting measurement error, but still not impair the ICC value as long as the change/measurement error is too small to shift the position of each subject in the group. Because of this potential risk of undiscovered genuine measurement errors, relative reliability statistic is considered legitimate only as a supplemental reliability analysis (16). Random error is detected more accurately with absolute reliability statistic. This statistic addresses the consistency of subjects' scores between repeated tests (15,18), thus accounts for any change/measurement error regardless of the position within the group. Absolute reliability can be calculated for a group using Standard Error of Measurement (SEM) or for a single individual utilizing Minimal Detectable Difference (MDD). The difference between relative- and absolute reliability is exemplified in Figure 6. Figure 6 illustrates measurement error because the

subjects muscle strength has changed from T1 to T2 within a short timeframe where true change is unlikely. The measurement error is not captured by relative reliability statistics with an excellent ICC score (ICC = 1.00) because the rank of subjects (ID 1 to ID 4) is unchanged at T2, despite different muscle strength-values at T2. In contrast, absolute reliability (SEM or MDD) captures the difference in muscle strength values between T1 and T2. Thus, it is exemplified that relative reliability measures the consistency of rank of subjects, whereas absolute reliability measures the consistency of values (15). Moreover, absolute reliability expresses measurement error in absolute values (e.g. seconds), which makes it easily understandable and feasible for practical implications.

Figure 6: Illustration of the difference between relative- and absolute reliability



Validity is defined as a tool's quality of accuracy and measuring what it is supposed to measure (19) (Figure 4, column T1). Different approaches of validity exist. Criterion-related validity is the degree of agreement between methods (target test and criterion/gold standard) that measure the same variable in the same units, and it is the most objective tool to define validity (19). However, a direct comparable tool does often not exist for a specific method, thus validity is obliged to be illustrated by construct validity. Construct validity establishes a method's ability to measure a construct or concept (19), in practice by a method's ability to predict results in another method that assesses the same theoretical concept.

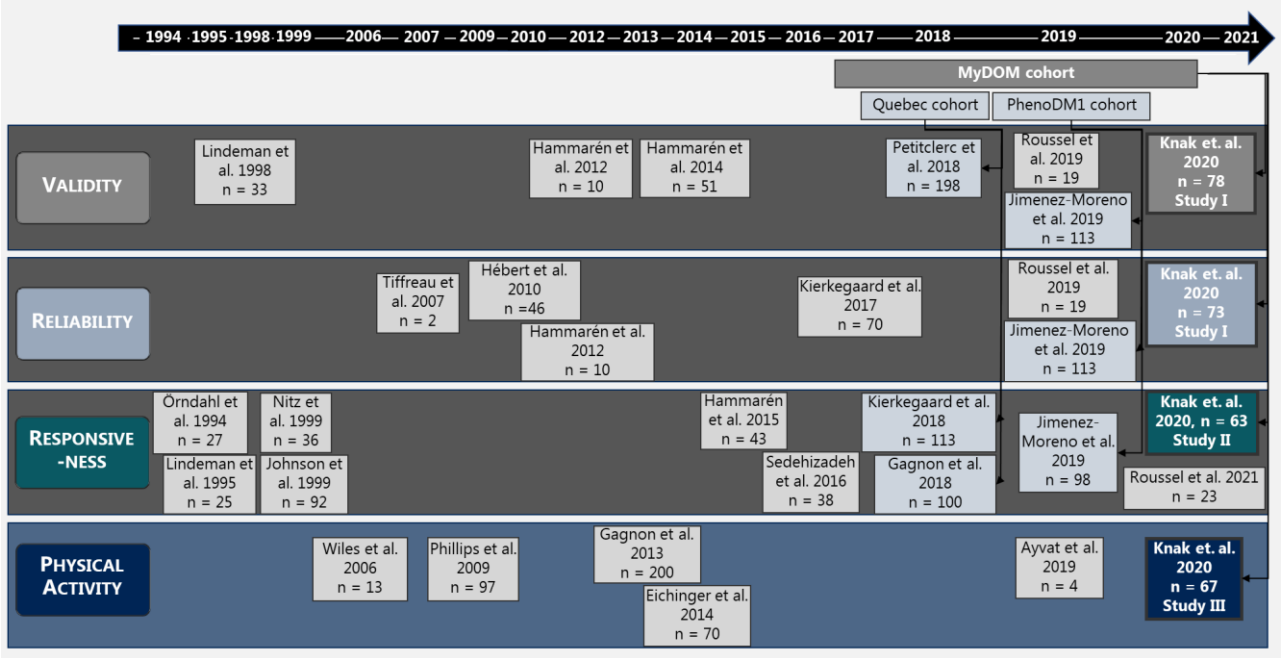
Responsiveness is based on the quality of a method to capture genuine altered conditions (19) (Figure 4, column T3). Various statistical indices exist to measure this concept. The distribution-based approach for responsiveness expresses the magnitude of change in a group at follow-up, and one of the tools for this purpose is statistical significance testing (20). The anchor-based approach investigates the degree of change according to a criterion/anchor, and for this purpose, global rating of change is often applied (20). Global rating of change is the individual's perception of whether a certain clinical condition has improved, deteriorated or is unchanged at follow-up (20). In general, the anchor-based approach is preferred over the distribution-based approach, because the anchor-based approach encompasses a criterion of what is considered important and it assesses the validity of change and not only significance of change. However, if the applied criterion in the anchor-based approach is a poor measurement, conclusions are problematic. In addition to the psychometric concepts, the measurements' feasibility should also be considered (Figure 3). Feasibility encompasses a method's time-efficiency, easiness to conduct and inexpensiveness (3).

There is consensus that psychometric properties of a wide spectrum of outcome measures in PwDM1 should

be established prior to interventional trials, thus methodological studies in larger DM1 samples from a larger research project in Quebec Canada (21–23) and other samples (3,24) have progressed during recent years. However, the evidence of several outcome measures is still not thoroughly established.

With the above methodological knowledge, the existing literature on validity, reliability and responsiveness for muscle strength, balance and functional mobility measurements in PwDM1 will be summarized below. This thesis will address the absent knowledge within this area and enhance the credibility of the currently available methodological studies. A continuum of published methodological studies is shown in Figure 7.

Figure 7: Timeline of studies of outcome measures and physical activity in PwDM1



6.2.1 Muscle strength measurements

Even though stationary dynamometry is considered gold standard within muscle strength measurement and muscle weakness is a cardinal symptom in PwDM1, the psychometric evidence of stationary dynamometry in PwDM1 is weak. This is due to only a small study of high methodological quality regarding validity (25) and reliability (25) in PwDM1, and unspecified DM types in the responsiveness studies (26,27). Moreover, only the knee muscles of the lower limbs have been investigated and documentation of responsiveness is limited to isokinetic muscle strength in PwDM1. For more details, see Supplementals (Supplementals, Table 1).

The easier to use Hand-Held Dynamometry (HHD) device has been the subject of more investigations (see Supplementals, Table 1), which probably is due to its easier application in clinical practice. The psychometric evidence of this device in PwDM1 ranges from weak for validity due to a small study of high methodological quality (25), moderate evidence for intra-rater reliability based on a moderate sample size of high quality (28) to strong evidence for responsiveness due to large cohorts (22,23) (Supplementals, Table 1). The high methodological quality of validity and reliability is restricted to few muscle groups of the lower limbs in PwDM1 and responsiveness has yet to be investigated for a 1-year follow-up period which is the typically

duration of clinical trials.

6.2.2 Balance measurements

Methodological specifications on previously published studies of balance are presented in Supplementals, Table 1. In recent years, focus on balance in PwDM1 has emerged, predominantly concerning dynamic balance. Hammarén et al. (2,29,30) have provided psychometric data on step test in PwDM1 ranging from weak intra-rater reliability evidence to moderate validity and responsiveness evidence based on sample sizes. Responsiveness of step test in PwDM1 has not been reported for a shorter follow-up period. The psychometric evidence of the widely applied Timed-Up-and-Go test (TUG) in PwDM1 is weak for intra-rater reliability and strong for validity and responsiveness based on sample sizes and certainty of statistical approaches (Supplementals, Table 1). Short-term follow-up is still lacking for TUG in PwDM1.
For static balance, only one small study (29) has been conducted in PwDM1. Thus, the psychometric evidence of feet-together stance in this patient group is weak for validity and intra-rater reliability and absent for responsiveness. For tandem stance and one-leg stance, only weak evidence of intra-rater reliability exists, and the evidence of validity and responsiveness in PwDM1 is unknown. The above-mentioned balance measurements are feasible, but they lack information on the quality of balance (e.g. unsteady stand). The information on small postural sway changes can instead be captured by a balance platform. However, no methodological studies on modified Clinical Test of Sensory and Integration and Balance (m-CTSIB) on a balance platform has been conducted in PwDM1.

6.2.3 Functional mobility measurements

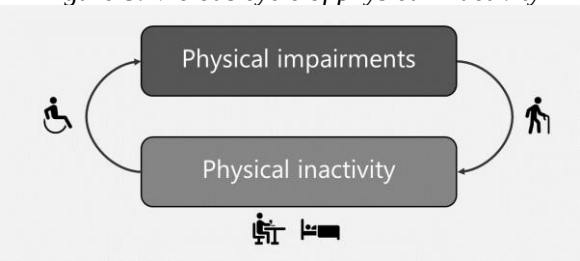
Methodological studies of 10-meter Walk Test (10mWT) (3,21,24,29–31) have provided weak intra-rater reliability evidence and strong evidence of construct validity and responsiveness. Thus, this thesis will provide stronger intra-rater reliability and strengthen the credibility of responsiveness and validity. The psychometric evidence of Sit-To-Stand test (STS) is absent for 10-times STS concerning validity, intra-rater reliability and responsiveness, and the evidence of 30-second STS is weak for validity based on statistical analysis and strong for responsiveness. For more details on the existing literature within functional mobility see Supplementals (Supplementals, Table 1).

6.3 Study III: Physical activity

Physical activity is defined as all movements (32), including everyday activities such as house-cleaning and bicycle transportation, and is not restricted to structured exercise in the gym. There is strong evidence of benefits of regular physical activity in the general population (32). This includes weight maintenance and prevention of lifestyle diseases (32). Not only does a sedentary lifestyle affect the individual, it also has economic consequences for the society (33).
It has been shown in healthy persons that physical inactivity causes loss of muscle strength (34), and a study in PwDM1 (35) found a positive relationship (but not causality) between physical activity and muscle strength. Furthermore, it is likely that muscle wasting caused by physical inactivity further decreases the physical activity level and a vicious cycle may be created (Figure 8). Since one of the cardinal features of DM1 is

muscular atrophy, the impact of disuse atrophy may be even more pronounced in PwDM1 (34) leading to additional physical impairment in these people.

Figure 8: Vicious cycle of physical inactivity



The WHO recommendations on physical activity in 18-64 year old adults can also serve as guideline for persons with disabilities (32). Albeit high level evidence of exercise in PwDM1 is lacking (25,36–39) and has proven problematic to achieve due to the rarity of the disease, the currently available literature on exercise therapies seems promising as they do not seem to generate harmful effects (26,40–46) and some benefits of exercise has been suggested (30,37,42–51). Given the WHO recommendations and the encouraging literature on exercise in PwDM1, physical activity is also important in the DM1 population.

Physical inactivity has been reported in PwDM1 (52–54). These results (52–54), however, are based on subjective questionnaires, which may overestimate the actual physical activity level due to inherent social desirability- and recall bias. The decreased physical activity level in PwDM1 has been supported by objective measurements in a combined cohort of 40 individuals with muscle diseases (including 4 DM patients, type unspecified) (55) and in 13 PwDM1 (56). Yet, objective measurement of physical activity level in a large cohort of PwDM1 is lacking. A timeline of the existing studies of physical activity is shown in Figure 7. To improve physical activity level in PwDM1, identification of risk factors is needed. At present, no studies have investigated predictors of physical activity level in PwDM1.

7. MATERIALS

For Studies I-III in the present thesis, a large DM1 cohort (MyDOM) was enrolled through November 2017 to September 2019. Out of necessity, the MyDOM cohort was selected from the area of Copenhagen via Rigshospitalet and from the area of Aarhus via Aarhus University Hospital in Denmark. The healthy controls for the physical activity study (Study III) were sampled from the area of Copenhagen, Denmark. The healthy controls were enrolled from public newspaper and Facebook advertisements but due to enrolment difficulties, secondary recruitment was conducted from the hospital staff. Eighty-four PwDM1 agreed to participate in Studies I-III. Seventy-eight PwDM1 completed visit 1 (Study I: validity), 73 PwDM1 completed visit 2 (Study I: intra-rater reliability), 63 PwDM1 completed visit 3 (Study II: responsiveness), and 67 PwDM1 and 39 healthy controls completed the physical activity study (Study III). Enrolment of participants is visualized in Figure 9 (PwDM1) and Figure 10 (healthy controls) (57–59).

Figure 9: Flowchart of sampling process of PwDM1 (Studies I-III)

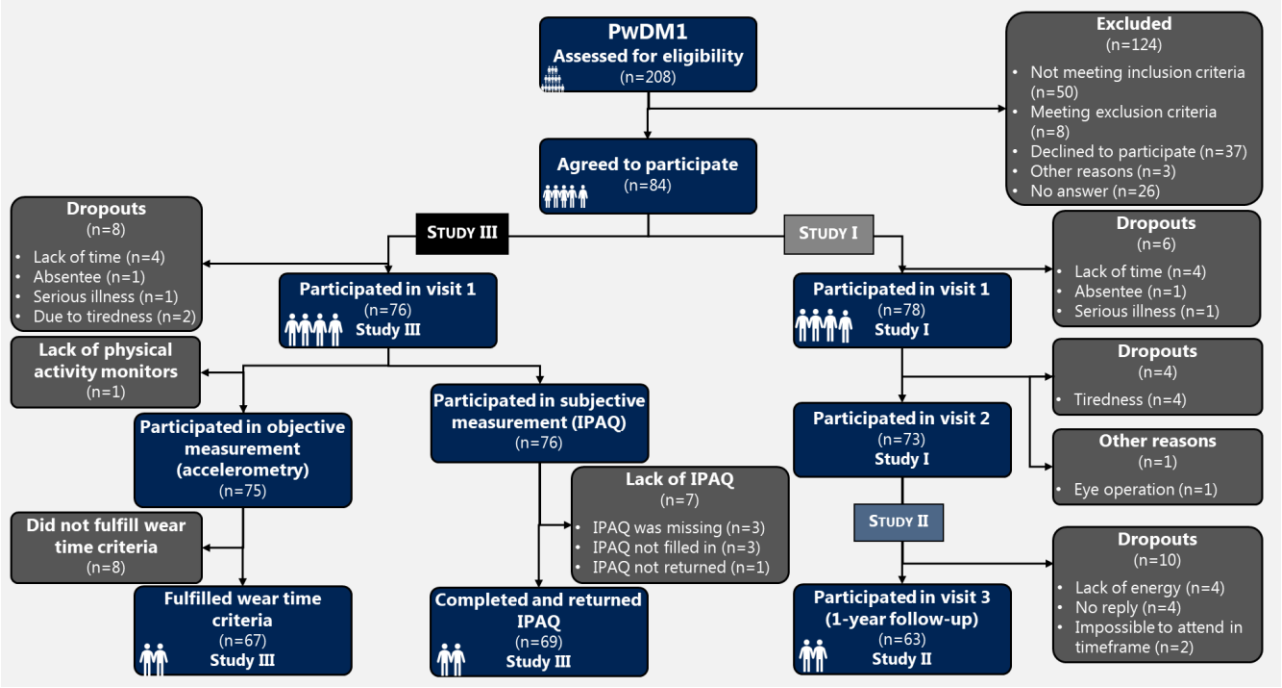
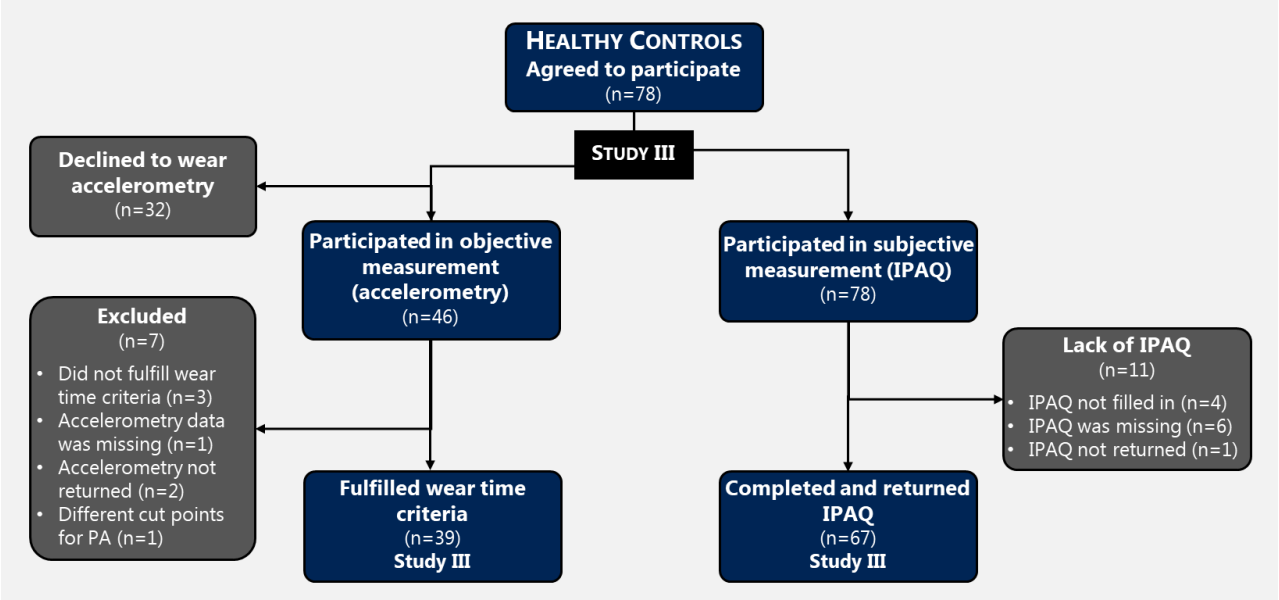


Figure 10: Flowchart of sampling process of healthy controls (Study III)



In- and exclusion criteria are described in Figure 11 for PwDM1 (Studies I-III) and in Figure 12 for healthy controls (Study III) (57–59).

Figure 11: In- and exclusion criteria for PwDM1 (Studies I-III)

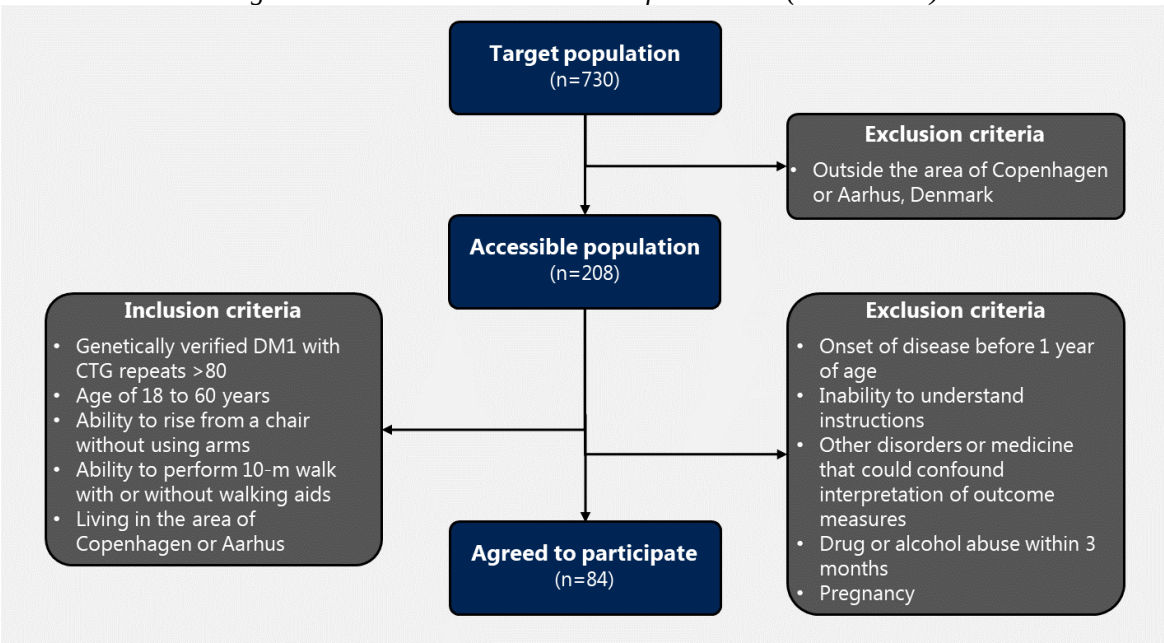
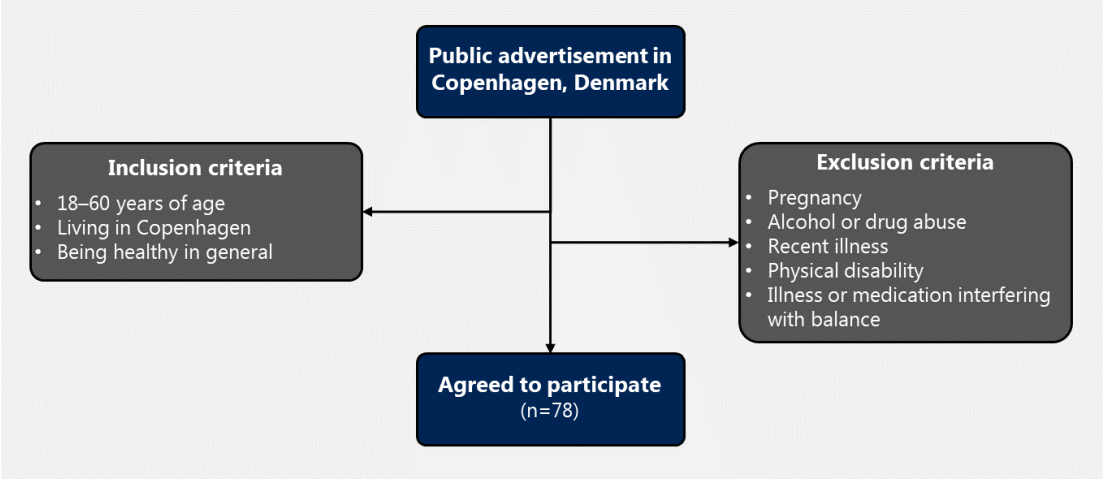


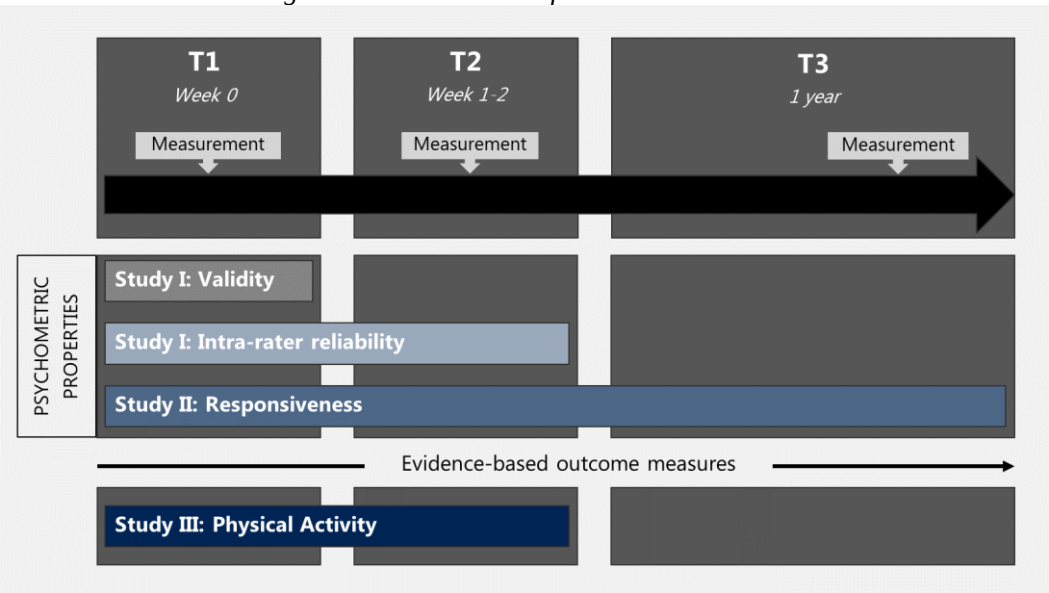
Figure 12: In- and exclusion criteria for healthy controls (Study III)



8. METHODS

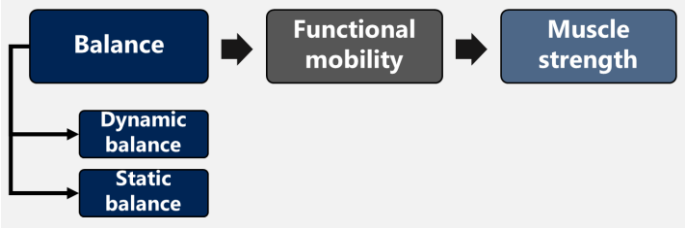
A continuum of the studies in this thesis is shown in Figure 13.

Figure 13: A continuum of studies in this thesis



8.1 Studies I + II: Outcome measures

The 3 visits at T1 (Study I: validity), T2 (Study I: intra-rater reliability) and T3 (Study III: responsiveness) were performed at Rigshospitalet and Aarhus University Hospital. Two raters collected the data. The time difference was 1-2 weeks between T1 and T2 to avoid true change, and 1 year between T2 and T3 to make true change possible but still within the timeframe of clinical trials. For all 3 visits, assessment biases were limited by: (I) same time of day measurements to avoid circadian variation, (II) same rater for each patient to avoid variation between raters, (III) same test order for all visits to avoid a shift in motivation- or fatigue for each measurement, (IV) same procedures to avoid variation in performance due to different instructions, (V) same assistive devices to minimize different prerequisites (unless safety required otherwise), (VI) validation of equipment calibration with 2 weeks interval, and (VII) no checking of previous scores before execution of assessments to avoid confounding of the participant's motivation and the rater's scoring (57). The patients were prior to assessments requested to: (I) wear comfortable, flat, closed shoes to avoid influence of inappropriate footwear and (II) restrain from exhausting or uncommon physical activities 24 hours before assessments to cancel out exercise-induced muscle soreness or fatigue. Due to the possibility of cognitive impairment in PwDM1, verbal instructions were short and accompanied by nonverbal illustrations. Potential questions were answered before test execution. The measurement categories were performed in the following order to reduce the risk of falls due to fatigue:



Block-randomization (excel) was conducted of the measurements within the 3 measurement categories (balance, functional mobility and muscle strength) to eliminate systematic bias; that is fatigue and demotivation are more plausible in the measurements at the end of the test battery. For unilateral measurements, the dominant leg was assessed (the leg used for kicking a ball). The measurement protocol specific to each measurement category is described below.

8.1.1 Muscle strength measurements

Maximal voluntary isometric muscle strength was tested in the lower limb flexor- and extensor muscle groups using HHD (microFET2; Hoggan Scientific, LLC, Salt Lake City, UT) and stationary dynamometry (Biodex System 3 and 4 PRO; Biodex Medical Systems, Upton, NY) (57). Test positions, instructions, encouragement, number of trials, contraction- and relaxation time and muscle strength unit were similar for both devices to ensure comparability of the two muscle strength devices (57). Muscle strength was registered as torque (Newton-meter, Nm), which is automatically generated by the stationary dynamometry. To convert the Newton (N) output from the HHD to torque, the lever arm was measured and the following equation of torque was calculated (57):

$$Nm = N * m$$

For both muscle strength devices, three trials (100% of maximum strength) were conducted after two practice trials (50% of maximum strength). Standardized, verbal encouragement was conducted for the three maximum trials (“push, push, push”). The protocol is detailly described in Figure 14, the test positions are provided in Figure 15+16, and the test positions are visualized in Figure 17 (57).

Figure 14: Protocol specifications for muscle strength measurements

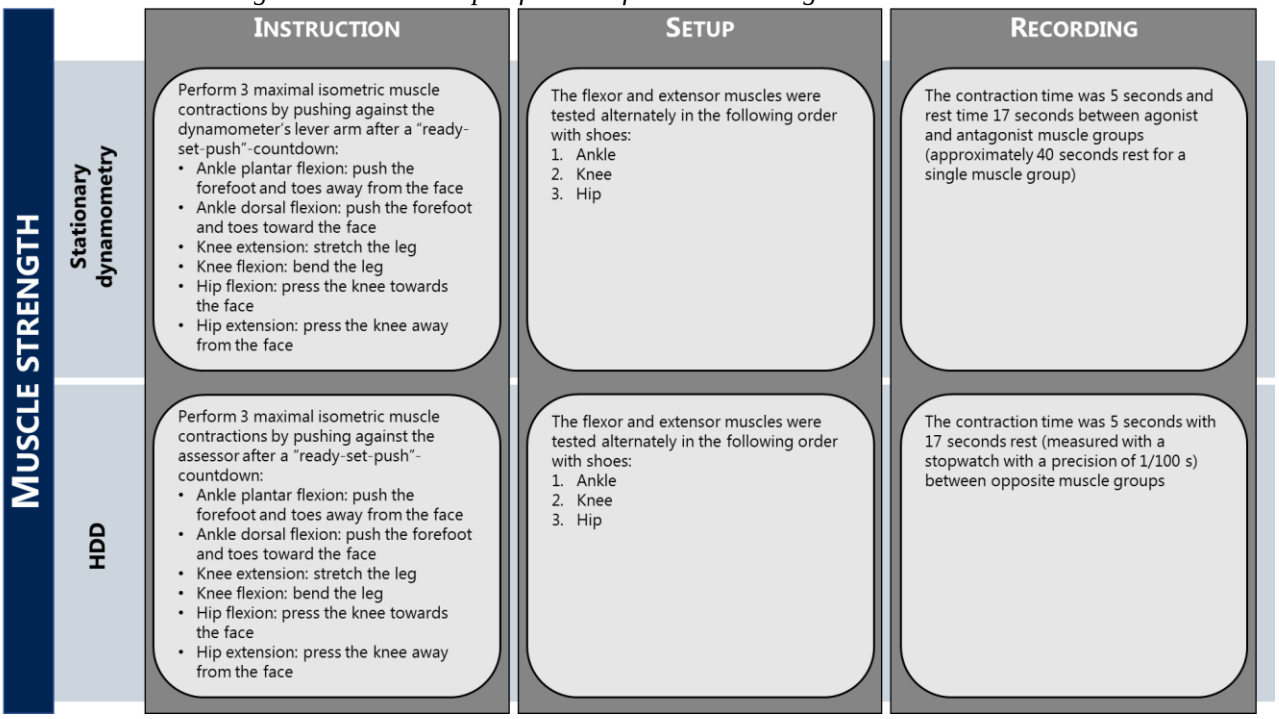


Figure 15: Test positions for stationary dynamometry

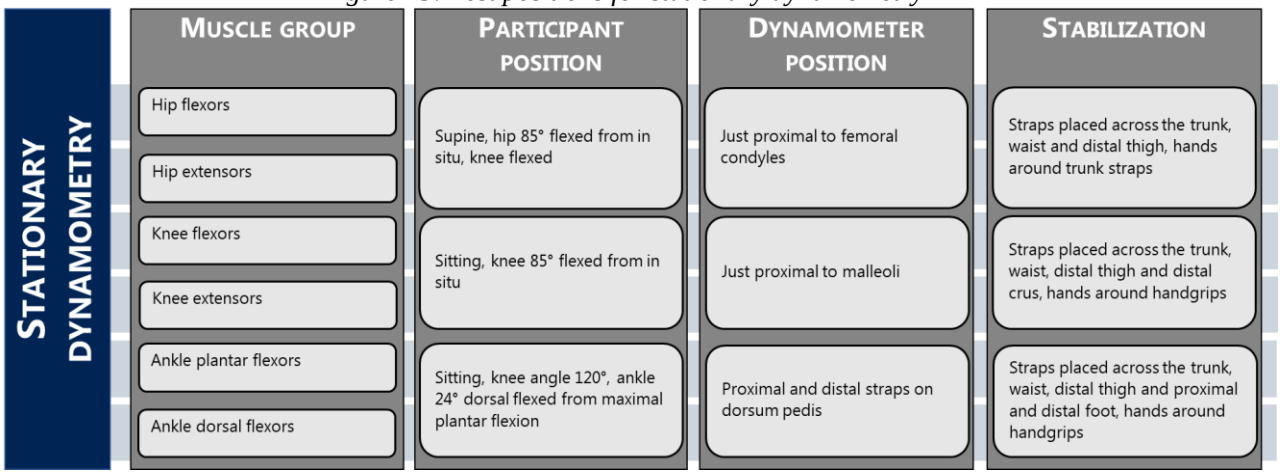
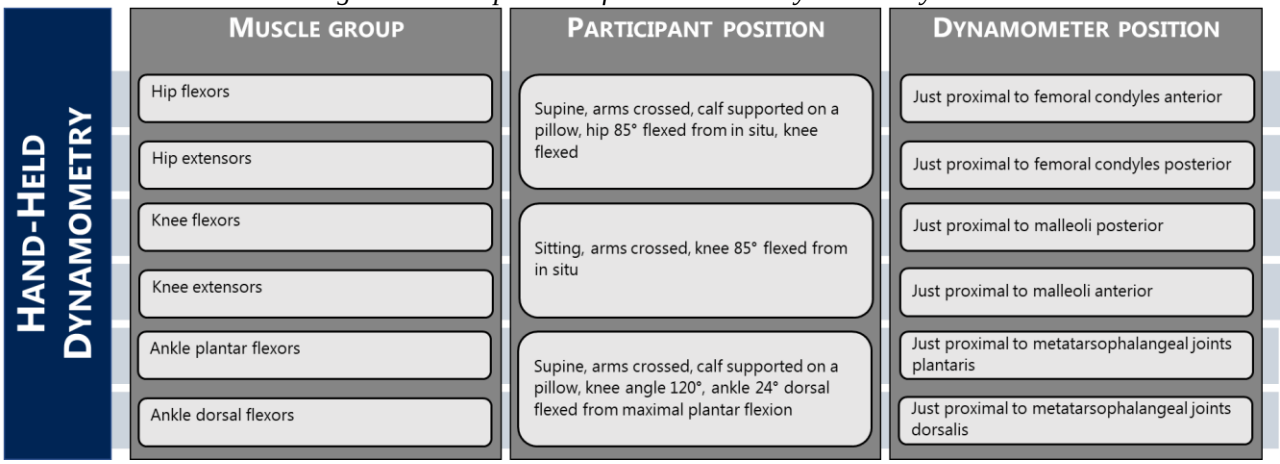


Figure 16: Test positions for Hand-Held Dynamometry



The degrees were visually estimated.

Figure 17: Visualization of test positions for stationary dynamometry and Hand-Held Dynamometry

The images are from Knak et al. (57).

8.1.2 Balance measurements

Both dynamic and static balance were assessed. The dynamic balance measurements consisted of TUG and step test, and habitual walking aids were allowed for the TUG. The static balance measurements included feet-together stance, tandem stance, one-leg-stance (eyes open and eyes closed), and m-CTSIB on a balance platform (Biosway Portable Balance System 950-460, Biodex Medical Systems, NY). Stance of ≥ 30 seconds in one-leg-stance eyes open test was required to qualify for the one-leg-stance eyes closed test (29). The Biosway balance platform is a platform on top of four strain gauges mounted to a frame (60). The m-CTSIB measures sway angle in degrees derived from the center of pressure (COP) and the center of mass (COM) with a sampling rate of 20 Hz (60). The COP is defined as the projection of the individual's COM on the platform, and the COM is calculated by the following equation (60):

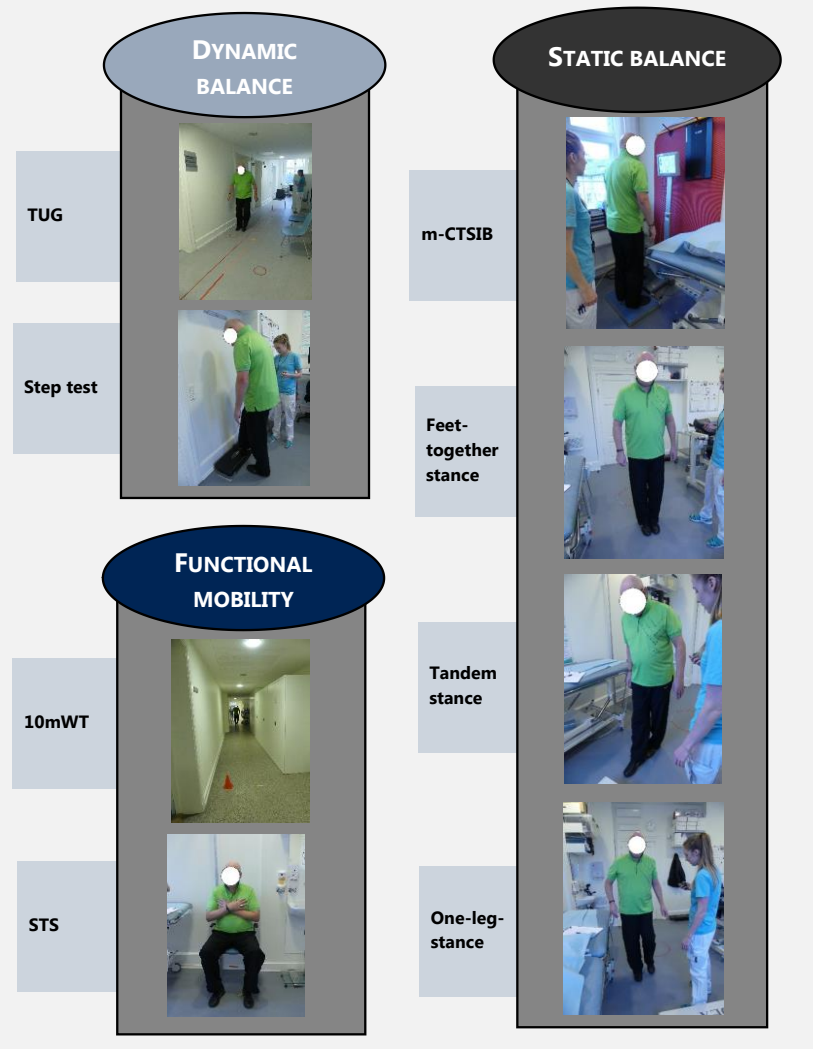
$$COM = 0.55 * patient\ height\ (inches)$$

The output of the m-CTSIB test is reported as sway index, which is defined as the standard deviation of the mean sway angle (60). The higher sway index score, the more unsteady stance during testing (60). Ankle-Foot Orthosis (AFO) and insoles were permitted for the static balance measurements. For all balance measurements, two trials were performed without practice and encouragement. Methodological specifications are presented in Figure 18 and visual demonstration in Figure 19 (57).

Figure 18: Protocol specifications for balance measurements

		INSTRUCTION	SETUP	RECORDING
BALANCE	TUG	Stand up from a chair, walk three meters, turn around the taped mark, walk back and sit down on the chair in a comfortable, safe pace	• Chair height: 45 cm • The starting position was with the back against the backrest and both arms on armrests	The stopwatch (recording 100th of a second) was started when the subject initiated movement and stopped when the subject's rear touched the chair
	Step test	Stand unsupported with feet parallel and place the foot of the dominant leg on the step bench and down on the floor as many times as possible in 15 seconds	• Step bench height: 10 cm • Stand position approximately 5 cm from the step bench • Definition of one step: foot moving from the floor, on to the step bench, and back down on the floor	The stopwatch (recording 100th of a second) was started when the foot began to lift from the floor and stopped after 15 seconds or if the subject needed support or the foot on the standing leg moved
	m-CTSIB	• Stand as quiet as possible with feet hip width apart, arms alongside the body, and look straight ahead in four different conditions: 1. Firm surface, eyes open 2. Firm surface, eyes closed 3. Foam surface, eyes open 4. Foam surface, eyes closed • The subject was informed of each condition before execution • Each condition was initiated by "3-2-1-go"	• Stand on a balance platform • Cursor off	Duration of each condition was 20 seconds with a 10 seconds rest between each condition
	Feet-together stance	Stand as long as possible with feet together and arms alongside the body with eyes open	NA	The stopwatch (recording 100th of a second) was started when the subject was ready and stood unsupported in the correct position and stopped after 60 seconds or if the subject lost balance (defined as a step or arm elevation above approximately 45 degrees)
	Tandem stance	Stand as long as possible with the dominant foot in front of the non-dominant foot so the toes of the rear foot touched the heel of the front foot. Arms alongside the body and eyes open	NA	The stopwatch (recording 100th of a second) was started when the subject was ready and stood unsupported in the correct position and stopped after 40 seconds or if the subject lost balance (defined as a step or arm elevation above approximately 45 degrees)
	One-leg stance eyes open and eyes closed	Stand as long as possible on the dominant leg with eyes open. Raise the non-dominant leg from the floor without touching the standing leg. Arms alongside the body	• If the subject maintained balance for at least 30 seconds with eyes open, the same test position was repeated with eyes closed • Each trial was separated by a 15-seconds rest to avoid muscle fatigue	

Figure 19: Visualization of balance and functional mobility measurements



The images are from Knak et al. (57).

8.1.3 Functional mobility measurements

Functional mobility was assessed by the 10mWT (fast pace) and the 10-times STS (see Figure 19+20) (57). Habitual walking aids were allowed for the 10mWT. All functional mobility measurements were tested twice. Neither practice trials nor encouragement were offered.

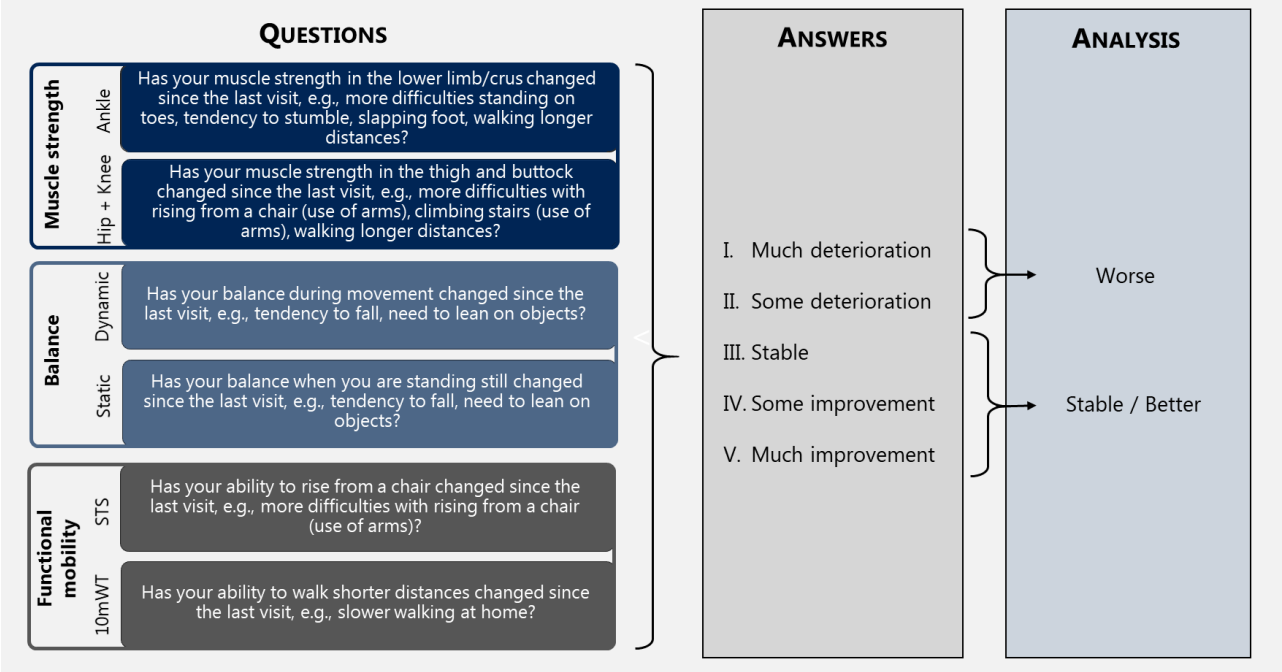
Figure 20: Protocol specifications for functional mobility measurements

FUNCTIONAL MOBILITY	INSTRUCTION	SETUP	RECORDING
10mWT	Walk as fast as possible, yet safely, on a 13-meter walkway on "ready-set-go"	The 10mWT was performed on a 13-meters walkway with 1.5-meters for acceleration and 1.5-meters for deceleration. The start- and finish lines were marked with tape and cones	The stopwatch (recording 100th of a second) was started when the subject's trunk passed the 1.5-meters mark and stopped when the subject's trunk passed the 11.5-meters mark
10-times STS	Stand up from the chair in a fully upright stance and sit down on the chair 10 times as fast as possible with arms crossed in front of the chest	Chair height: 45 cm	The stopwatch (precision of 100th of a second) was started when the subject's back left the backrest and stopped when the rear touched the chair

8.1.4 Study II: Questionnaires

At 1-year follow-up, the PwDM1 filled in a subjective Global Rating Scale (GRS) questionnaire, which reflected their own perception of whether their physical condition was stable, worse or better compared to baseline in relation to the objective measurements of muscle strength, balance and functional mobility (see Figure 21) (23). The short-version International Physical Activity Questionnaire (IPAQ) was administered at baseline and follow-up. This was done to register their self-reported physical activity level and to account for a potential shift in physical activity level at follow-up.

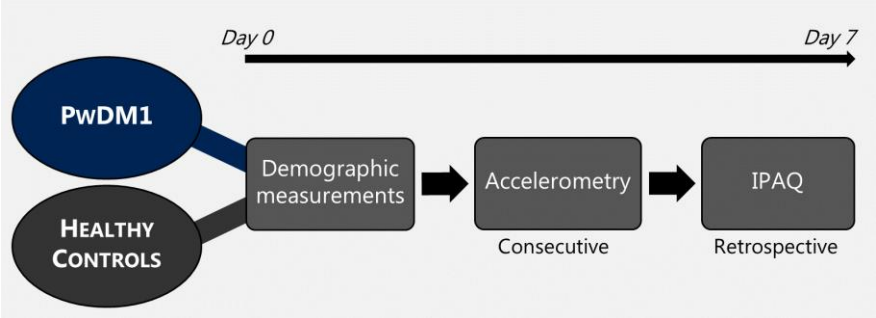
Figure 21: Global Rating Scale



8.2 Study III: Physical activity

The design to study physical activity level in PwDM1 and in a healthy control group is illustrated in Figure 22.

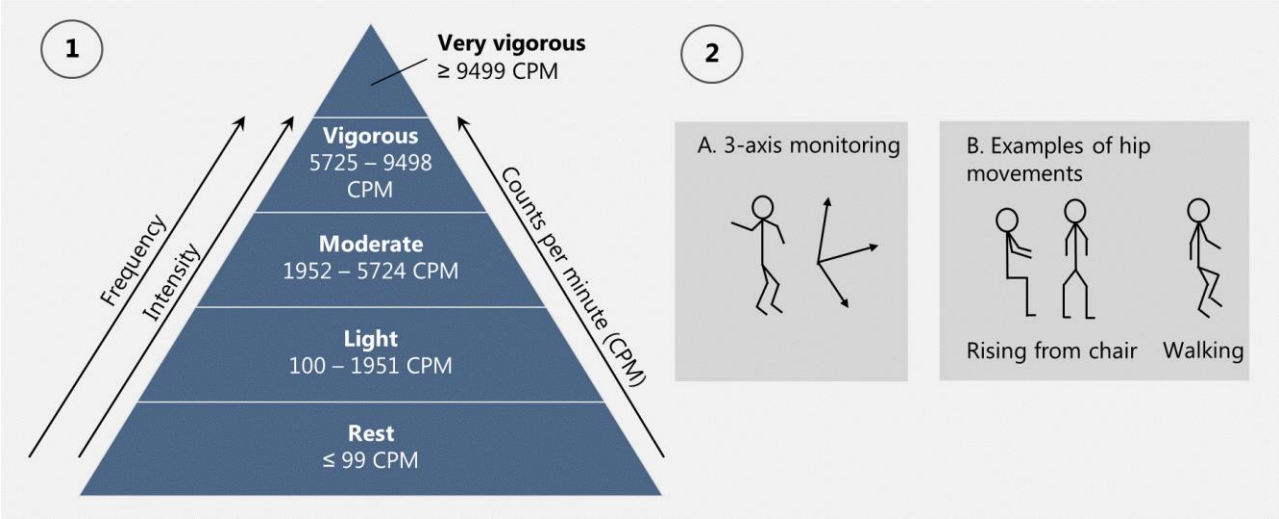
Figure 22: Design of the physical activity study (Study III)



Information was recorded about age, marital status (cohabitant or not), education (none, elementary school, high school, college with three different education lengths, postgraduate), apathy (Apathy Evaluation Scale-Self-rated, AES-S), fatigue (Fatigue Severity Scale-7 items, FSS-7), and ankle dorsal flexor muscle strength, because these variables are recognized as possible contributors to physical activity (59). Stationary dynamometry (Biodex System 3 and 4 PRO; Biodex Medical Systems, Upton, NY) was applied to assess isometric ankle dorsal flexor muscle strength, because our validity findings demonstrated a superior validity of

the stationary dynamometry compared to the HHD (although the validity was equal for the ankle dorsal flexors). An accelerometer (wGT3X-BT, Timik Medical, Herlev, Denmark) worn at the hips was applied to record physical activity level objectively in the PwDM1 and the healthy controls (59). To obtain a representative picture of the physical activity level across weekdays, the accelerometer was worn for 7 consecutive days 24 hours a day. Removal was only instructed for contact with water. The accelerometer device monitors 3-axis motion of the hips (up-down, left-right, forward-back) as to intensity and duration of movements (Figure 23 [2]) (59). Figure 22 [1] visualizes the applied predefined intensity categories (61). Zero acceleration of ≥ 1 hour was recorded as non-wear time (62), and a minimum of at least 4 days daytime wearing was required to be included in analyses (59). The short-version IPAQ was administered on the last day (day 7) of accelerometer monitoring to study the participants subjective perception of their physical activity level for the retrospective 7 days (59).

Figure 23: Accelerometry specifications on intensity [1] and motion [2]



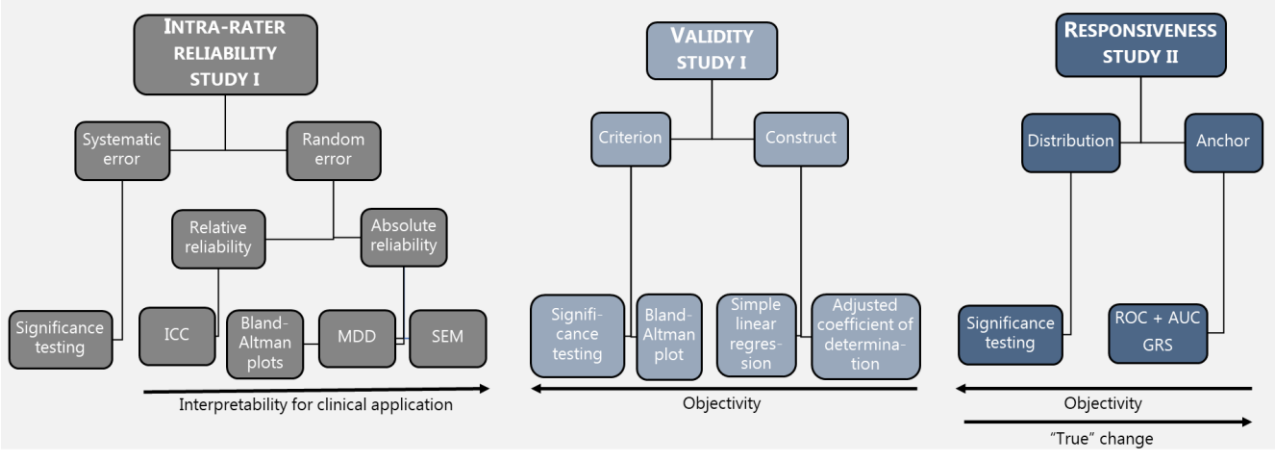
8.3 Statistics

Statistical analyses for Study I was performed using IBM SPSS Statistics 25, and analyses in Studies II+III were conducted with SAS Enterprise Guide 7.1. Statistical significance of $p \leq 0.05$ was determined a priori. If model assumptions for analyses were violated, model assumptions were checked on log-transformed data. Analyses were conducted on the data that fulfilled the model assumptions most properly (raw data or log-transformed data). When log-arithmetic transformation was needed, raw data were log-transformed, analysed and then back-transformed with the equivalent anti-log to interpret the results converted into ratios (percentages). In the presence of genuine outliers, sensitivity analysis was conducted both including and excluding the outliers. Genuine outliers were included because the scores were real and not due to error, but results are also presented without outliers since the influence on the results of few outlier patients is remarkable compared to most of the patients.

8.3.1 Study I

An overview of the statistical analyses related to each psychometric property is presented in Figure 24.

Figure 24: Overview of the statistical analyses for the psychometric properties in studies I-II



First, intra-rater reliability statistics will be addressed. Systematic measurement error was tested with a parametric two-tailed paired-samples t-test (57). Random error was analysed using relative- and absolute statistics. Intraclass Correlation Coefficient, ICC (2,1) (two-way random-effects, absolute agreement, 95% CI, single measurement) was used to assess relative reliability (63):

$$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

MS_R = mean square for rows, MS_E = mean square for error, k = number of measurements, n = number of subjects, MS_C = mean square for columns. The ICC represents a continuum ranging from 0.00 to 1.00, and the following ICC guideline (64) was applied with proviso in the present study:

$$\leq 0.75 = poor\ to\ moderate$$
$$> 0.75 = good$$

Absolute reliability was analysed by different statistical approaches. Agreement was tested by Bland-Altman plots (17,65):

$$S(x,y) = \left(\frac{S_1 + S_2}{2}, S_1 - S_2\right)$$

S_1 = visit 1, S_2 = visit 2. The closer the difference scores between visit 1 and visit 2 are to zero, the lower absolute measurement error between repeated measurements, which indicates higher agreement and higher degree of absolute reliability. Absolute measurement error on a group level was calculated by Standard Error of Measurement (SEM) (15,66):

$$SEM = \frac{SD_{diff}}{\sqrt{2}} \qquad SEM\% = SEM/mean * 100$$

SD_{diff} = standard deviation of visit 1 – visit 2, mean = [(visit 1 + visit 2)/2]. SEM is the value that should be exceeded at retest in a group to indicate a genuine change. The previously applied cut-off of $SEM_{\%} \leq 15\%$,

that has been suggested acceptable in a group, (67) was used as guidance in the present study. Minimal Detectable Difference (MDD) with 95% CI was conducted to estimate absolute measurement error on an individual level (15,66):

$$MDD_{95} = SEM * 1.96 * \sqrt{2} \qquad MDD_{95\%} = \frac{MDD}{mean} * 100$$

Mean = [(visit 1 + visit 2)/2]. MDD is the value that should be exceeded at retest for a single individual to indicate a genuine change. Measurement error of $MDD_{95\%} \leq 30\%$ has previously been considered reasonable for individuals (68) and did, thus, serve as guidance in the present study. In addition, the Minimal Clinically Important Difference (MCID) was estimated to evaluate the MDD scores based on the following guideline (69):

$$MDD > MCID = troublesome$$

MCID is the minimal difference of patient importance (70), and the mathematically concept of this is as follows (71):

$$MCID = \frac{SD_{baseline}}{2} \qquad MCID\% = \frac{MCID}{baseline} * 100$$

In contrast to the ICC, the SEM, MDD and Bland-Altman plot statistics are strengthened by being independent of sample heterogeneity (16). These latter analyses were, therefore, prioritized as the primary reliability analyses in the present study.

Second, validity statistics will be presented. Criterion-related concurrent validity was assessed for the HHD and the stationary dynamometry by paired-samples t-test to determine systematic difference between the devices and by the Bland-Altman plots to visualize both overall and individual agreement between the devices. The equation for conducting the Bland-Altman plots is stated above in the reliability section. The Y-axis was defined as *stationary dynamometry - HHD*. Construct validity was calculated for balance and functional mobility measurements utilizing simple linear regression and adjusted coefficient of determination (R^2) to test the ability of one measurement to predict the output of another measurement which shares similar aspects (57). Secondary, Pearson product-moment correlation coefficient (*r)* was conducted to assess the relationship between the measurements that reflect the same underlying theoretical concept (57). The following correlation coefficient guideline was applied with proviso (72):

$$0.00 - 0.25 = no\ or\ little$$
$$0.25 - 0.50 = fair$$
$$0.50 - 0.75 = moderate\ to\ good$$
$$> 0.75 = good\ to\ excellent$$

Ceiling- and floor effects were investigated to evaluate whether the measurements were unable to capture performances outside the eligible performance range. The previously used definition of more than 15% of all

participants with highest possible scores (ceiling effect) or lowest possible scores (floor effect) (73) was applied as guideline in the present thesis.

It was investigated if there were differences between patients who completed visit 1 and visit 2 versus dropouts by conducting unpaired t-tests for the variables age and body mass index, Mann-Whitney test for Muscular Impairment Rating Scale (MIRS), and by X^2-test of homogeneity for the variable sex (57).

8.3.2 Study II

The statistical analyses are visualized in Figure 24. To measure responsiveness, we used a distribution-based and anchor-based approach. The distribution-based analysis with statistical significance testing was the primary responsiveness analysis in this study. Statistically significant change at follow-up was calculated by linear mixed model with family as a random effect, visit as a covariate and with unstructured covariance (58). This was done to adjust for family members in the cohort, repeated measurements in the same individuals and missing values/dropouts. The results are presented as mean ± SE. Wherever model assumptions were violated, log-arithmetic transformation and back-transformation were applied (58). As in study I, the cut-off of either ceiling- or floor effects was >15% of the patients with either maximum or minimum performance scores. Anchor-based analyses with the subjective GRS as anchor were used secondary to describe responsiveness. Because of unavailable objective anchors and the recognition of the limitations of utilizing subjective anchors, especially in PwDM1 where self-awareness can be limited, the anchor-based approach was used supplementary in this study, although, in theory this approach is superior. Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) were conducted to estimate the target tests' (muscle strength, balance, and functional mobility measurements) ability to correctly register change or no change at follow-up in accordance with the anchor (GRS) (20). The discrimination ability was evaluated from the following guideline (74):

$$0.50 = no$$
$$0.50 - 0.70 = poor$$
$$0.70 - 0.80 = acceptable$$
$$0.80 - 0.90 = excellent$$
$$0.90 - 1.00 = outstanding$$

Both the target tests (muscle strength, balance, and functional mobility measurements) and the anchor (GRS) were dichotomized into worse and stable/better for the ROC and AUC analyses (see Figure 21). The continuous output from the target tests were dichotomized according to the MDD-values established in Study I.

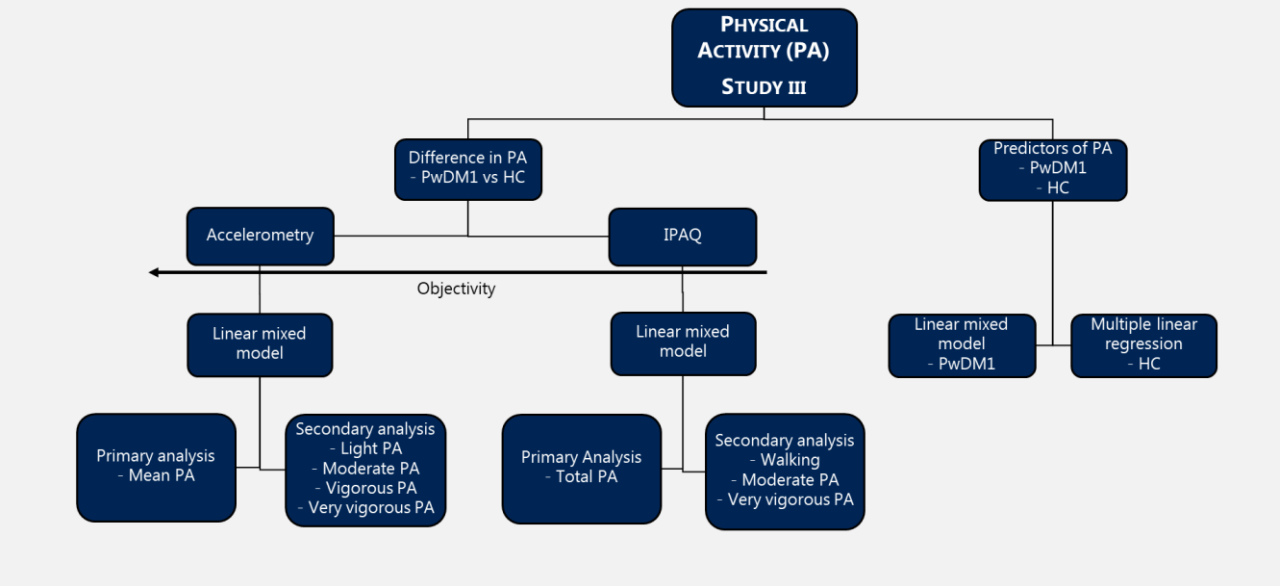
For clinical trials relevance regarding inclusion criteria, sub-analyses of statistical significance testing, ROC and AUC for ankle dorsal muscle strength with stationary dynamometry were done in the PwDM1, excluding those patients who were unable to activate the stationary dynamometry threshold (21% of the PwDM1) (58). For generalizability, it was analyzed whether there was a divergence between the patients who completed the study from the patients who declined to be enrolled in the study and the dropouts by unpaired t-test for continuous data, Mann-Whitney test for ordinal data and Fisher's exact test for dichotomous data (58). To

explore a difference in cognition and apathy between the two groups of patients with either agreement or disagreement between the subjective GRS scores and the objective scores, Mann-Whitney tests was conducted for the six GRS questions (58). These six Mann-Whitney tests were not Bonferroni-corrected, because the tests were only secondary and exploratory (58). The HHD was chosen as the objective outcome of muscle strength for these analyses, because the HHD was more reliable than the stationary dynamometry based on the previous reliability findings.

8.3.3 Study III

In Figure 25, the statistical analyses to investigate physical activity are illustrated.

Figure 25: Overview of the statistical analyses in the physical activity study III of PwDM1 and healthy controls (HC)



A difference in physical activity level (min/week) between PwDM1 and healthy controls was analysed by linear mixed models. The family variable was inserted as random effect to adjust for family members in the DM1 group (59). The variables sex, age, BMI, marital status and education was inserted as covariates to adjust for possible differences on these aspects between the DM1 and healthy control groups (59). The primary confirmatory linear mixed model analyses were conducted on the mean physical activity level ([light intensity + moderate intensity + vigorous intensity + very vigorous intensity] / 4) for accelerometry, and for the total activity level (walking + moderate intensity + very vigorous intensity) for IPAQ. These two primary outcomes were Bonferroni-corrected (p 0.05 / 2 = $p_{Bonferroni}$ 0.025) (59). Secondary linear mixed model analyses were tested on each physical activity level for both accelerometry and IPAQ but were not Bonferroni-corrected due to the exploratory nature, as this would be too conservative (59).

Predictors of physical activity in PwDM1 were analysed by linear mixed model with family as a random effect (correction for family members) and age, marital status, education, apathy, fatigue and ankle dorsal flexor strength as covariates/predictors (59). To analyze predictors of physical activity in healthy controls, multiple linear regression was conducted with age, marital status and education as covariates (there were no family members in this group). To avoid mass significance, the number of predictors/levels for PwDM1 and healthy controls did not exceed the following guideline (recommendation from statistical support) (59):

$$Sample\ size\ /\ 10$$

A continuous variable counts only as one predictor/level, but each score of ordinal data and nominal data counts as one level. Hence, ordinal and nominal data types count as several levels, which are troublesome statistically. Age, BMI and ankle dorsal muscle strength are naturally continuous variables, but to avoid mass predictors/levels, the variables education, apathy and fatigue were considered "continuous" since a rank order exists, and this is a common statistical maneuver. However, marital status cannot be considered as ranked data and, therefore, this variable counts as several levels. To reduce the number of levels, marital status was dichotomized into cohabitant or not cohabitant.

It was investigated whether demographic data on subjects who adhered to wear time of accelerometry differed from the subjects who did not adhere using unpaired t-test (continuous variables), Mann-Whitney test (ordinal variables) and Fisher's exact test (dichotomous variables) (59).

8.4 Ethics

Studies I-III are approved by the Regional Committee on Health Research Ethics in Denmark (H-17017556), and informed written consent was obtained from all participants (PwDM1 and healthy controls) (57–59).

9. RESULTS

9.1 Participants

Demographic data for PwDM1 and healthy controls are provided in Table 1 and missing values for demographic data are available in Supplementals (Supplementals, Table 2) for Studies I-III (57–59).

Table 1: Demographic data for PwDM1 and healthy controls (Studies I-III)

	STUDY I		STUDY II	STUDY III	
	VALIDITY	RELIABILITY	RESPONSIVENESS	PWDM1	HEALTHY CONTROLS
Sex (no.) Women Men	78 39 39	73 36 37	63 30 33	67 32 35	39 21 18
Age (years), mean (SD)	40 (10)	40 (10)	41 (10)	41 (10)	39 (11)
BMI[1], median (IQR)	24 (21-28)	24 (21-27)	24 (21-27)	24.3 (5.1)[2]	23.5 (2.6)[2]
MIRS[3] (no.) Grade 1 Grade 2 Grade 3 Grade 4 Grade 5	 0 17 3 51 7	 0 17 2 47 7	 0 13 2 42 6	NA	NA
Walking aid (no.) Insoles AFO[4] Cane Walker Three-wheeled scooter	 1 8 1 1 1	 1 8 1 1 1	 1 8 1 1 1	NA	NA
AES-S[5], median (IQR) *Apathy (no.)*	12 (9-17) 0	12 (9-17) 0	12 (8-16) 0	12 (9-17) 0	NA
STROOP[6] (T-Scores), median (IQR)					
Word score *Cognitive impairment (no.)*	32 (27-39) 21	32 (27-39) 21	32 (27-37) 24	32 (27-39) 20	NA
Colour score *Cognitive Impairment (no.)*	34 (31-40) 13	34 (31-40) 14	34 (31-40) 11	34 (31-40) 11	NA
Colour-Word score *Cognitive impairment (no.)*	38 (35-46) 2	38 (35-46) 3	37 (34-45) 3	38 (35-45) 1	NA
Interference score *Cognitive Impairment (no.)*	50 (50-51) 0	50 (50-51) 0	50 (50-51) 0	50 (50-51) 0	NA

FSS-7[7], median (IQR) *Abnormal fatigue (no.)*	NA	NA	NA	4 (3-5) 39	NA
Ankle dorsal flexor muscle strength (Nm), median (IQR) *Too weak to activate the test, no.*	NA	NA	NA	17 (8-26) 13	NA
Marital status (no.) Cohabitant No cohabitant	NA	NA	NA	 50 17	 31 8
Accomplished education (no.) None Elementary school Youth education Short-cycle higher education (2-2½ yrs) Medium-cycle higher education (3-4½ yrs) Long-cycle higher education (5-6 yrs) Postgraduate education	NA	NA	NA	 1 10 9 17 16 11 3	 0 1 3 6 13 15 1

Study III: The demographic data are presented for the participants (PwDM1 and HC) who wore accelerometry (objective data) and fulfilled wear time criteria because the objective accelerometry data were applied to calculate predictors of PA for both groups. Age (mean ±SD, 40 ± 8 years for PwDM1 and 38 ± 10 years for HC) and BMI (24.4 ± 4.2 for PwDM1 and 23.2 ± 1.96 for HC)) were similar for the groups of PwDM1 (n=69) and HC (n=67) that completed and returned IPAQ (subjective data).

[1]BMI (Body Mass Index) (kg/m^2), $\frac{weight\ (kg)}{height\ (m)^2}$, BMI <18.5=underweight; BMI 18.5-24.9=normal weight; BMI 25.0-29.5=overweight; BMI 30.0-34.9=obesity.

[2]mean (SD)

[3]MIRS=Muscular Impairment Rating Scale. Grade 1=no muscular impairment; grade 2=minimal signs; grade 3=distal weakness; grade 4=mild to moderate proximal weakness; grade 5=severe proximal weakness.

[4]AFO=Ankle-Foot Orthosis

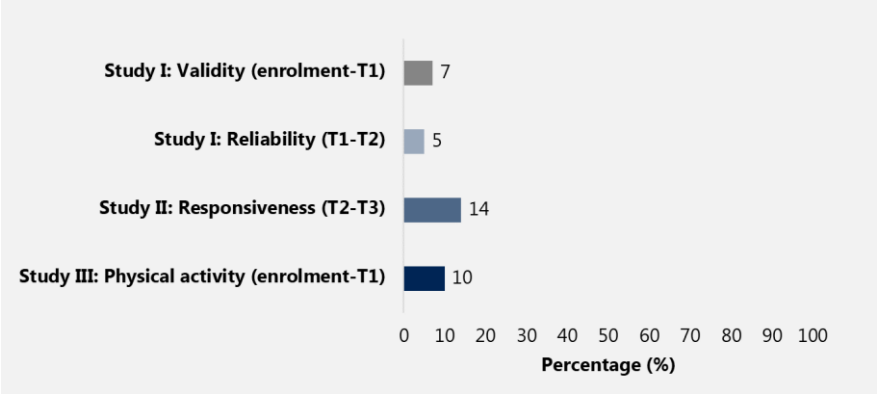
[5]AES-S=Apathy Evaluation Scale (Self-rated). Score >34=apathy

[6]Verbal STROOP Colour and Word Test (Adult version). 95% CI for STROOP normative data: 30.4 to 69.91. For all STROOP scores, higher scores reflect better performance.

[7]FSS-7=Fatigue Severity Scale (7 items)

Figure 26 shows percentage of PwDM1-dropouts in Studies I-III.

Figure 26: Withdrawal of PwDM1 (Studies I-III)



9.1.1 Studies I + II: Outcome measures

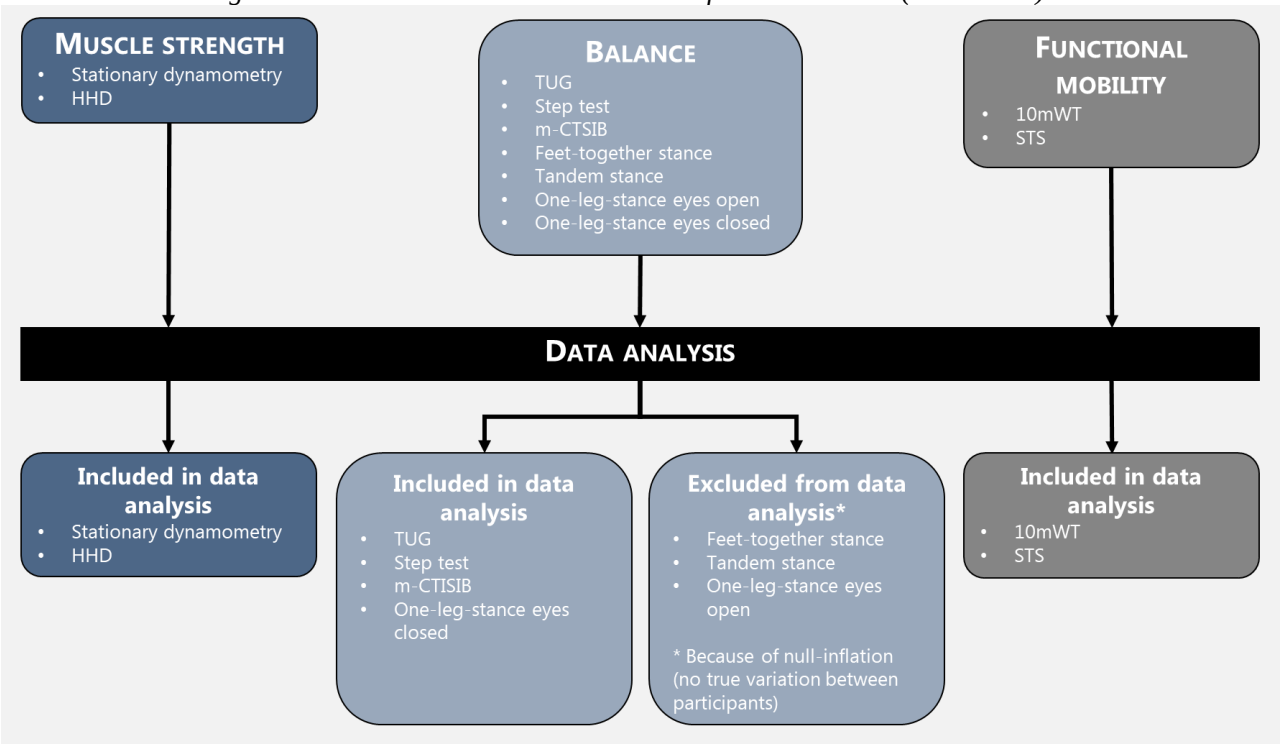
Both samples from the validity study (Study I) and the intra-rater reliability study (Study I) adhered to the methodological sample size recommendation of 50 participants (75,76). There was no difference between PwDM1-dropouts and the PwDM1 who completed visit 2 (Study I: Reliability) nor between PwDM1-dropouts and the PwDM1 who completed visit 3 (Study II: Responsiveness) regarding age (p = 0.08 for visit 2 and p = 0.15 for visit 3), sex (p = 0.61 for visit 2 and p = 0.52 for visit 3), muscle affection (MIRS) (p = 0.48 for visit 2 and p = 0.36 for visit 3) and BMI for the PwDM1 who completed visit 3 (p = 0.74). However, the PwDM1 who completed visit 2 showed lower BMI (median, IQR; 23.6 kg/m^2, 20.7 to 27.5 kg/m^2) compared to the PwDM1-dropouts at visit 2 (32.8 kg/m^2, 28.7 to 36.9 kg/m^2) (p = 0.004) (Study I: Reliability). There was no difference in sex between the PwDM1 who declined to participate in visits 1-3 (Studies I-II) and the PwDM1 who completed visit 2 (Study I: Reliability) (p = 1.00) and visit 3 (Study II: Responsiveness) (p = 0.49). However, the PwDM1 who declined to participate in visits 1-3 (Studies I-II) were older (mean ± SD, 47 ± 9 years) than the PwDM1 who completed visit 2 (40 ± 10 years, p = 0.004) and visit 3 (41 ± 10 years, p = 0.008).

9.2 Outcomes

9.2.1 Studies I + II: Outcome measures

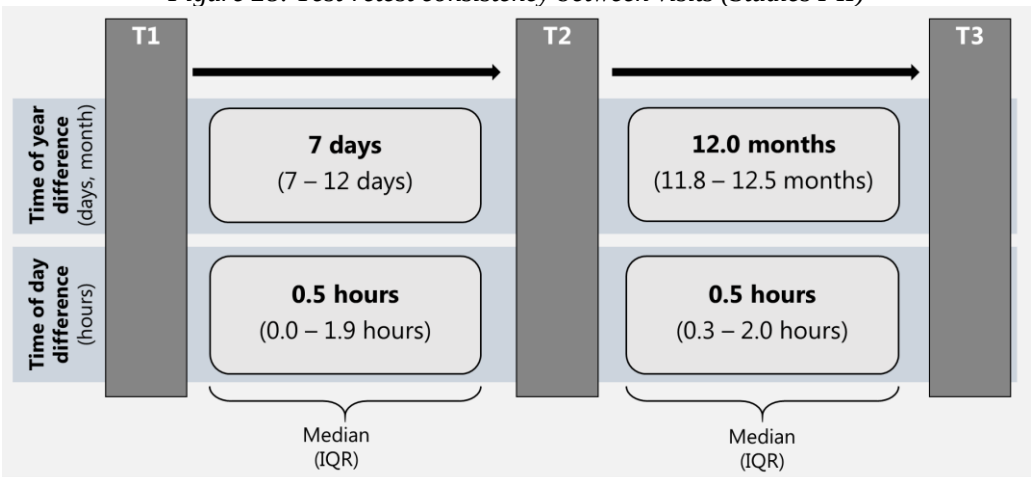
Analyses of all outcome measures were conducted on the best value to obtain the best achieved performance by the participants. Figure 27 provides an overview of the outcome measures that qualified for inference statistics, and a flowchart of PwDM1 included in data analyses is provided in Supplementals (Supplementals, Figure 1) (57,58).

Figure 27: Outcome measures included in inference statistics (Studies I-II)



Missing values for outcome measures are available in Supplementals (Supplementals, Table 3) (57,58) and test-retest consistency between visits (T1-T3) is shown in Figure 28.

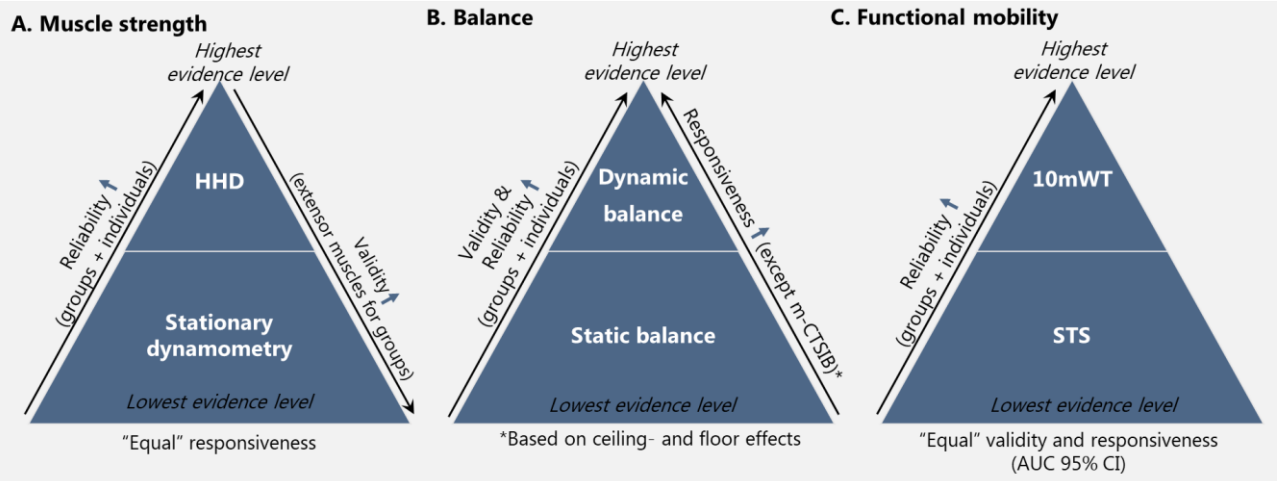
Figure 28: Test-retest consistency between visits (Studies I-II)



Visualization of results has been prioritized in this thesis, but the exact values and absolute values for SEM, MDD and MCID are available in the published articles attached for Studies I-II (57,58). The physical activity level was unchanged from visit 2 (median, IQR; 375, 200 to 620 total PA-intensities min/week) to follow-up (372.5, 180.0 to 810.0 total PA-intensities min/week) (p = 0.32) (58).

A simplification of the overall psychometric evidence of outcome measures in PwDM1 is shown in Figure 29. The rationale behind the evaluation is elaborated on in the following sections.

Figure 29: The psychometric evidence of the outcome measures in PwDM1 (Studies I-II)

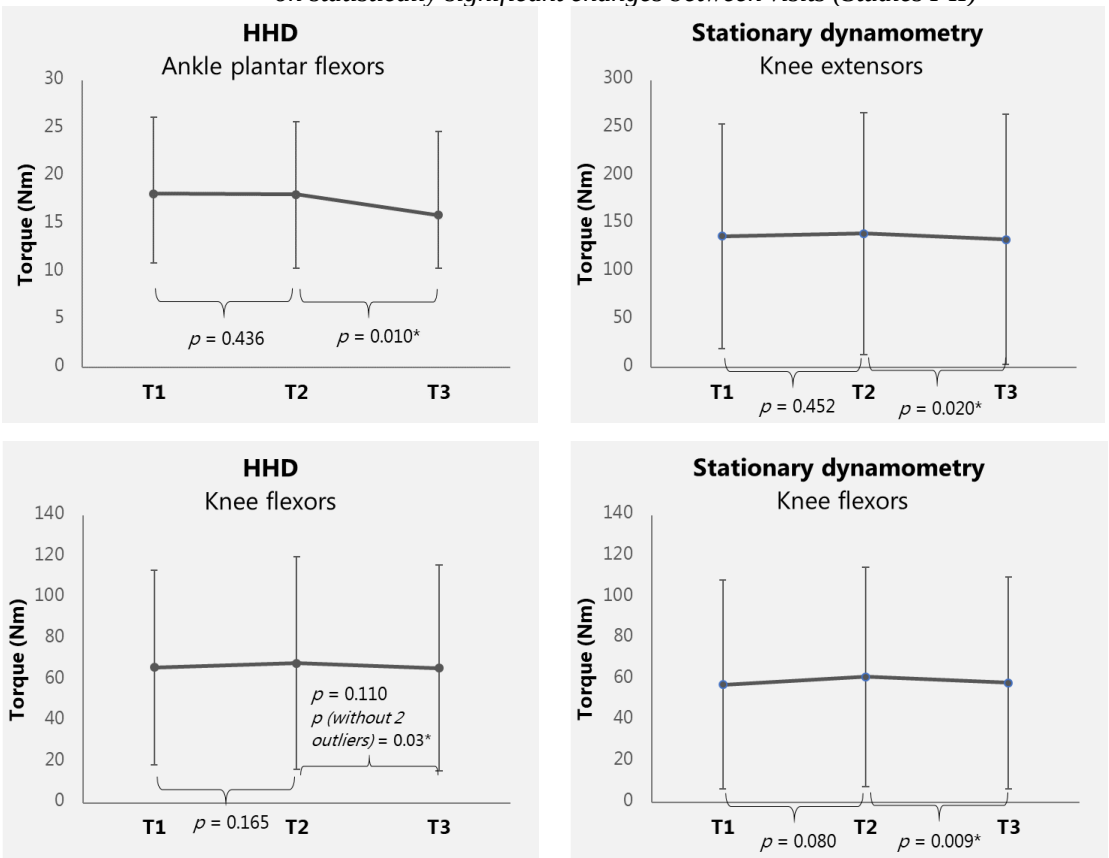


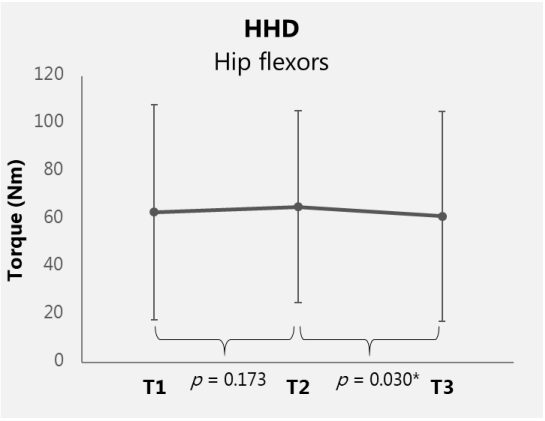
Muscle strength measurements

Intra-rater reliability

HHD was more reliable than stationary dynamometry because the HHD showed a lower or similar degree of measurement error compared to the stationary dynamometry both on (I) a group level as to systematic error (difference between visit 1 and visit 2) (Figure 30+31) and absolute measurement error (SEM) (Figure 32), and on (II) a single individual level as to absolute measurement error (MDD) (Figure 33) (57). Measurement error for single individuals is visualized in Figure 34. The MDD for single individuals exceeded the MCID for both HHD and stationary dynamometry (Figure 33).

Figure 30: Reliability (T1-T2) and responsiveness (T2-T3) of muscle strength measurements in PwDM1 based on statistically significant changes between visits (Studies I-II)





**Statistically significant. HHD (ankle plantar flexors): median (IQR) because mean – (1.96*SD) resulted in a negative value which is meaningless. Stationary dynamometry (knee extensors and -flexors) and HHD (flexors in knee and hip): mean (95% CI). Graphs are only provided for the outcome measures that showed no change at T2 but a change at T3. Graphs for the remaining outcome measures are available in Supplementals (Supplementals, Figure 2).*

Figure 31: Matrix of simplified reliability (T1-T2) and responsiveness (T2-T3) of outcome measures in PwDM1 based on statistically significant changes between visits (Studies I-II)

	RELIABLE	UNRELIABLE
RESPONSIVE	**Stationary dynamometry:** • Knee extensors • Knee flexors **HHD** • Ankle plantar flexors • Knee flexors (without 2 outliers) • Hip flexors **m-CTSIB**	**Stationary dynamometry:** • Hip extensors • Hip flexors **TUG**
UNRESPONSIVE	**Stationary dynamometry:** • Ankle plantar flexors • Ankle dorsal flexors **HHD** • Ankle dorsal flexors • Knee extensors • Knee flexors (with 2 outliers) • Hip extensors **One-leg-stance eyes closed** **10mWT**	**Step test** **STS**

Figure 32: Absolute reliability on a group level (SEM) of outcome measures in PwDM1 (Study I)

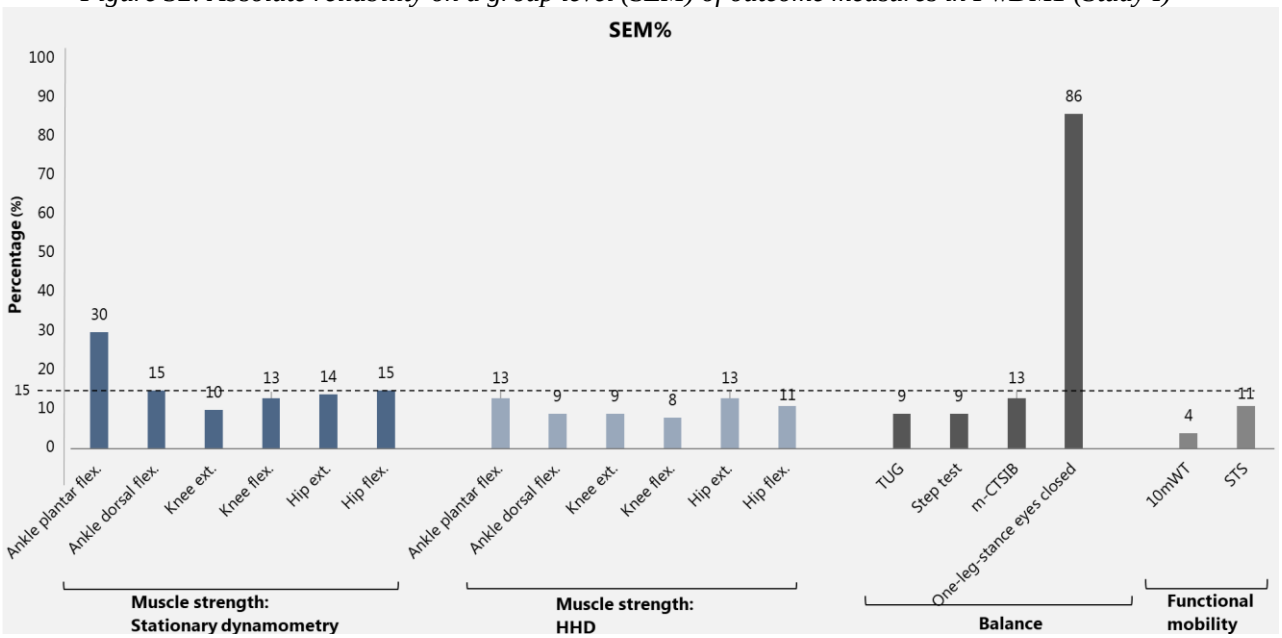


Figure 33: Absolute reliability on an individual level (MDD, MCID) and responsiveness (percent of PwDM1 with change values >MCID) of outcome measures in PwDM1 (Studies I-II)

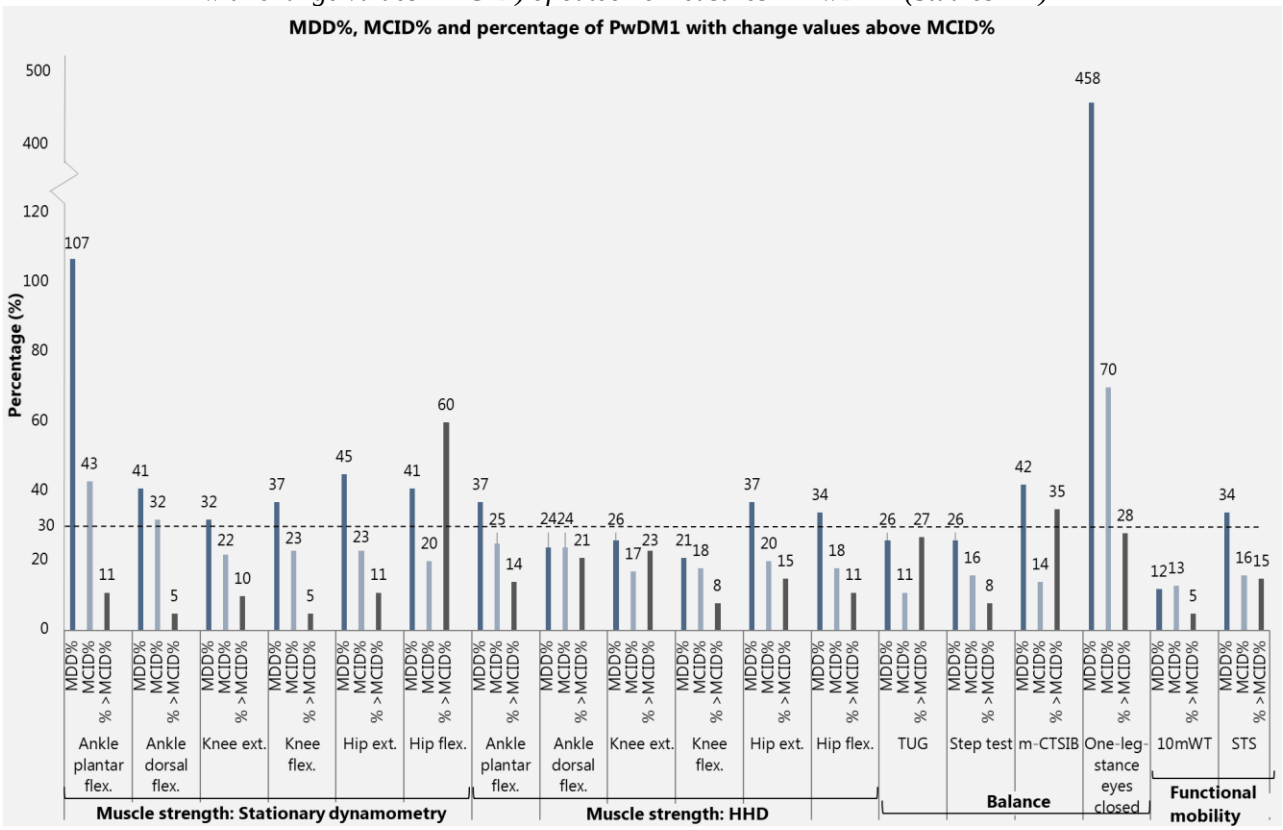
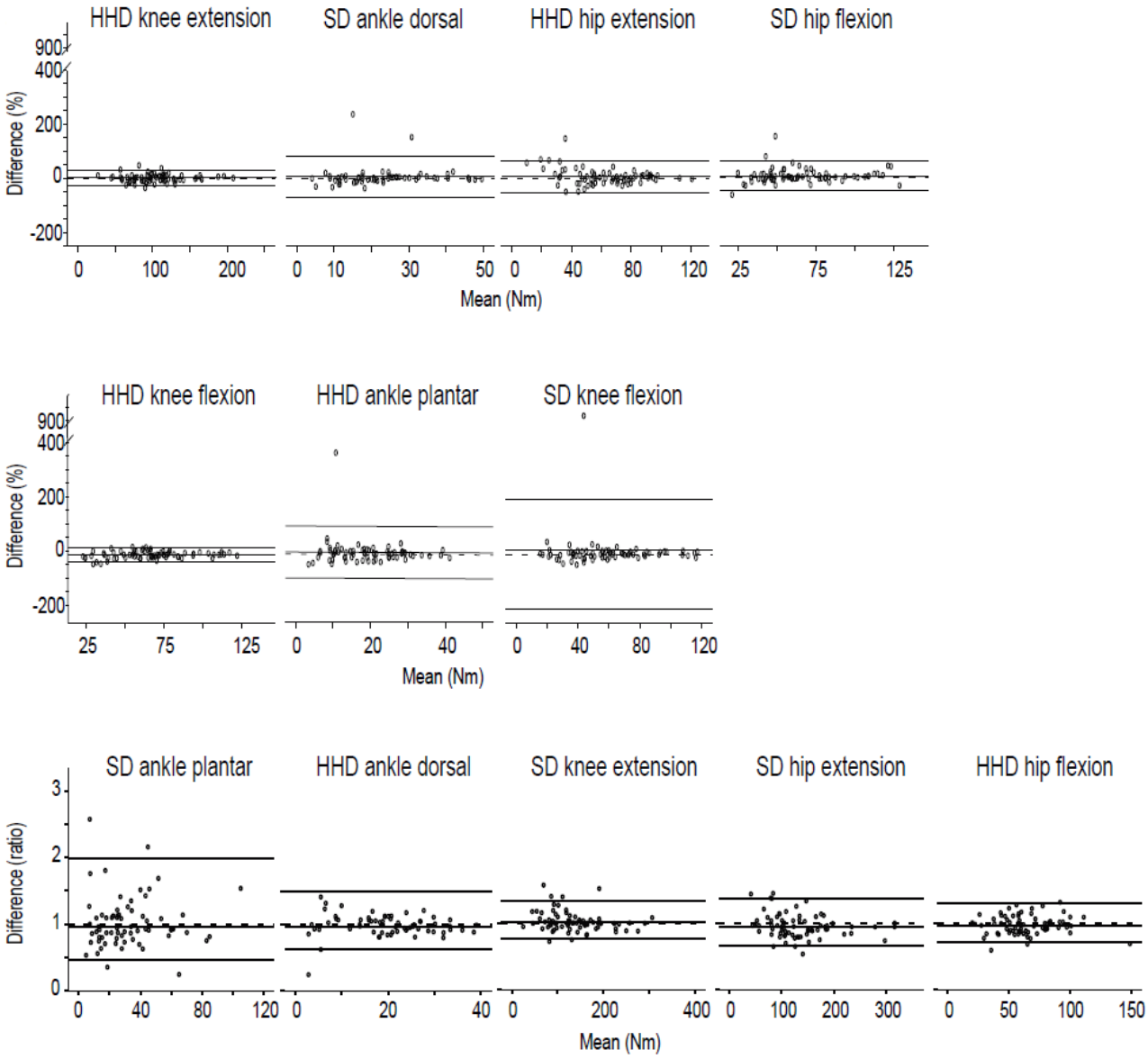


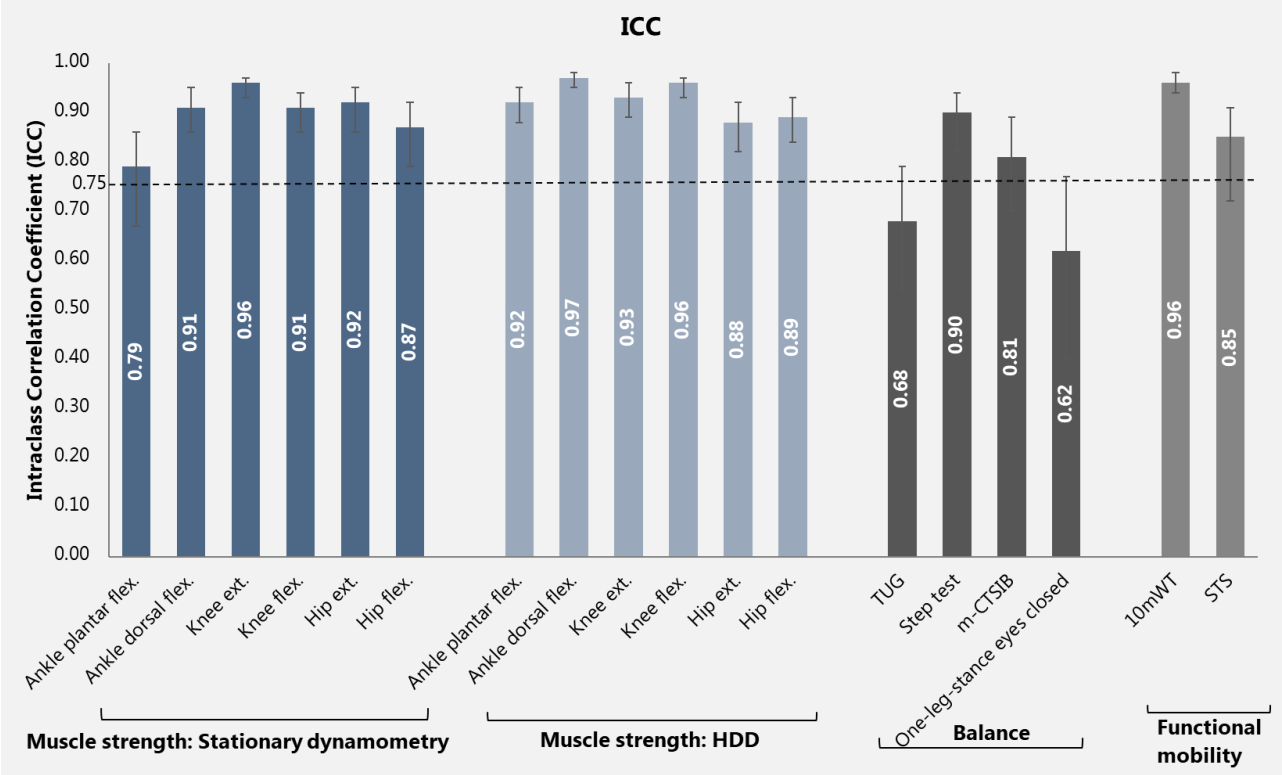
Figure 34: Bland-Altman plots of reliability on an individual level of muscle strength measurements in PwDM1 (Study I)



HHD = Hand-Held Dynamometry; SD = Stationary Dynamometry.
X-axis: (visit 1 + visit 2)/2; Y-axis: visit 1 − visit 2. Difference (percent) is calculated on original data. Difference (ratio) is calculated on log10-arithmetic transformed data and back-transformed into ratio. A circle represents 1 participant. A black dot represents >1 participant. The dotted lines indicate no difference of the repeated measurements. The three full lines indicate mean difference ±1.96 SD. The closer the circles/black dots (participants) are to 0 for untransformed difference (percent) and to 1 for log10-arithmetic back-transformed difference (ratio), the higher agreement between visits 1 and 2, which indicates less absolute measurement error. This figure is from Knak et al. (57).

The secondary relative reliability results showed good to excellent relative reliability for all muscle groups for both HHD and stationary dynamometry, except ankle plantar flexors with stationary dynamometry (Figure 35).

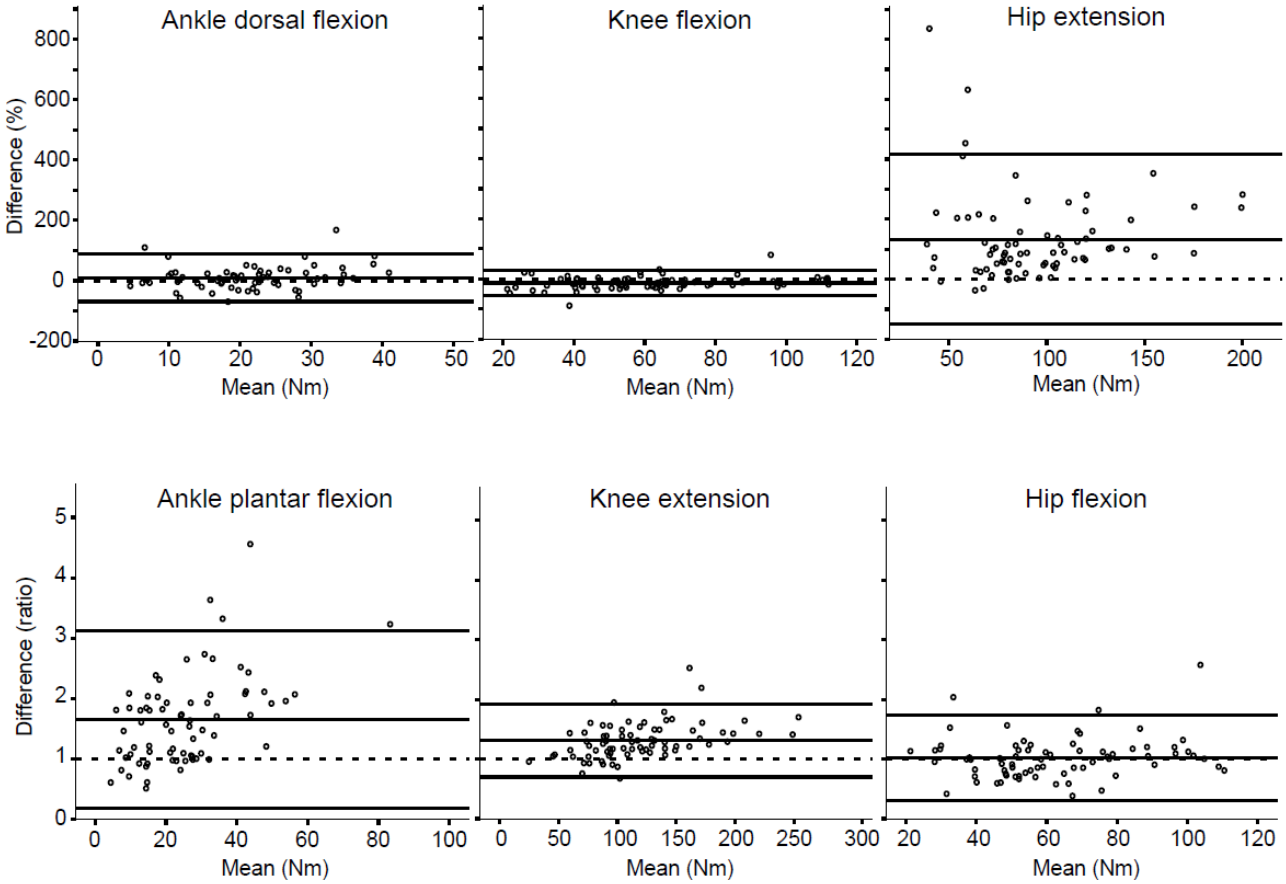
Figure 35: Relative reliability (ICC) of outcome measures in PwDM1 (Study I)



Criterion validity

On a group level, a higher torque was captured by the stationary dynamometry for the ankle plantar flexors (mean, 95% CI; 12.53 Nm, 8.65 to 16.41 Nm, $p < 0.0005$), knee extensors (32.93 Nm, 25.04 to 40.82 Nm, $p < 0.0005$), and hip extensors (63.05 Nm, 51.12 to 74.98 Nm, $p < 0.0005$) compared to the HHD (57). A lower torque was registered by the stationary dynamometry for the knee flexors (–8.71 Nm, -11.81 to -5.53 Nm, $p < 0.0005$) compared to the HHD (57). Muscle strength of the remaining lower limb flexor muscles did not differ between the two muscle strength devices ($p \geq 0.22$). On an individual level, the registered muscle strength was inconsistent between the HHD and the stationary dynamometry for all muscle groups (especially for the hip extensors and ankle plantar flexors), but none of the devices were favored with 95% confidence (Figure 36) (57).

Figure 36: Bland-Altman plots of validity of muscle strength measurements in PwDM1 (Study I)

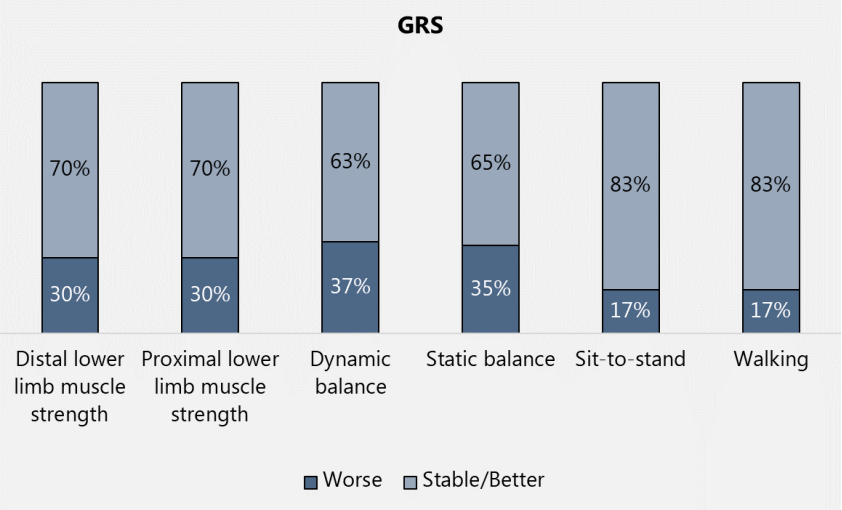


X-axis: (stationary dynamometry + HHD)/2; Y-axis: Stationary dynamometry – HHD.
Difference (percent) is calculated on original data. Difference (ratio) is calculated on log10-arithmetic transformed data and back-transformed into ratio. A circle represents 1 participant. A black dot represents >1 participant. The dotted lines indicate no difference of the repeated measurements. The three full lines indicate mean difference ±1.96 SD. The closer the circles/black dots (participants) are to 0 for untransformed difference (percent) and to 1 for log10-arithmetic back-transformed difference (ratio), the higher agreement between stationary dynamometry and HHD, which indicates stronger validity. This figure is from Knak et al. (57).

Responsiveness

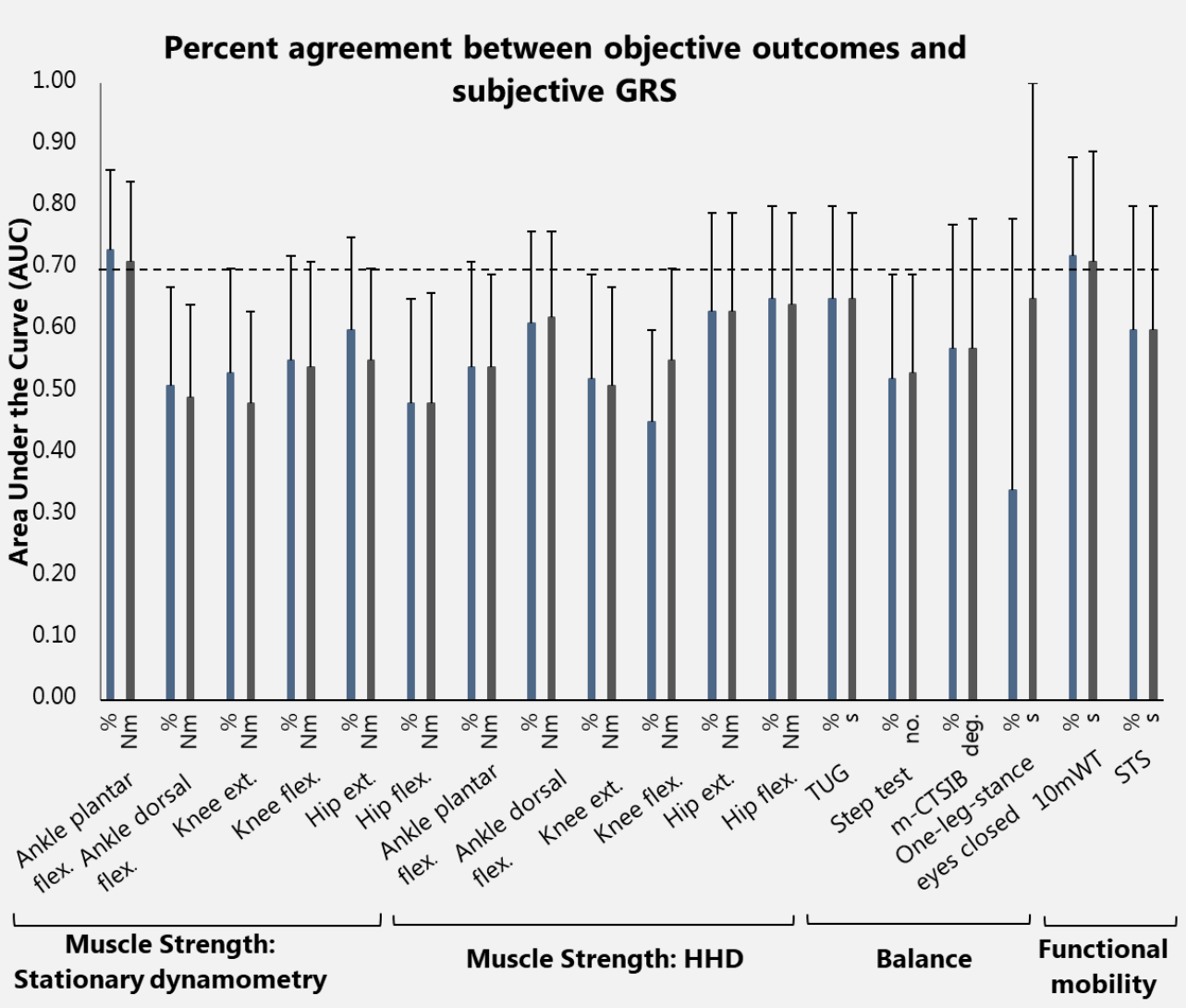
At follow-up, stationary dynamometry detected an alteration in muscle strength in the flexor- and extensor muscles of the hip and knee ($p \leq 0.02$), and HHD registered a change in the hip- and knee flexors (without 2 outliers for the knee flexors) and in the ankle plantar flexors ($p \leq 0.03$) (Figure 30+31) (58). The lack of change in the ankle dorsal flexors using stationary dynamometry persisted when the participants who were unable to generate sufficient muscle strength to activate the device (21% of the participants) were excluded from the analysis. The subjective perception of change in muscle strength at follow-up compared to "baseline" (T2) is visualized in Figure 37. The proportion of PwDM1 who demonstrated clinically important change from visit 2 to visit 3 (change value > MCID) is shown in Figure 33.

Figure 37: Global Rating Scale (GRS) reported by PwDM1 at follow-up (Study II)



Ankle plantar flexor strength using stationary dynamometry was the only objective muscle strength outcome that agreed with the subjective GRS perception of change in muscle strength at follow-up based on the best estimate (men AUC > 0.70), but agreement is possible when the 95% CI is accounted for, primarily for the extensor muscles using stationary dynamometry and for nearly all muscle groups using HHD (Figure 38) (58).

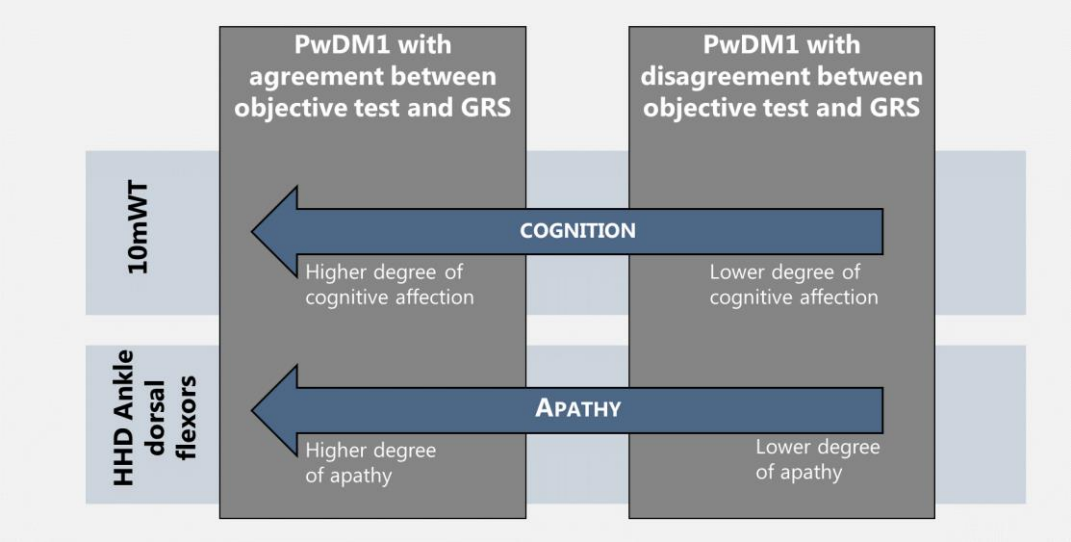
Figure 38: Agreement between the objective outcome measures and the subjective GRS in PwDM1 (Study II)



AUC is presented for both absolute values (Nm, s, no., deg.) and relative values (%) with 95% CI (positive error bars are shown). This figure is from Knak et al. (58). Please notice that I have made a correction to the Figure in Knak et al. (58) which has displayed the AUC 95% confidence intervals erroneously, but not the mean AUC. The correct 95% confidence intervals are shown in this thesis.

The PwDM1 with disagreement between change in objective ankle muscle strength measurement (HHD) and change in subjective ankle muscle strength measurement (GRS) showed a lower degree of apathy (median, IQR; 9, 6-12 AES-score) than the PwDM1 with agreement between the HHD and the GRS (13, 10-17 AES-score) ($p = 0.048$) (Figure 39). Otherwise, there was no difference in apathy or cognition between the two groups as to change in proximal muscle strength (knee and hip) ($p \geq 0.24$).

Figure 39: Difference in cognition and apathy in PwDM1 (Study II)



Feasibility

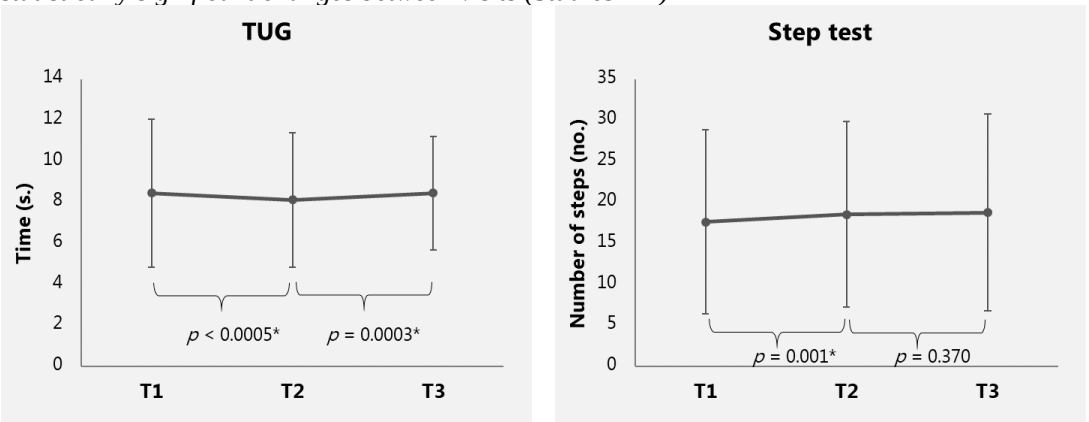
The HHD was more feasible than the stationary dynamometry regarding easiness of application, execution time and economics.

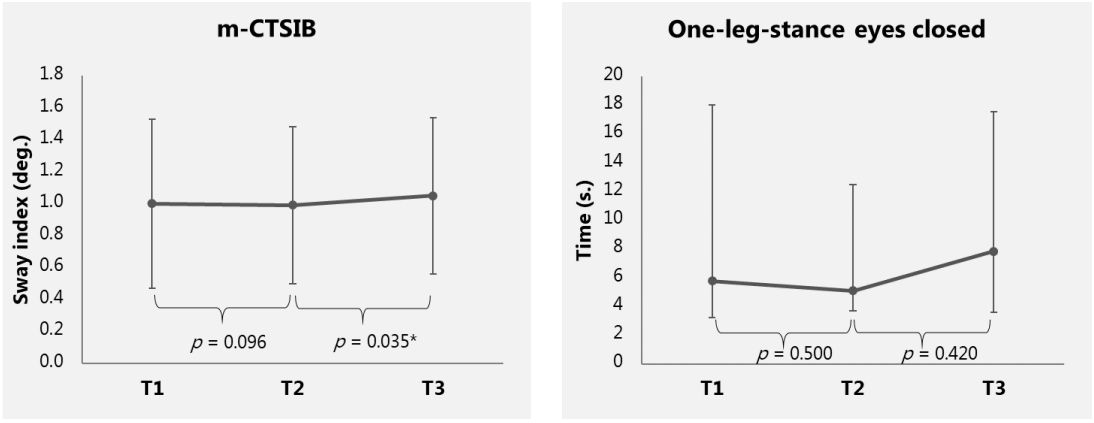
Balance measurements

Intra-rater reliability

The dynamic TUG and step test showed lower absolute measurement error in groups (SEM) (Figure 32) and single individuals (MDD) (Figure 33) compared to the static m-CTSIB and especially the one-leg-stance eyes closed test. However, only the dynamic balance tests were associated with learning effects ($p \leq 0.001$) (Figure 31+40).

Figure 40: Reliability (T1-T2) and responsiveness (T2-T3) of balance measurements in PwDM1 based on statistically significant changes between visits (Studies I-II)

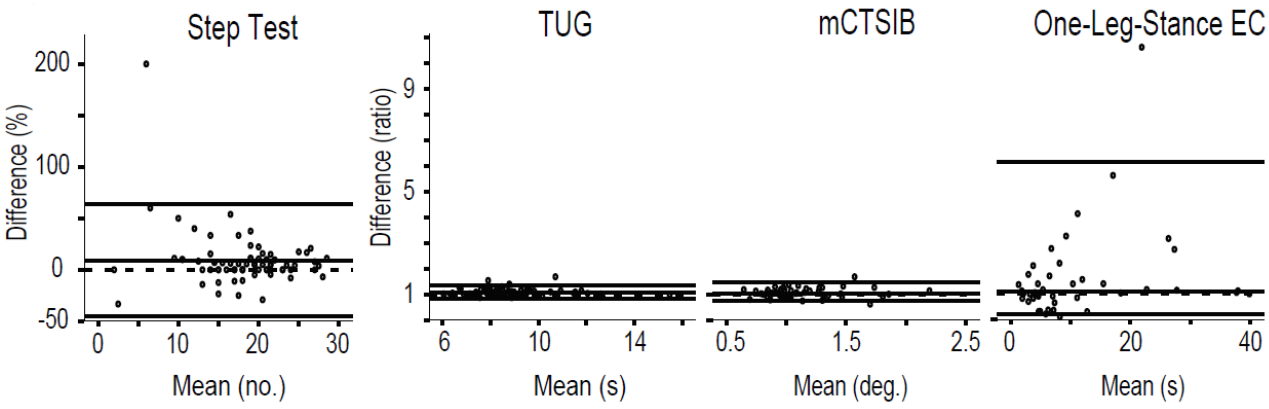




**Statistically significant. One-leg-stance eyes closed: median (IQR) because mean – (1.96*SD) resulted in a negative value which is meaningless. TUG, step test and m-CTSIB: mean (95% CI). Graphs for feet-together stance, tandem stance and one-leg-stance eyes open with no true variation are provided in Supplementals (Supplementals, Figure 3).*

Measurement error for single individuals is visualized in Figure 41. MDD exceeded the MCID for both the dynamic- and static balance tests (Figure 33).

Figure 41: Bland-Altman plots of reliability on an individual level of balance measurements in PwDM1 (Study I)



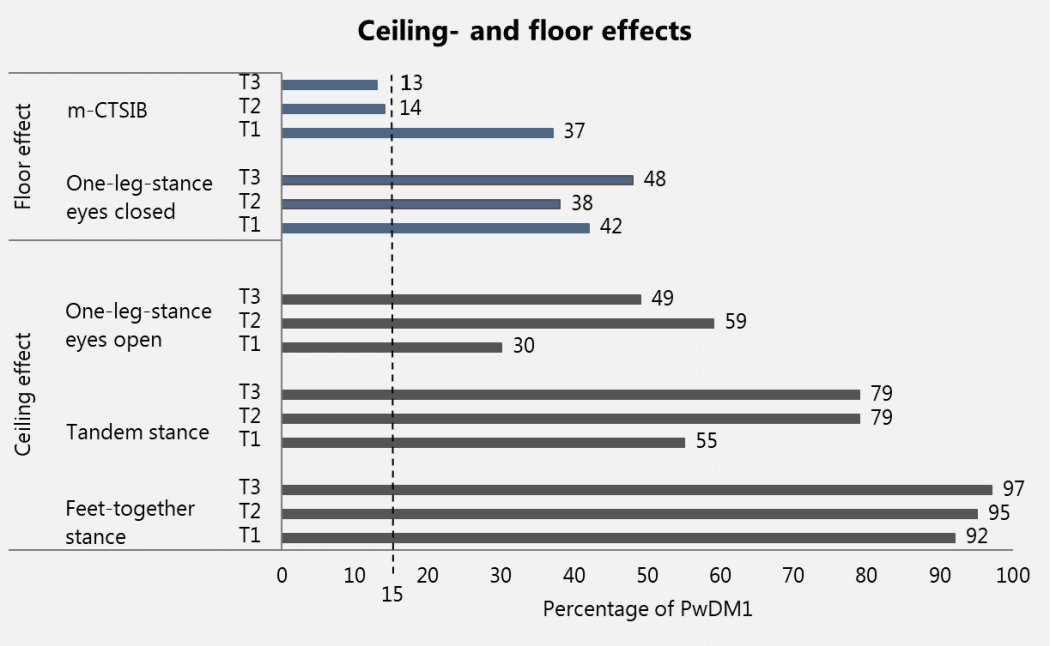
TUG = Timed-Up-and-Go test; mCTSIB = modified Clinical Test of Sensory Integration and Balance; EC = Eyes Closed.
X-axis: (visit 1 + visit 2)/2; Y-axis: visit 1 − visit 2. Difference (percent) is calculated on original data. Difference (ratio) is calculated on log10-arithmetic transformed data and back-transformed into ratio. A thick circle represents >1 participant. The dotted lines indicate no difference of the repeated measurements. The three undotted lines indicate mean difference ±1.96 SD. The closer the circles (participants) are to 0 for untransformed difference (percent) and to 1 for log10-arithmetic back-transformed difference (ratio), the higher agreement between visits 1 and 2, which indicates less absolute measurement error. This figure is from Knak et al. (57).

The secondary relative reliability results showed good to excellent reliability for the step test, moderate to good reliability for the m-CTSIB and the TUG, and fair to good reliability of the one-leg-stance eyes closed (Figure 35).

Construct validity

Ceiling effects were present in the static balance tests feet-together stance, tandem stance and one-leg-stance eyes open. Floor effects were detected in the remaining static balance tests m-CTSIB and one-leg-stance eyes closed (e.g. 42% of the PwDM1 did not qualify for one-leg-stance eyes closed because of < 30 s stance in one-leg-stance test eyes open (29), thus a surrogate measure was applied) (Figure 42 [T1]) (57).

Figure 42: Ceiling- and floor effects in balance measurements regarding validity (T1) and responsiveness (T2-T3) in PwDM1 (Studies I-II)



This figure is from Knak et al. (58) (with minor modifications).

Since the one-leg-stance eyes closed test demonstrated very poor reliability, validity analysis was not conducted because validity is automatically compromised (75). A poorer balance performance of 1-second increase in TUG predicted a poorer balance performance with a reduction of 1.66 steps (95% CI, -2.35 to -0.98 steps) in the step test and explained 22.6% of the step test performance (Table 2) (57). A poorer balance performance of doubling in TUG time predicted a poorer muscle strength performance in lower limbs with a muscle strength reduction of -37.9% (-58.8 to -6.4%) and explained 7.0% of the lower limb muscle strength performance (Table 2) (57). A better balance performance of 1 step increase in step test predicted a better ankle muscle strength performance with an ankle muscle strength improvement of 5.5% (3.2 to 7.7%) and explained 28.6% of the ankle muscle strength performance (Table 2) (57).

Table 2: Regression and adjusted coefficient of determination (R^2) of balance and functional mobility measurements in PwDM1 (Study I)

INDEPENDENT VARIABLE, PREDICTOR (X)	DEPENDENT VARIABLE, OUTCOME (Y)	SLOPE OF REGRESSION LINE (β)	95% CI FOR β	P-VALUE FOR β	ADJUSTED R^2	P-VALUE FOR ADJUSTED R^2
TUG	Step Test	-1.664	-2.352 to -0.997	<0.0005*	0.226	<0.0005*
TUG	Muscle strength (hip extensors, knee extensors and ankle dorsal flexors)	0.621[1]	0.412[1] to 0.936[1]	0.023*	0.070	0.023*
Step Test	Muscle strength (ankle plantar- and dorsal flexors)	1.055[1]	1.032[1] to 1.077[1]	<0.0005*	0.286	<0.0005*
STS	10mWT	0.168	0.127 to 0.209	<0.0005*	0.465	<0.0005*
STS	Muscle strength (hip- and knee extensors)	0.975[1]	0.9599[1] to 0.992[1]	0.003*	0.103	0.003*
10mWT	Muscle strength (hip extensors, knee extensors, ankle plantar- and dorsal flexors)	0.861[1]	0.802[1] to 0.925[1]	<0.0005*	0.224	<0.0005*

The muscle strength data are from stationary dynamometry (torque, mean of the muscle groups' best values). The data are original data unless otherwise stated. *Statistically significant. [1]Antilog$_2$: ratio. This table is from Knak et al. (57).

The balance tests correlated little (τ_b=-0.220 to τ_b=-0.251), fairly (r=-0.313 to r=-0.486) and moderately (r=0.546) with tests assessing similar components (Table 3). The dynamic balance tests showed higher correlations (r=-0.313 to r=-0.486) than the static balance test (τ_b=-0.220 to τ_b=-0.251) (Table 3).

Table 3: Correlation of balance and functional mobility measurements in PwDM1 (Study I)

OUTCOME MEASURES	PEARSON'S CORRELATION COEFFICIENT	P-VALUE[1]	95% CI
TUG correlation with:			
Step test	-0.486	<0.0005*	-0.687 to -0.286
STS	0.439	<0.0005*	0.292 to 0.695
Muscle strength (hip extensors, knee extensors and ankle dorsal flexors)	-0.313[2]	0.015*	NA
Step test correlation with:			
Muscle strength (ankle plantar- and dorsal flexors)	0.546	<0.0005*	0.327 to 0.764
m-CTSIB (composite score) correlation with:			
Step test	-0.251[2]	0.008*	NA
Muscle strength (ankle plantar- and dorsal flexors)	-0.220[2]	0.022*	NA
STS correlation with:			
10mWT	0.687	<0.0005*	0.519 to 0.855
Muscle strength (hip- and knee extensors)	-0.340	0.003*	-0.563 to -0.117
10mWT correlation with:			
Muscle strength (hip extensors, knee extensors and ankle plantar- and dorsal flexors)	-0.488	<0.0005*	-0.719 to -0.256

NA = Not Applicable because 95% CI is not provided for the non-parametric Kendall's tau-b test.

The muscle strength measures are based on values from stationary dynamometry (torque, mean of the muscle groups' best values).

*Statistically significant

[1]Adjustment for multiple testing:

- Since TUG was correlated with three variables independently, the statistically significance level has been Bonferroni-corrected at α = 0.05/3 = 0.017.
- Since step test was correlated with three variables independently, the statistically significance level has been Bonferroni-corrected at α = 0.05/3 = 0.017.
- Since STS was correlated with three variables independently, the statistically significance level has been Bonferroni-corrected at α = 0.05/3 = 0.017.
- Since m-CTSIB (composite score) was correlated with two variables independently, the statistically significance level has been Bonferroni-corrected at α = 0.05/2 = 0.025.
- Since 10mWT was correlated with two variables independently, the statistically significance level has been Bonferroni-corrected at α = 0.05/2 = 0.025.

[2]Kendall's tau-b

This table is from Knak et al. (57).

Responsiveness

Ceiling- or floor effects were found in all static balance tests at visit 2 (T2) and follow-up (T3), except for m-CTSIB (Figure 42 [T2, T3]). TUG and m-CTSIB captured a change in balance at follow-up ($p \leq 0.035$) in contrast to the other balance tests ($p \geq 0.37$) (Figure 31+40) (58). The proportion of PwDM1 with clinically relevant T2-T3 change values is shown in Figure 33. For the subjective perception of change/no change in balance ability, see Figure 37. The change in objective static- and dynamic balance tests did not agree

acceptably with the subjective perception of change in dynamic and static balance at follow-up based on the best estimate (mean AUC < 0.70) but considering the 95% CI, agreement is possible for almost all the balance tests (Figure 38). There was no difference in apathy or cognition between the PwDM1 for whom the objective balance test (step test) agreed with the subjective balance GRS and the PwDM1 for whom the step test disagreed with subjective balance GRS at follow-up ($p \geq 0.078$).

Feasibility

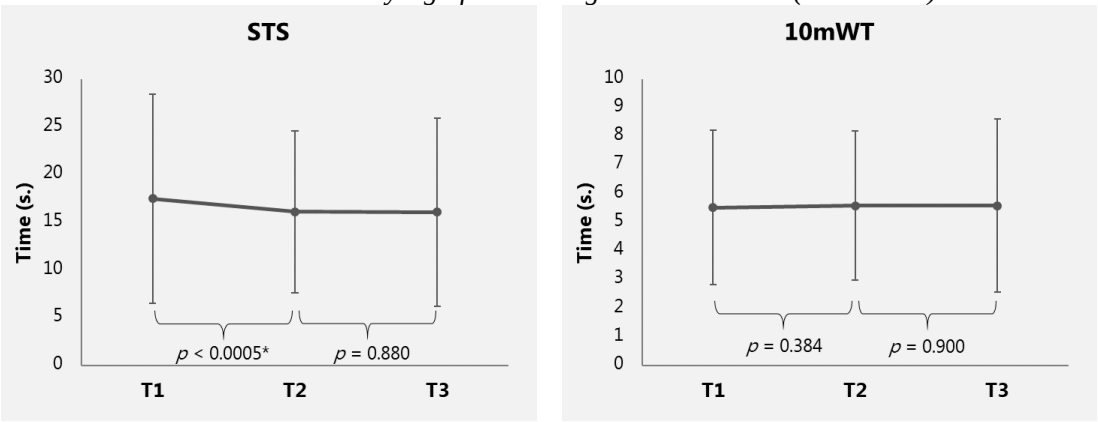
The m-CTSIB was less feasible than the other balance tests as to easiness of application, execution time, economics and interpretation of results. The dynamic balance tests were feasible but a little less feasible than the static balance tests (except for m-CTSIB), because TUG required walking space and step test required a step bench.

Functional mobility measurements

Intra-rater reliability

Both functional mobility tests were acceptably reliable on a group- (SEM) (Figure 32) and individual level (MDD) (MDD of 27% for STS with exclusion of outliers, not shown) (Figure 33). The 10mWT showed the lowest absolute measurement errors (Figure 32+33), an MDD < MCID (Figure 33) and was not associated with learning effects ($p = 0.38$) in contrast to the STS ($p < 0.0005$) (Figures 31+43).

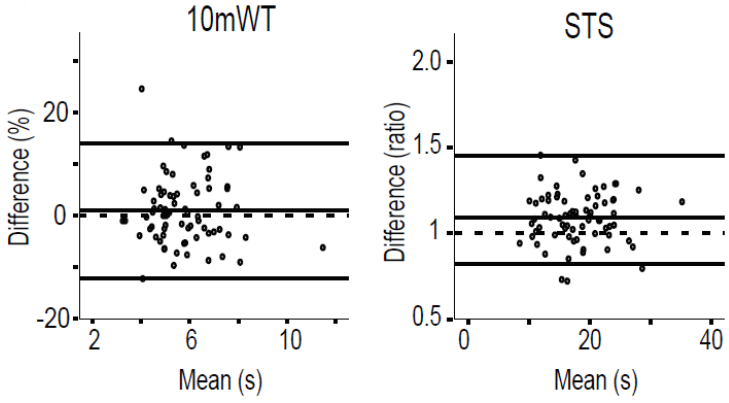
Figure 43: Reliability (T1-T2) and responsiveness (T2-T3) of functional mobility measurements in PwDM1 based on statistically significant changes between visits (Studies I-II)



**Statistically significant. STS and 10mWT: mean (95% CI).*

Measurement error for single individuals are visualized in Figure 44.

Figure 44: Bland-Altman plots of reliability on an individual level of functional mobility measurements in PwDM1 (Study I)



10mWT = 10-meter Walk Test; STS = Sit-To-Stand test.
X-axis: (visit 1 + visit 2)/2; Y-axis: visit 1 − visit 2. Difference (percent) is calculated on original data. Difference (ratio) is calculated on log10-arithmetic transformed data and back-transformed into ratio. A thick circle represents >1 participant. The dotted lines indicate no difference of the repeated measurements. The three undotted lines indicate mean difference ±1.96 SD. The closer the circles (participants) are to 0 for untransformed difference (percent) and to 1 for log10-arithmetic back-transformed difference (ratio), the higher agreement between visits 1 and 2, which indicates less absolute measurement error. This figure is from Knak et al. (57).

The secondary relative reliability results showed moderate to excellent reliability for the STS and excellent reliability for the 10mWT (Figure 35).

Construct validity

A poorer functional mobility performance of 1-second increase in STS predicted (I) a poorer functional mobility performance with an increase of 0.168 s (95% CI, 0.127 to 0.209 s) in 10mWT and explained 46.5% of the 10mWT performance (II) a poorer proximal lower limb muscle strength performance with a reduction in muscle strength of 2.5% (-4.1 to -0.8%) and explained 10.3% of the muscle strength performance (Table 2) (57). A poorer functional mobility performance of a 1-second slower 10mWT predicted a poorer lower limb muscle strength performance with a muscle strength reduction of 13.9% (-19.8 to -7.5%) and explained 22.4% of the lower limb muscle strength performance (Table 2) (57). Fair (r=-0.340 to r=-0.488) to good (r=0.687) correlations were found between the functional mobility tests and tests within the same aspect (Table 3) (57).

Responsiveness

The 10mWT and STS did not register a change in functional mobility at follow-up ($p \geq 0.88$) (Figures 31+43), but acceptable agreement with functional mobility GRS was reached with the 10mWT (mean and 95% CI AUC > 0.70) and was possible for the STS (95% CI AUC > 0.70) (Figure 38) (58). The percent of PwDM1 who demonstrated clinically important change from visit 2 to visit 3 is shown in Figure 33. The subjective perception of change/no change in functional mobility from "baseline" (T2) to follow-up (T3) is shown in Figure 37. The PwDM1 with disagreement between change in the objective 10mWT and change in the subjective functional mobility GRS showed better cognition (median, IQR; 43.00, 39.25 to 48.00 STROOP mean score) than the PwDM1 with agreement between the 10mWT and the GRS (38.25, 35.50 to 41.13 STROOP mean score) (p = 0.02), but apathy did not differ between the two groups (p = 0.37) (Figure 39) (58).

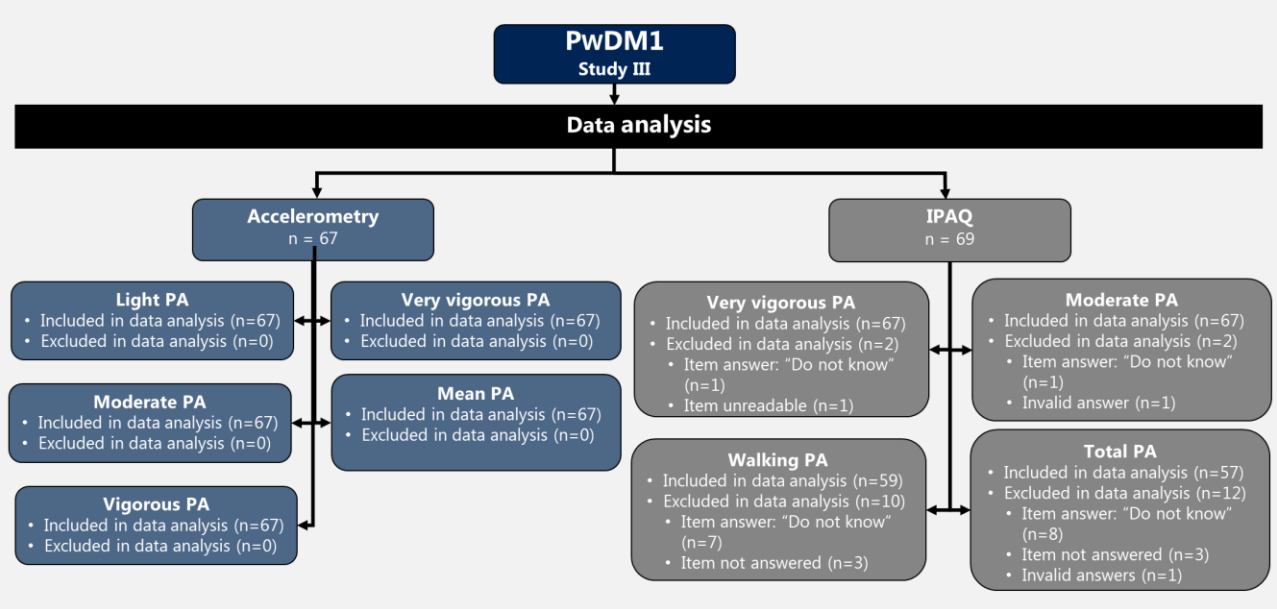
Feasibility

Both the 10mWT and STS were feasible, but the 10mWT was minorly less feasible than the STS because of walking space requirement.

9.2.2 Study III: Physical activity

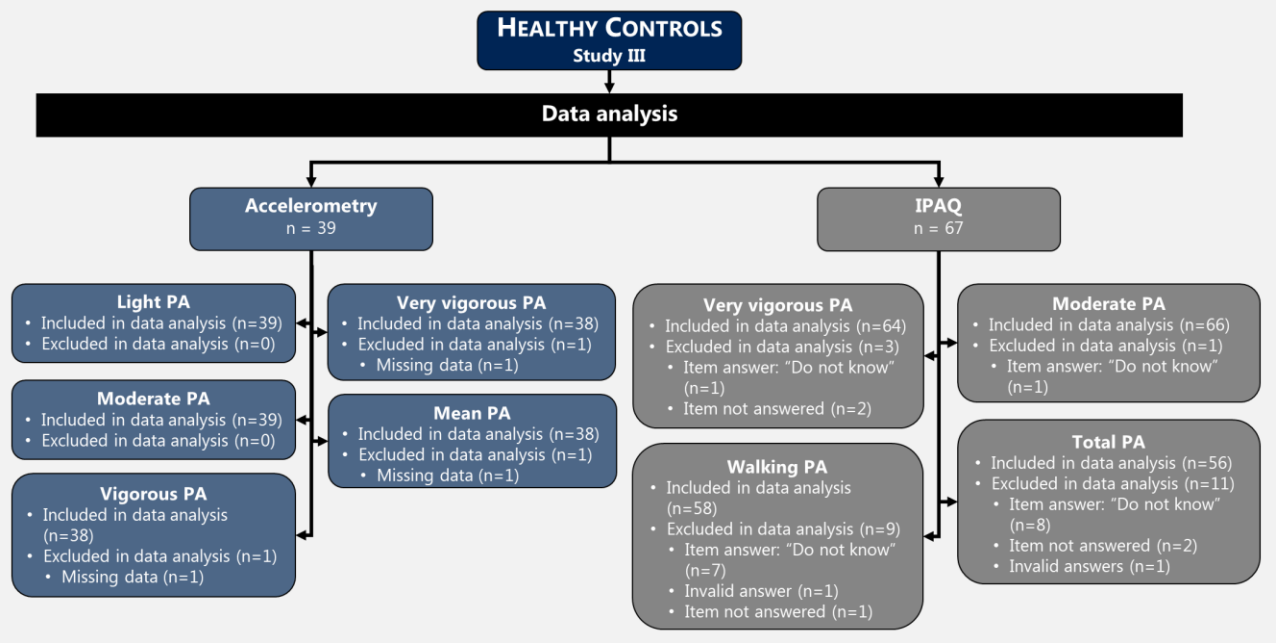
Flowcharts of number of participants included in data analyses are provided for PwDM1 (Figure 45) and healthy controls (Figure 46).

Figure 45: Flowchart of number of PwDM1 included in data analyses (Study III)



PA = Physical Activity.

Figure 46: Flowchart of number of healthy controls included in data analyses (Study III)

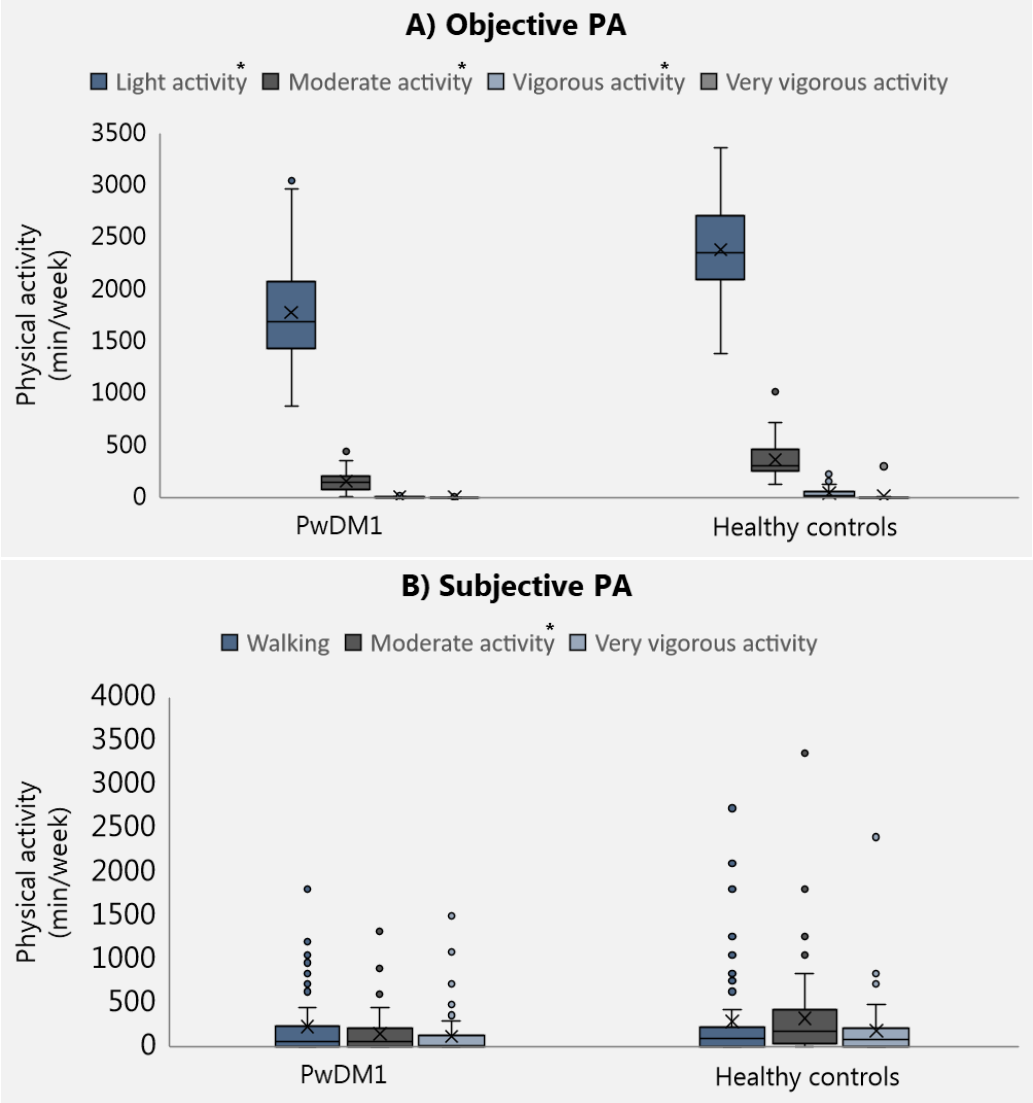


PA = Physical Activity.

The physical activity level (mean PA of all intensities/week), monitored objectively by accelerometry, was -187 min/week (95% CI, -248 to -127 min/week, $p < 0.00001$) lower in PwDM1 (mean ± SD, 485 ± 144 min/week) compared to healthy controls (695 ± 138 min/week) (59). The physical activity level (total PA of all intensities/week), measured subjectively by IPAQ, was -48%/week (95% CI, -65 to -23%/week, $p = 0.001$) lower in PwDM1 (median IQR, 380, 215 to 720 min/week) compared to healthy controls (550, 368 to 983 min/week) (59). In addition, secondary analyses showed that the PwDM1 demonstrated a lower objective physical activity level than healthy controls for light intensity ($p < 0.00001$), moderate intensity ($p < 0.00001$) and vigorous intensity ($p = 0.00005$), but not for very vigorous intensity ($p = 0.08$) (Figure 47 [A]). The

PwDM1 were less physically active subjectively compared to healthy controls for moderate intensity (p = 0.04), but not for walking (p = 0.13) and very vigorous intensity (p = 0.37) (Figure 47 [B]).

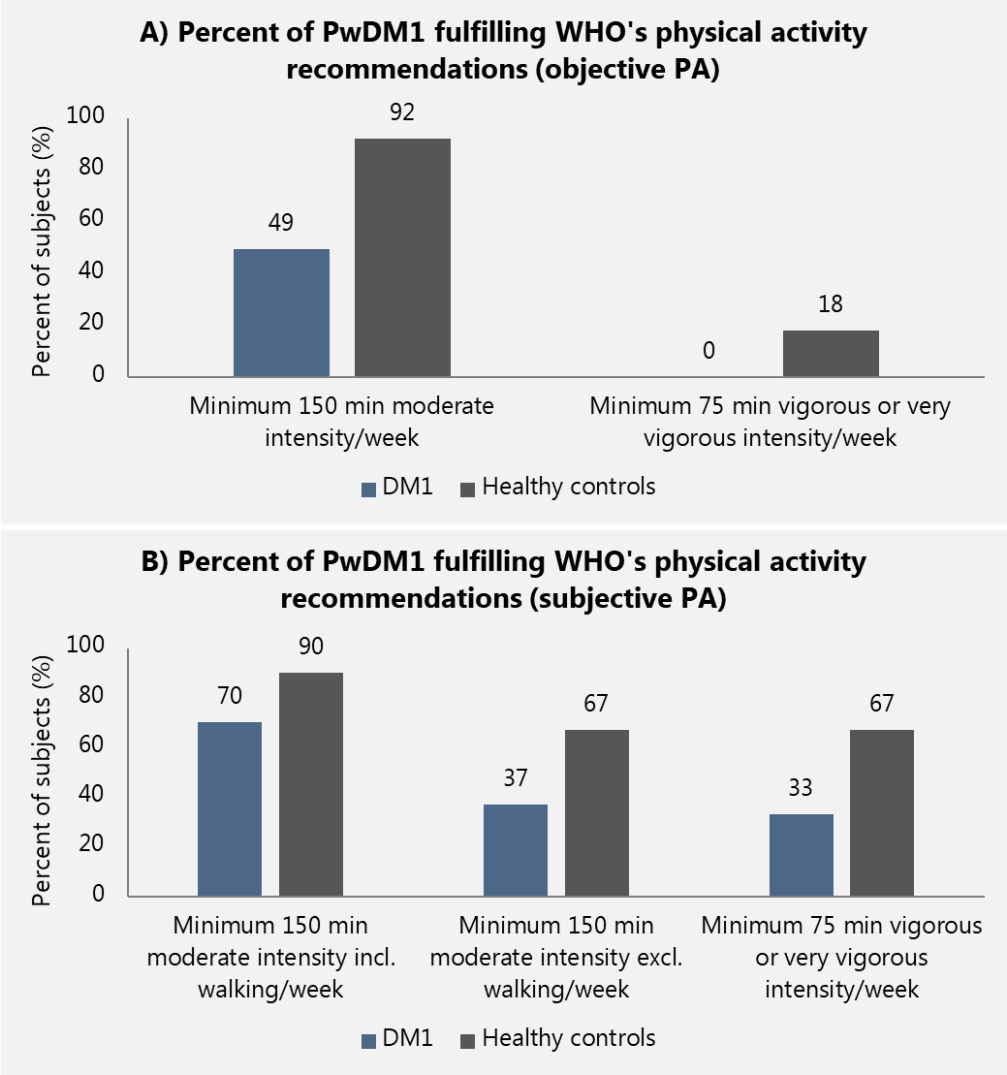
Figure 47: The objective and subjective physical activity intensities in PwDM1 and healthy controls (Study III)



**Statistically significant. Box = 25-75% percentiles. Cross = mean. Horizontal line = median/50% percentile. Whisker = minimum and maximum values that are not outliers. Dot = outlier, defined as data points > 1.5 box-length from the edge of their box. This figure is from Knak et al. (59).*

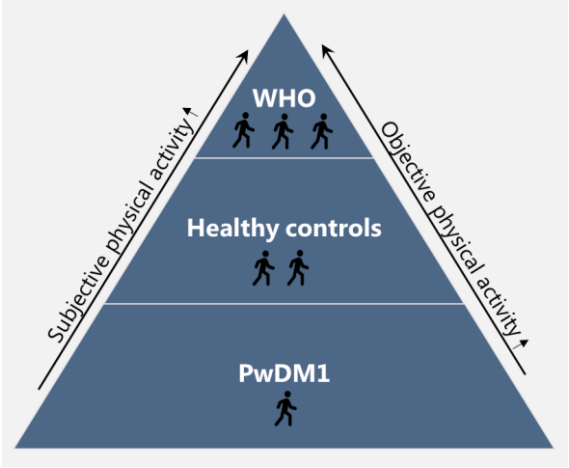
Likewise, did the PwDM1 adhere less to the WHO recommendations on physical activity, both objectively and subjectively, compared to healthy controls (Figure 48+49). Overall, the adherence to the WHO's physical activity recommendations was poor in PwDM1 and moderate to good in healthy controls (Figures 48+49).

Figure 48: The percent of PwDM1 and healthy controls fulfilling the WHO's physical activity recommendations objectively and subjectively (Study III)



This figure is from Knak et al. (59).

Figure 49: Hierarchy of physical activity levels in PwDM1 and healthy controls (Study III)



The objective physical activity level in PwDM1 was only predicted by the level of education with one higher level of education predicting an increase of mean physical intensities per week by 29 min/week (95% CI, 5 to

53 min/week, $p = 0.02$) (Table 4) (59). The objective physical activity level in healthy controls was not statistical significantly predicted by age, marital status or education ($p \geq 0.20$) (59).

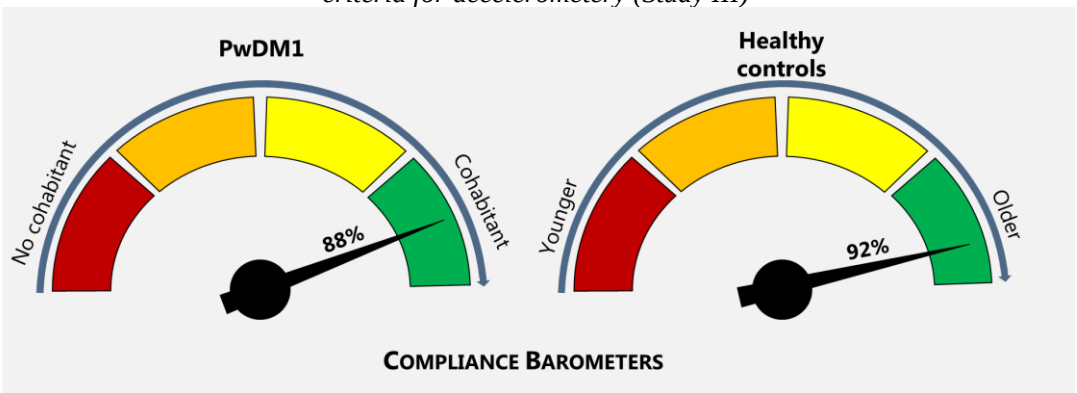
Table 4: Predictors of physical activity in PwDM1

PREDICTOR VARIABLE	β	SE_β	95% CI_β	P-VALUE
Age	-3.05	1.68	-6.34; 0.25	0.07
Marital status Cohabitant No cohabitant	59.03	42.39	-24.05; 142.11	0.16
Education	28.93	12.24	4.95; 52.91	0.02*
AES-S	-4.36	3.11	-10.46; 1.73	0.16
FSS-7	-24.55	12.87	-49.78; 0.68	0.06
Ankle dorsal flexor muscle strength (stationary dynamometry)	-0.47	1.40	-3.21; 2.27	0.74

*Statistically significant. AES-S=Apathy Evaluation Scale (Self-rated). FSS-7=Fatigue Severity Scale (7 items). This figure is from Knak et al. (59).

The compliance for wearing the accelerometer for at least 4 days was deemed acceptable for both PwDM1 and healthy controls (Figure 50). Only marital status for PwDM1 ($p = 0.009$) and age for healthy controls ($p = 0.014$) differed statistically significant between the participants who adhered to wear time criteria and the participants who did not adhere to wear time criteria (Figure 50) (59). Age ($p = 0.053$), BMI ($p = 0.53$), education ($p = 0.08$), apathy ($p = 0.49$), fatigue ($p = 0.31$) and ankle dorsal flexor muscle strength ($p = 0.91$) did not differ between the two groups for PwDM1. There was no difference in BMI ($p = 0.70$), marital status ($p = 0.47$) and education ($p = 0.25$) between the two groups for healthy controls.

Figure 50: Compliance barometer of percentage of PwDM1 and healthy controls fulfilling the wear time criteria for accelerometery (Study III)



10. DISCUSSION

10.1 Studies I-III: Main findings

In this thesis, a main finding was that HHD is an adequate alternative to stationary dynamometry for muscle strength testing in a non-congenital cohort of PwDM1. The HHD was more reliable, equally valid (except for the extensor muscles on a group level), and similarly responsive compared to the stationary dynamometry. Considering the assessment of balance, the main finding was that the dynamic balance measurements, in general, demonstrated superior reliability, validity and responsiveness compared to the static balance measurements in PwDM1. All static balance tests were flawed by ceiling- or floor effects. For functional mobility assessments, both the STS and the 10mWT were adequately valid and reliable, but unable to detect change in PwDM1 in the present time frame. The 10mWT was, however, most reliable. A simplified evaluation of all psychometric properties for each outcome measure is provided in Table 5.
This prospective, large study of objective and subjective physical activity monitoring in PwDM1 showed that PwDM1 were less physically active compared to healthy controls. In addition, the present novel study of physical activity predictors showed that the physical activity level was only predicted by educational level in PwDM1.

Table 5: Evaluation of the psychometric properties of muscle strength, balance, and functional mobility measurements in PwDM1 (Studies I-II)

	RELIABILITY			VALIDITY		RESPONSIVENESS		
	Systematic error	Random error		Criterion or construct		Distribution	Anchor	
		Group	Single	Group	Single		AUC mean	AUC 95% CI
Muscle strength								
A. Stationary dynamometry								
Ankle plantar flexors	✓	✗	✗	✓	✓	✗	✓	✓
Ankle dorsal flexors	✓	✓	✗	✓	✓	✗	✗	✗
Knee extensors	✓	✓	✗	✓	✓	✓	✗	✓
Knee flexors	✓	✓	✗	✗	✓	✓	✗	✓
Hip extensors	✗	✓	✗	✓	✓	✓	✗	✓
Hip flexors	✗	✓	✗	✓	✓	✓	✗	✗
B. HHD								
Ankle plantar flexors	✓	✓	✗	✗	✓	✓	✗	✓
Ankle dorsal flexors	✓	✓	✓	✓	✓	✗	✗	✓
Knee extensors	✓	✓	✓	✗	✓	✗	✗	✗
Knee flexors	✓	✓	✓	✓	✓	✓ ✗ - outliers	✗	✓
Hip extensors	✓	✓	✗	✗	✓	✗	✗	✓
Hip flexors	✓	✓	✗	✓	✓	✓	✗	✓
Balance								
TUG	✗	✓	✓	✓		✓	✗	✓
Step test	✗	✓	✓	✓		✗	✗	✗
m-CTSIB	✓	✓	✗	✗		✓	✗	✓
Feet together stance	N.A.	N.A.	N.A.	✗		✗	N.A.	N.A.
Tandem stance	N.A.	N.A.	N.A.	✗		✗	N.A.	N.A.
One-leg-stance eyes open	N.A.	N.A.	N.A.	✗		✗	N.A.	N.A.
One-leg-stance eyes closed	✓	✗	✗	✗		✗	✗	✓
Functional mobility								
STS	✗	✓	✗	✓		✗	✗	✓
10mWT	✓	✓	✓	✓		✗	✓	✓

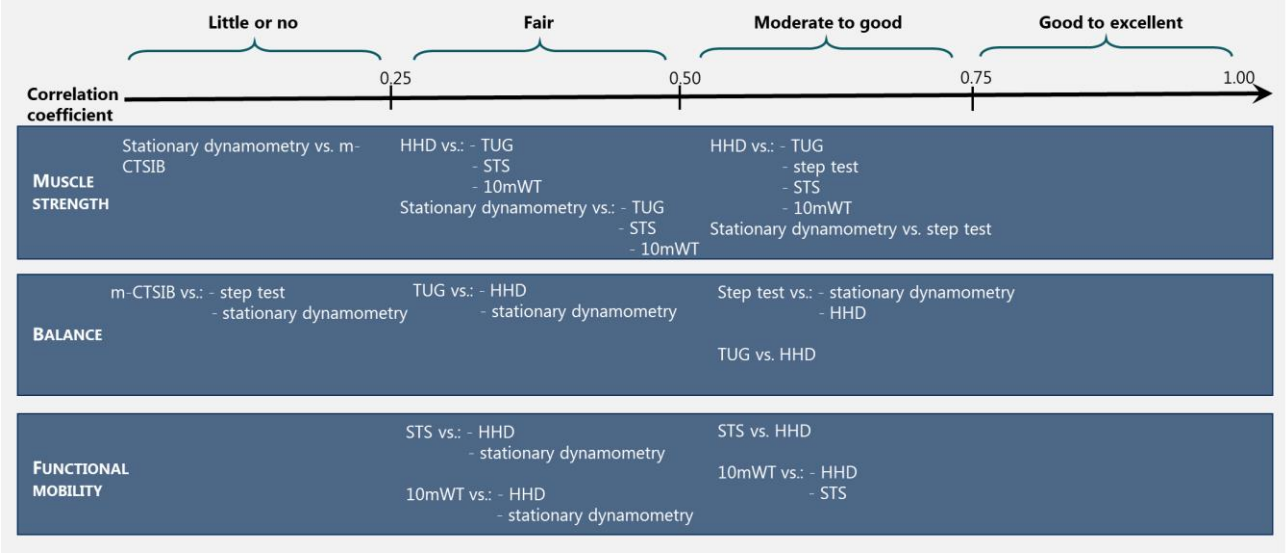
10.2 Studies I-II: Outcome measures

10.2.1 Construct validity challenges

For the interpretation of the validity of the outcome measures (Study I), it is important to understand that validity is limited by the poorest test of the compared tests (75) and by the reliability of the test (17). For the balance and functional mobility assessments where no direct comparable gold standard exists, the validity

statistic is limited to construct validity, which is difficult to interpret. Even small to moderate regression- and correlation coefficients may be reasonable for construct validity. This is reasoned by the comparison of tests, which shares some common elements and, therefore, high correlation is less likely. Figure 51 visualizes the challenge of reaching high construct correlations (good to excellent correlation) of muscle strength-, balance- and functional mobility measurements in PwDM1 based on the present- (57) and previous studies (2,24). Moreover, larger sample sizes reduce the correlation coefficients (72), and because the present sample of 78 participants is considered large, the correlation coefficients in the present study are reduced.

Figure 51: Construct validity challenges (Study I)

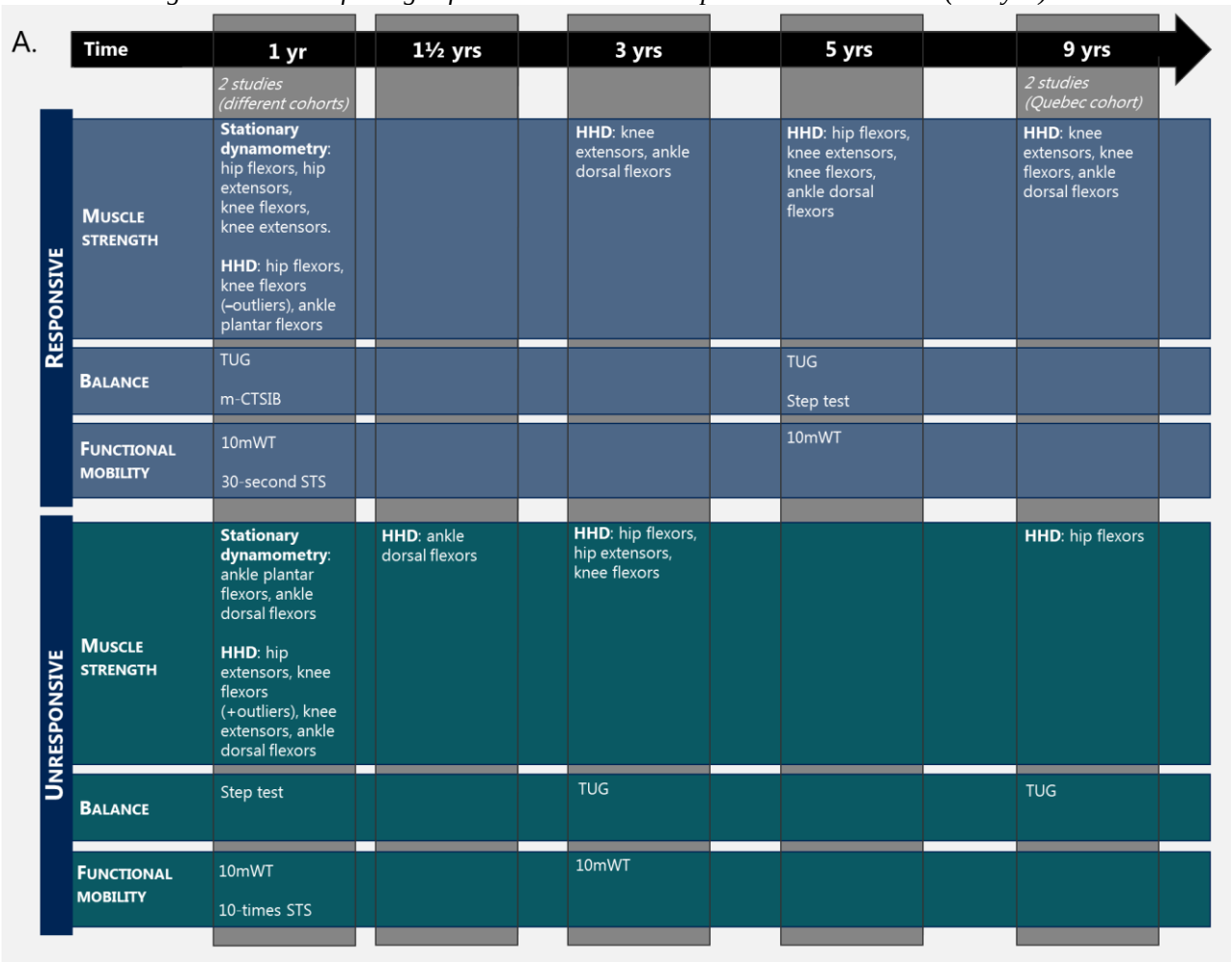


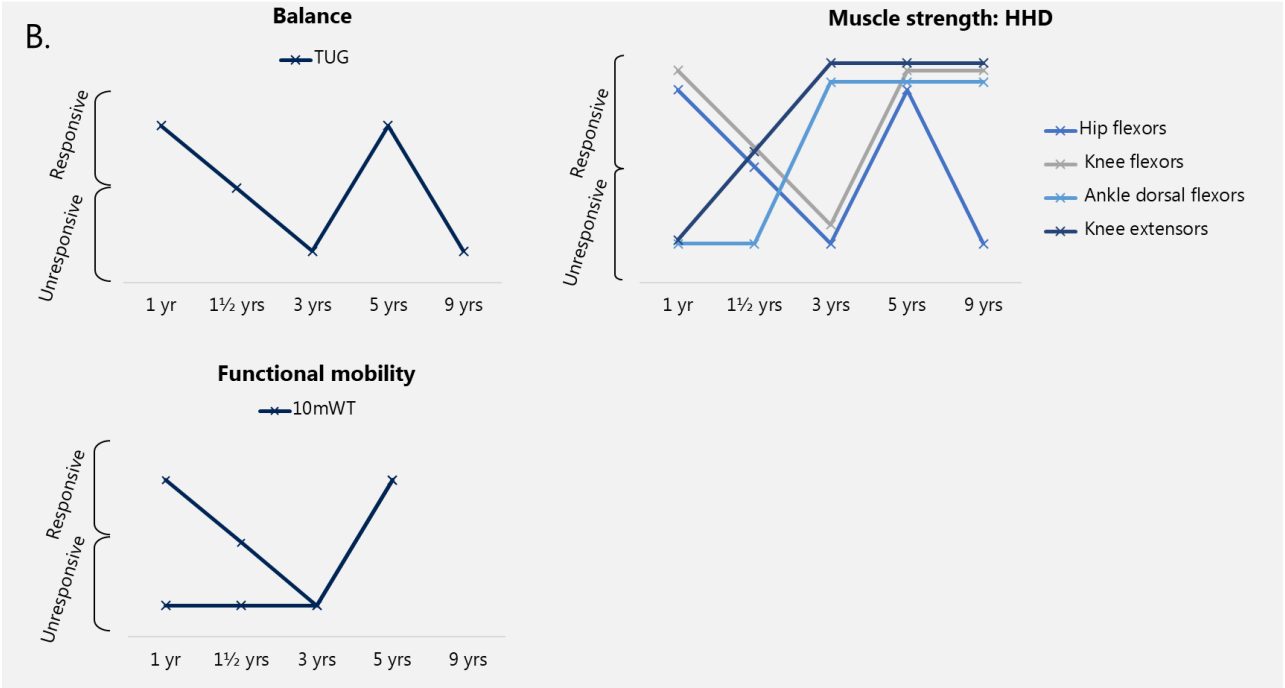
10.2.2 Responsiveness challenges

The possible impaired mental functions in PwDM1 may compromise the findings of the psychometric properties in Studies I-II. The mental impairments cover apathy, lack of insight, attention deficit, demotivation and cognitive dysfunction. It may especially problematize the application of the patient-reported GRS as an anchor to evaluate the responsiveness of the outcome measures within muscle strength, balance and functional mobility. Kierkegaard et al. (23) has previously introduced the limitations of the use of GRS in PwDM1 presenting with mental deficits. Therefore, cognitive impairments and apathy were screened for in the present studies. Depending on the specific cognitive ability, cognitive deficits were present in 5-38% of the PwDM1. Despite variation in apathy scores, clinical apathy according to the Apathy Evaluation Scale (77) was not detected. However, apathy may be underestimated in the present study using the patient-rated version of the AES in contrast to informant- or clinician-rated versions of AES as indicated by a previous study in PwDM1 (78). Unexpectedly, the responsiveness findings indicated that better cognition and lower apathy score were associated with disagreements between the objective measurements of change versus the patient-reported perception of change. Cognitive impairment may, thus, not violate measurements in PwDM1 as much as expected. Moreover, deficits in the mental aspects are inherently associated with the DM1-disease and should be acknowledged. It has previously been suggested that homogeneity improves the chance of a successful response to changes in a variable over time (79). A more homogeneous sample in the present studies with exclusion of PwDM1 who presented with dysfunction on the mental parameters would, therefore, probably improve the psychometric properties of the outcome measures, especially responsiveness. However, such

exclusion would compromise the generalizability to the broader DM1-population. In addition to the mental-dysfunction-shortcomings of the patient-reported GRS, other challenges concerning responsiveness are the slow nature of the progression in DM1 disease and the non-DM1 specific challenges such as recall bias (the memory differ from reality) and response-shift bias (adaptation to the new situation and adjustment of expectations), which previously has been stressed (23). A paradox of response-shift has previously been documented in PwDM1 (80) with findings of improved Quality of Life but deteriorated muscle strength at follow-up. These shortcomings may partly account for the overall poor agreement between the objective outcome measures and the corresponding subjective GRS perceptions after 1 year in Study II, because the objective outcomes are anchor-dependent and cannot be better than the anchor. Besides, the findings of agreement between the objective and subjective outcomes should be viewed with caution due to the wide 95% confidence intervals ranging from no agreement to different levels of agreement. Certain conclusions about the anchor-based responsiveness is, therefore, not established. The subjective GRS was selected as the anchor to evaluate responsiveness because of the unavailability of objective anchors to define true change. Nevertheless, a clinician-rated GRS can be considered as a substitute to the patient-rated GRS to account for the above-mentioned obstacles with patient-reporting in PwDM1. The patients' own perspectives should, however, still be recognized (23). Moreover, the clinician-rated GRS may not necessarily differ from the patient-rated version since a previous report in PwDM1 (81) stated that an apathy score was similar for the clinician and the patient-rated versions. The AUC-values differed between relative and absolute changes in Study II. In accordance with Kierkegaard et al. (23), there was a tendency of higher AUC-values for the relative changes compared to absolute changes. This supports the general use of relative changes for follow-up measurements. Responsiveness has proven challenging after follow-up of 1 year in PwDM1 due to the unavailability of objective anchors to define true change and due to the slow decline of the DM1 disease. The non-significant changes after 1 year in the present study may either be caused by no genuine change within this short timeframe or by the tools' inability to detect subtle changes. The measurements may, however, still be responsive to interventional effects within 1 year or to larger natural history changes, which often presupposes a longer follow-up. The latter suggestion is supported by the finding of responsiveness of muscle strength, balance and functional mobility measurements in long-term follow-up studies in PwDM1 (23,24,30,82,83). However, as shown in Figure 52 outcome measures within muscle strength, balance and functional mobility are not consistently more responsive over time (22–24,30,58,82,84).

Figure 52: Mixed findings of distribution-based responsiveness over time (Study II)

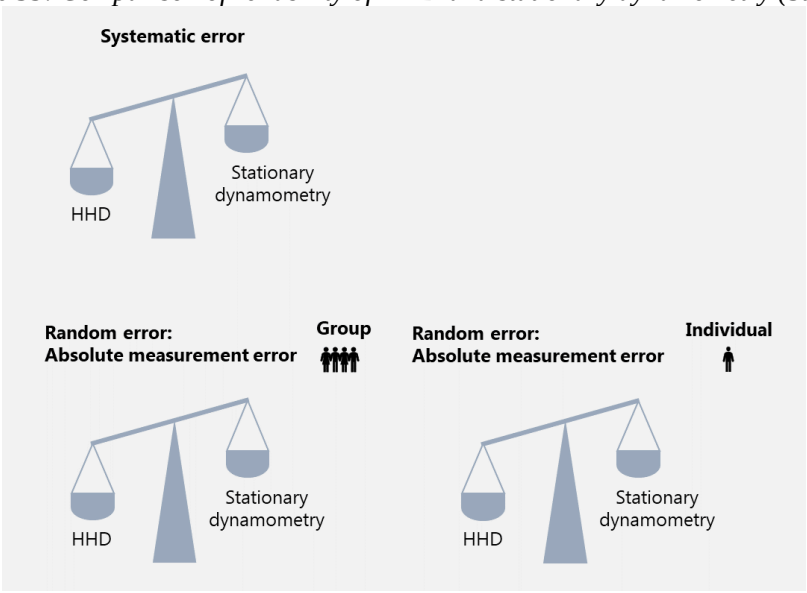




10.2.3 Muscle strength measurements

The unforeseen, positive psychometric findings of the HHD, compared to the stationary dynamometry (Figure 53-55), is likely influenced by the assessors' standardized, fixed position for the measurements with the HHD, which reduced the assessor-dependency. This finding is in line with a previous report in PwDM1 (25), which suggested that stronger muscle groups are not necessarily assessor-limited by larger measurement errors as long as suitable fixation is applied for the HHD measurements. Hence, appropriate fixation when using HHD is strongly recommended.

Figure 53: Comparison of reliability of HHD and stationary dynamometry (Study I)

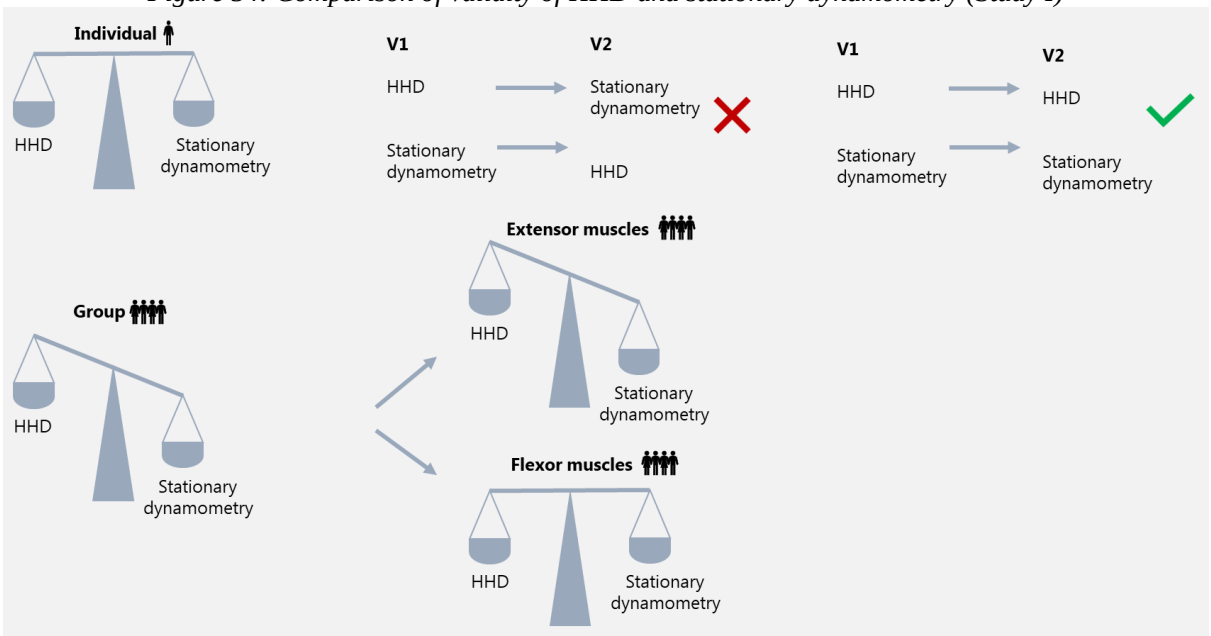


The heaviest weight scale = the best psychometric property.

Regarding absolute reliability, the measurement errors were acceptable for both the HHD and the stationary dynamometry on a group level ($SEM_{\%} \leq 15\%$) and on an individual level ($MDD_{95\%} \leq 30\%$, with exclusion of outliers). These findings are overall consistent with the previously established SEM-values for stationary dynamometry and HHD in PwDM1 (25,28) and healthy individuals (69,85–88) and also for the MDD-values for HHD in PwDM1 (85–91) and for stationary dynamometry in healthy individuals (69). For both muscle strength devices, the hip- and ankle muscles were associated with larger individual measurement errors than the knee muscles. This is likely attributed to the more unaccustomed positions with the hip- and ankle assessments. In line with this, the MDD-values exceeded the MCID-values mostly for the hip muscles for both muscle strength devices, and extremely for the ankle plantar flexors using stationary dynamometry. For the remaining muscle groups, the difference between the MDD-values and the MCID-values were not remarkable when outliers were excluded. Only the HHD ankle dorsal flexors demonstrated MDD-values equal to the MCID-values. Hence, for both devices, individual clinical important alterations in muscle strength may overall be hidden. Nevertheless, changes close to the MCID-values may be captured for the knee muscles and most of the ankle muscles. Concerning reliability as to systematic bias, a statistically significant learning effect was only demonstrated in the hip muscle groups using stationary dynamometry. The lack of systematic bias in the knee extensors with both stationary dynamometry and HHD is consistent with a previous study of the knee extensors using stationary dynamometry in PwDM1 (25).

For criterion validity on an individual level, none of the muscle strength devices were superior to another. Thus, muscle strength measurement in a single individual in the clinic can either be performed with the HHD or the stationary dynamometry as long as the same muscle strength device is used for repeated measurements (Figure 54). This finding is in accordance with a previous study in healthy people (86), which found that neither the belt-stabilized HHD nor the stationary dynamometry was superior for single individuals. Concerning criterion validity on a group level, the torque was only significantly higher for the extensor muscles with the stationary dynamometry compared to the HHD (Figure 54). This finding is consistent with a previous publication in PwDM1 (25), which showed that the stationary dynamometry captured a higher torque than the HHD for the knee extensors. The lower torque captured by the HHD in the stronger extensor muscles is possibly assigned to assessor-dependency. A study in healthy individuals (85) showed a higher muscle force with an additional fixation device, which suggests that the assessor-dependency can be counteracted by appropriate fixation. Fixation with straps may, therefore, improve the criterion validity of the HHD for the extensor muscles in a group of PwDM1.

Figure 54: Comparison of validity of HHD and stationary dynamometry (Study I)



The heaviest weight scale = the best psychometric property.

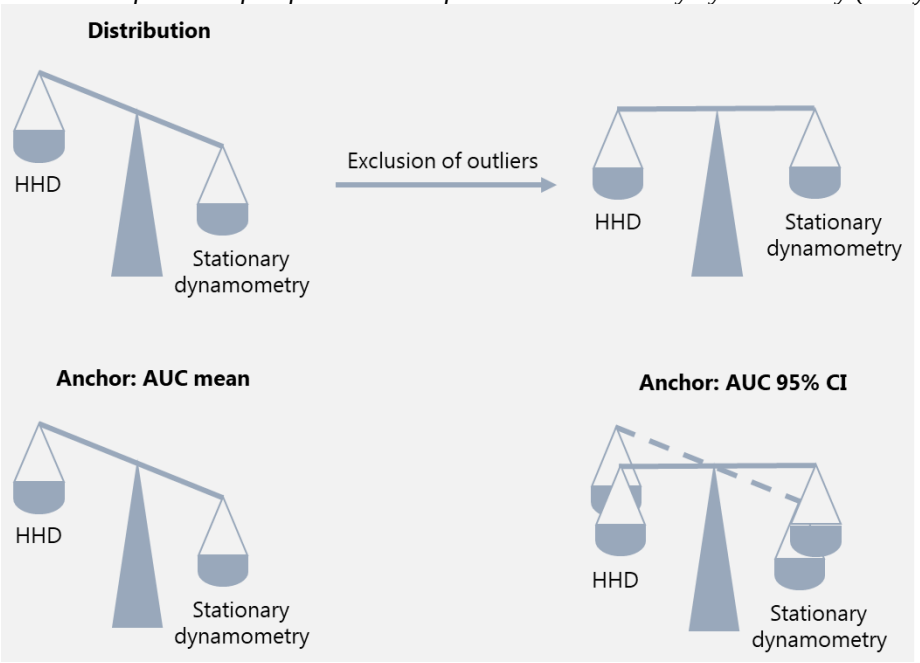
For distribution-based responsiveness, the stationary dynamometry detected significant changes in more muscle groups than the HHD after 1 year (Figure 55). This difference was, however, minor considering the exclusion of outliers in the knee flexors and the almost significant *P*-value of 0.065 in the hip extensors with the HHD. Both muscle strength devices captured a change in the proximal muscles but not in the distal ankle dorsal flexors. This is likely attributed to the large proportion (21%) of the participants with weak ankle dorsal flexors unable to activate the stationary dynamometer, and, thus, were unresponsive to detect deteriorations. This floor effect was not seen with the HHD, hence, the HHD is strengthened by the capability of monitoring very weak muscle strength in PwDM1. Albeit proximal muscle affection is less pronounced than the distal muscles in PwDM1, proximal weakness of MIRS ≥ 4 was present in 76% of the PwDM1 in the present studies. Responsiveness of the dynamometers in the proximal muscles was therefore possible. Regarding the stationary dynamometry, a significantly higher torque was registered in the hip flexors at 1-year follow-up.

This improvement was unexpected but consistent with a previous study of PwDM1 (23), which demonstrated a median increase in the hip flexor strength using HHD after 9 years. In contrast, a deterioration of the hip flexors measured by HHD has been shown in both the present study and in a previous study of PwDM1 after 5 years (30). Considering this and the fact that the physical activity level was not increased at follow-up in the present study, the detected increase in the hip flexors with stationary dynamometry may be false. The stationary dynamometry registered a change in all hip- and knee muscles, which contrasts with a previous 1-year follow-up study in PwDM1 (26), which did not detect a change in the knee flexors using stationary dynamometry. The inconsistent findings may be attributed to different DM-types, isometric versus isokinetic muscle work, but especially sample size. The smaller sample size in Lindeman et al. (26) increases the risk of a false non-statistically significant finding.

Concerning the HHD, the HHD captured a change in muscle strength in the hip-, knee- (without outliers) and ankle plantar flexors after 1 year. Previous and recent findings of responsiveness in the knee flexors using HHD (22,23,30,82) are mixed with the longest follow-up studies (22,23,30) detecting a change and the shorter 3-years follow-up study (82) detecting no change. Thus, detection of change in the knee flexors using HHD is possible at short-term follow-up but more likely at long-term follow-up. The previous and recent findings of the hip flexors using HHD (22,23,30,82) are also mixed, but for this muscle group there was no change after 3-(82) and 9 years (22,23) but a change in between this time interval of 5 years (30). Hence, no pattern for the hip flexor strength is seen regarding follow-up time. Among the present study and previous/recent studies (22,23,30,82), the sample sizes varied but likely only Roussel et al. (82) had a sample size that risks being underpowered. Except for Roussel et al. (82), the non-responsive findings of the HHD hip flexors were unexpectedly shown in the long-term Quebec cohort after 9 years (22,23) whereas the shorter follow-up studies of 1 year in the present study and after 5 years in a previous study (30) demonstrated responsiveness in this muscle group. This discrepancy may partly be explained by the different DM1 phenotypes and degree of muscle impairment. The non-responsive outcomes (22,23) were found in PwDM1 with adult or late-onset phenotypes and milder muscle impairments (42-44% of the participants presenting with mild MIRS 1-3). In contrast, the responsive findings in the present and previous study (30) were shown in PwDM1 with non-congenital phenotypes, including the juvenile onset, who demonstrated slightly more severe muscle impairments (24-37% of the PwDM1 presenting with mild MIRS 1-3). Hence, responsiveness of the HHD hip flexors, which are generally less affected than the more distal muscle groups, may be better in PwDM1 with more severe muscle impairments due to progressed disease status or more affected phenotypes. Finally, the unaccustomed position of the HHD hip flexors may be a factor for the mixed results. In the present study, the HHD did not register a change in the hip- and knee extensors and in the ankle dorsal flexors after 1 year. These nonsignificant findings in the knee extensors and ankle dorsal flexors after a relatively short time is comparable with previous HHD-findings in a 1½-year follow-up study in PwDM1 (84). Long-term follow-up studies in PwDM1 of 3-years (82), 5-years (30) and 9-years (22,23) have shown that the HHD responded to change in these muscle groups. This suggests that the timeframe of follow-up is critical in PwDM1; 1-year of follow-up is possibly too short to respond to change in the knee extensors and ankle dorsal flexors in PwDM1 using HHD, and a prolonged follow-up time is likely needed. Previous findings in PwDM1 (92) suggested that MRI is responsive to muscle changes. However, it remains to be established whether MRI is more responsive

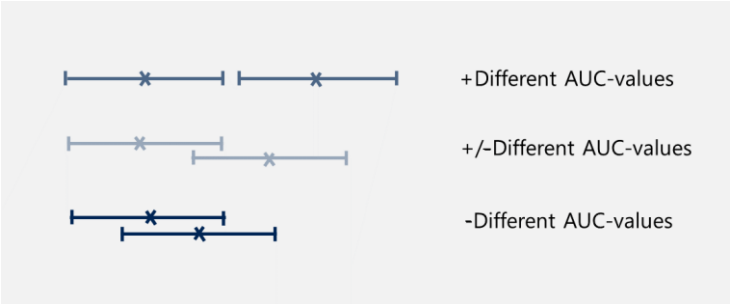
than quantitative muscle testing to evaluate muscle changes in PwDM1. Moreover, MRI is less feasible and does not measure muscle function in contrast to the quantitative muscle testing. Concerning the anchor-based responsiveness, comparison of the HHD and the stationary dynamometry showed that only the ankle plantar flexors using stationary dynamometry demonstrated a mean AUC ≥ 0.7. However, the true AUC may exceed the AUC threshold for both devices considering the 95% CI, except the ankle dorsal flexors and hip flexors with stationary dynamometry and the knee extensors with the HHD. The anchor-based responsiveness may be significantly superior for the stationary dynamometry for the ankle plantar flexors, but there was no significant difference between the stationary dynamometry and the HHD for the remaining muscle groups (Figure 55). A guideline for comparison of AUC-values is visualized in Figure 56. Compared to previous findings of anchor-based responsiveness of a HHD lower limb composite score after 9 years in PwDM1 (AUC mean, 95% CI; 0.60, 0.50-.070 for % and 0.70, 0.60-0.80 for Nm) (23), there was no significant difference between the AUC-values, except for the knee extensors which may be less responsiveness in the present study. Perception of change is most likely easier for larger changes, which can be expected after longer periods such as in a previous study (23). Anchor-based responsiveness may, hence, be improved in long-term follow-up studies, although this has not been clearly demonstrated.

Figure 55: Comparison of responsiveness of HHD and stationary dynamometry (Study II)



The heaviest weight scale = the best psychometric property.

Figure 56: Comparison of AUC-values (Study II)



One line represents one device or study; cross = AUC mean; line = AUC 95% CI.

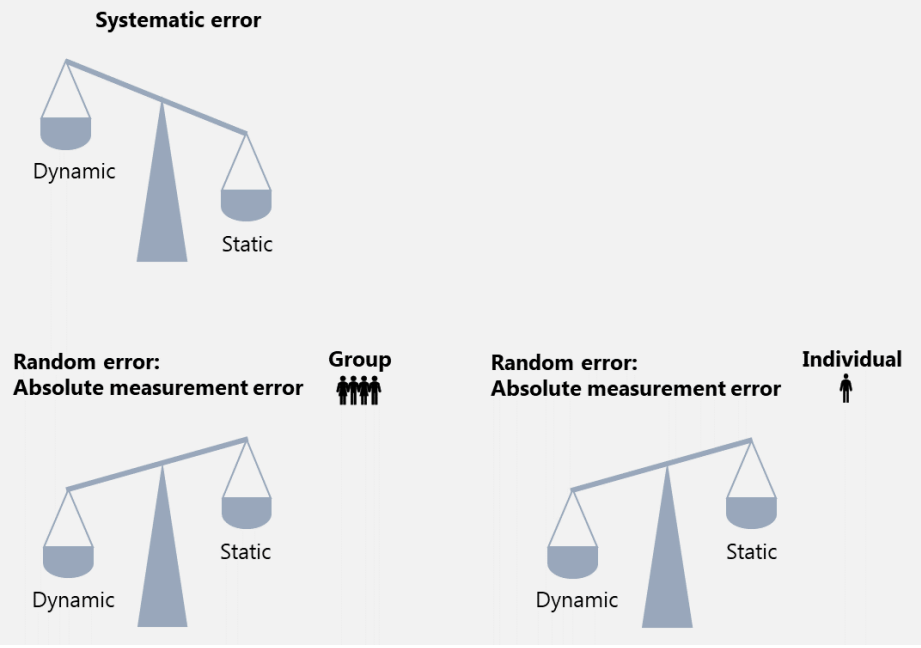
Regarding choice of muscle group for muscle strength assessments, measurement of the hip muscles has proven troublesome with both muscle strength devices. Overall, the hip muscles were associated with the highest absolute measurement errors (both muscle strength devices), a learning effect (stationary dynamometry), a spurious finding of increased hip flexor strength at follow-up (stationary dynamometry) and inconsistent responsive findings in the literature with no clear pattern over time (HHD). Thus, muscle strength assessment of the ankle- and knee muscles is recommended in preference to the hip muscles using either muscle strength device.

The level of evidence and the certainty of the present psychometric findings are in summary overall strengthened by the agreement with the literature regarding absolute reliability for both muscle strength devices and criterion validity for the knee extensors. The mixed findings with no clear pattern of responsiveness over time for the HHD knee flexors, and especially for the hip flexors, introduce uncertainty.

10.2.4 Balance measurements

First, the reliability findings will be discussed. The dynamic balance tests were more reliable, as to absolute measurement error, than the static balance tests for both groups and single individuals (Figure 57). Although, the discrepancy to the static m-CTSIB was less pronounced on a group level.

Figure 57: Comparison of reliability of dynamic- and static balance measurements (Study I)



The heaviest weight scale = the best psychometric property.

Despite that both the static and the dynamic balance assessments measure postural control, the different degrees of absolute measurement error may partly be reasoned by the different requirements for the two balance categories. The dynamic- and static balance assessments differ in the sense that the dynamic balance tests represent more complex functions, whereas the static balance tests can be considered as more "pure" balance outcomes. For static balance, imbalance is first corrected by an ankle strategy using the ankle dorsal- and plantar flexor muscles (93). The impact of the predominantly distal muscle weakness in PwDM1 may, therefore, be more pronounced in the static positions compared to the dynamic transitions where the step

strategy is used to regain balance, which also relies on the proximal muscle groups in the lower limb. In addition, the postural control requirements and the level of difficulty are different for the static- and dynamic balance assessments. Vision-, vestibular- and somatosensory (proprioceptive, cutaneous and joint receptors) inputs are available for the dynamic balance tests. In contrast, the visual cue is eliminated in the static balance tests with eyes closed, and the somatosensory input is disturbed in the static stance on a foam surface as part of the m-CTSIB (93). The challenging static one-leg-stance with eyes closed (no visual cue and a small base of support) demonstrated an extreme amount of measurement error and was less reliable than the less demanding dynamic balance tests. Likewise, the high balance requirements of the m-CTSIB subpart with standing on a foam surface with eyes closed were possibly responsible for the less reliability of the overall m-CTSIB on a balance platform compared to the dynamic balance tests. The described postural control theory and the present findings are supported by a previous study in healthy, elderly individuals (94), which suggested that increased task difficulty yields larger balance measurement errors. It is also possible that the high sensitivity of the balance platform, which the m-CTSIB was conducted on, limits the ability to achieve a high degree of reliability because even subtle changes are registered. However, this high sensitivity is advantageous to detect minor balance impairments that is undetected with balance tests monitoring time standing in a position. Albeit the m-CTSIB was less reliable than the dynamic TUG and step test, the measurement error on a group level was low ($SEM_{\%} \leq 15\%$) and close to reach the acceptable cut-off of $MDD_{95\%} \leq 30\%$ on an individual level when outliers were excluded. The m-CTSIB yielded $SEM_{\%}$-values and MDD-values in line with a previous study in elderly (94). The coordinates of the foot position for the m-CTSIB on a balance platform was not recorded in the present study, but the foot position instruction was similar for the repeated measurements. Moreover, consensus on the influence of the foot position remains to be established (95,96). For the dynamic balance measurements, the $MDD_{95\%}$-findings in the present study are overall consistent with previous TUG-findings in PwDM1 (29) and non-PwDM1 (68,97,98) and with findings of the step test in PwDM1 (29) and individuals with hip osteoarthritis (99). Despite the superiority of the dynamic balance tests as to absolute measurement error compared to the static balance tests, the dynamic balance tests were the only balance tests associated with systematic error by the detected learning effects (Figure 57). Hence, at visit 2 the participants benefitted from the experience of the first visit as to the dynamic balance assessments. The systematic error with learning effects may, however, be negligible by practice trials, whereas the random error with absolute measurement error is less adjustable. The learning effect findings are consistent with previous findings for the step test (left foot), but not with findings for the step test (right foot) nor the TUG in PwDM1 (29). Compared to a previous study (29), the present study had a substantially larger sample, which means a larger power and a higher chance of finding a significant difference between repeated measurements. The lack of systematic bias in the static balance tests in the present study is in accordance with a previous study in PwDM1 (29) that showed no statistically significant difference between visits for the feet-together stance, tandem stance and one-leg-stance eyes open and -closed. The MDD exceeded the MCID in all balance measurements but was only pronounced for the static balance tests (one-leg-stance eyes closed and m-CTSIB). Thus, minimal clinical important changes in balance may be concealed, especially for the static balance tests.

Second, the findings of construct validity will be discussed. The correlation coefficients were larger for the dynamic balance tests compared to the static balance tests; thus, the dynamic balance tests are favoured (Figure 58).

Figure 58: Comparison of validity of dynamic- and static balance measurements (Study I)

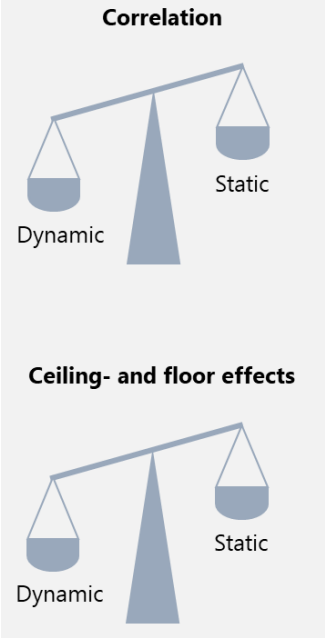
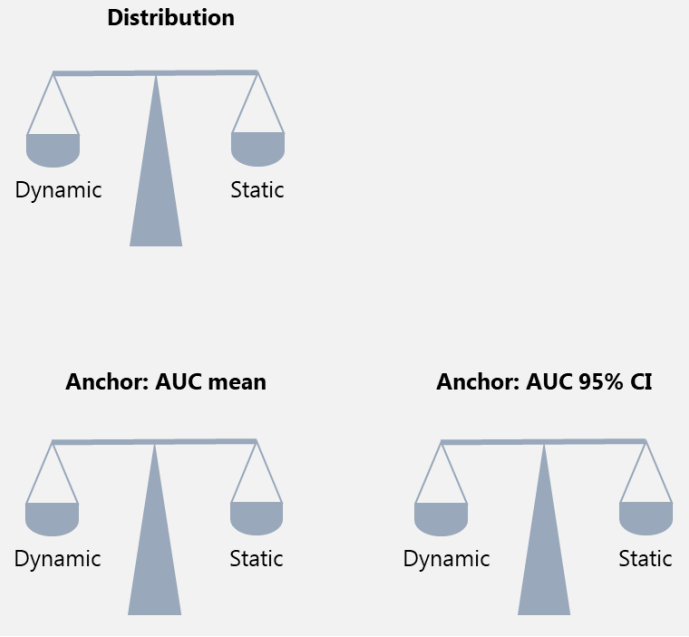


Figure 59: Comparison of responsiveness of dynamic- and static balance measurements (Study II)



The heaviest weight scale = the best psychometric property.

The correlation-findings of the static balance tests are similar to previous findings in PwDM1 (2). However, correlation does not allow a proper comparison across studies due to its sensitivity towards size and heterogeneity of samples (72) and should, therefore, only be done with reservations. Regression has not been reported for the balance tests in PwDM1 in the literature yet. Although the static balance tests covered a wide spectrum of difficulty levels, all static balance tests were flawed by floor effects or predominantly ceiling effects. According to the presented guideline for floor effects, the m-CTSIB showed only a floor effect for visit 1, but the m-CTSIB tended to demonstrate a floor effect for visits 2-3 as well with 13-14% of the PwDM1 being unable to complete the m-CTSIB. The ceiling effect in the feet-together stance and the one-leg-stance eyes open has previously been reported in PwDM1 (29) and in individuals with hip osteoarthritis (99).

Finally, responsiveness will be addressed. Concerning distribution-based responsiveness, the dynamic TUG and the static m-CTSIB monitored a change in balance after 1 year in the present study. Hence, responsiveness was proven for both dynamic- and static balance tests (Figure 59). In accordance, a deterioration in TUG was reported by Kierkegaard et al. (23) after 9 years and by Hammarén et al. (30) after 5 years, but not by Roussel et al. (82) after 3 years. The latter study (82), however, was limited by a small sample. The step test did not detect a change after 1 year as to distribution-based responsiveness, which contrasts with a previous 5-year follow-up study (30) that demonstrated a deterioration in dynamic balance with this test. This suggests that 1 year may be too short to capture changes in dynamic balance by the step test. Regarding anchor-based responsiveness, none of the balance tests were responsive based on the best estimate (AUC mean < 0.70) but all tests, except the step test, may be responsive when the 95% CI is accounted for (95% CI upper limit > 0.70

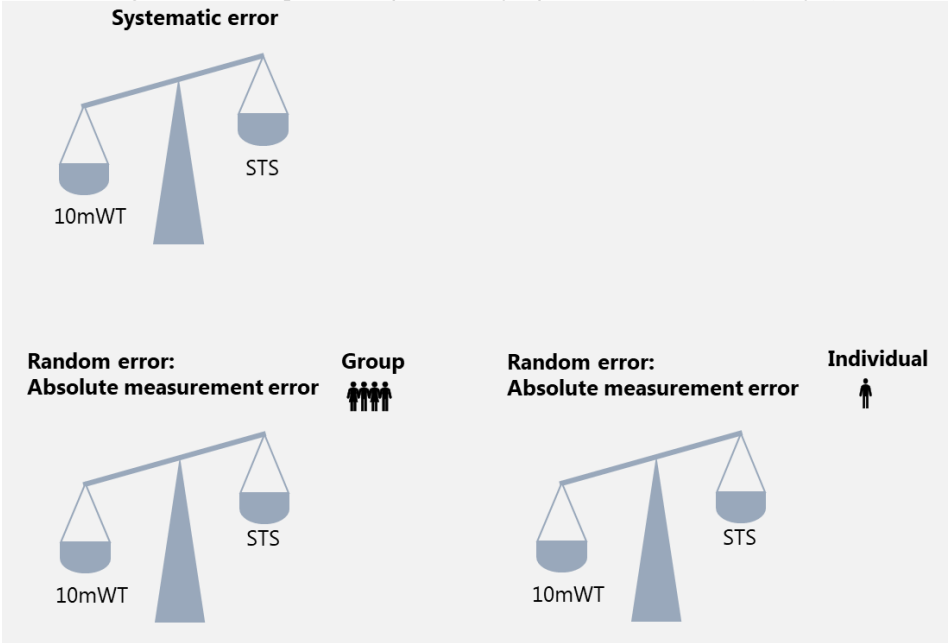
AUC threshold). However, the 95% CI for the step test did almost reach the AUC threshold (AUC upper limit = 0.69). It should be noticed that the 95% confidence interval is remarkably large for the one-leg-stance eyes closed, which introduces high uncertainty for this measurement. There was no significant difference for anchor-based responsiveness among the balance measurements (Figure 59). The anchor-based responsiveness for the TUG (AUC mean, 95% CI; 0.65, 0.51-0.80) did not differ significantly from the TUG finding (0.80, 0.70-0.90) in a previous 9-years follow-up study in PwDM1 (23). However, a tendency of a more pronounced anchor-based responsiveness for the TUG may occur after a longer period.

In summary, compared to the existing literature, the present studies showed consistent findings as to the m-CTSIB (absolute reliability), TUG (absolute reliability), step test (absolute reliability), the static feet-together stance, tandem stance and one-leg-stance eyes open and closed (no systematic measurement error, ceiling effect and construct validity). Hence, certainty of these findings is enhanced, which improves the level of evidence. Systematic measurement error for the TUG is associated with uncertainty due to divergent findings in the literature.

10.2.5 Functional mobility measurements

Concerning reliability, the 10mWT was more reliable than the STS because the STS demonstrated systematic error with a learning effect and showed a higher degree of random, absolute measurement error for both groups and single individuals (Figure 60). Nevertheless, the measurement errors for the STS were low for both groups ($SEM_{\%} \leq 15\%$) and single individuals ($MDD_{95\%} \leq 30\%$ with exclusion of outliers). In addition, familiarization trials may eliminate the learning effect in the STS. The lack of systematic bias for the 10mWT is consistent with a previous finding in PwDM1 (29). The finding of the very low measurement error for groups and single individuals for the 10mWT is generally comparable to previous studies in PwDM1 (29) and non-PwDM1 (98,100). Only the 10mWT showed an MDD-value below the MCID-value and is, hence, able to capture minimal clinical important alterations in functional mobility, which may be concealed by the STS.

Figure 60: Comparison of reliability of 10mWT and STS (Study I)



The heaviest weight scale = the best psychometric property.

Regarding construct validity, the correlations were similar for the 10mWT and the STS (Figure 61). The correlations for the functional mobility tests with other functional mobility measurements, muscle strength measurements and balance measurements ranged from fair to good, which is considered reasonable. Even though the STS correlated fairly (r = -0.34) with the proximal muscle strength, this correlation was lower than expected. The fair correlation between these two features is, however, almost consistent with a previous report on people with rheumatoid arthritis (101) and with a finding of the modified 30-second STS in PwDM1 (24). Thus, the different types of muscle contractions and influencing variables for the STS and the isometric muscle strength measurement may impact the correlation more than anticipated. The 10mWT-correlations with muscle strength and STS are overall in accordance with previous studies in PwDM1 (2,24) and non-PwDM1 (102). Nonetheless, precautions should be made for comparison of correlation across trials. No reference studies exist for regression of functional mobility tests in PwDM1.

Figure 61: Comparison of validity of 10mWT and STS (Study I)

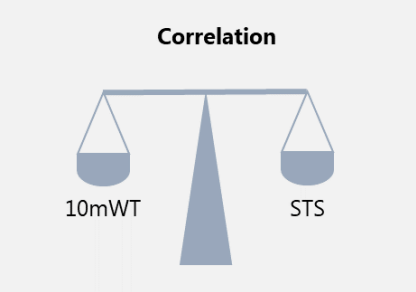
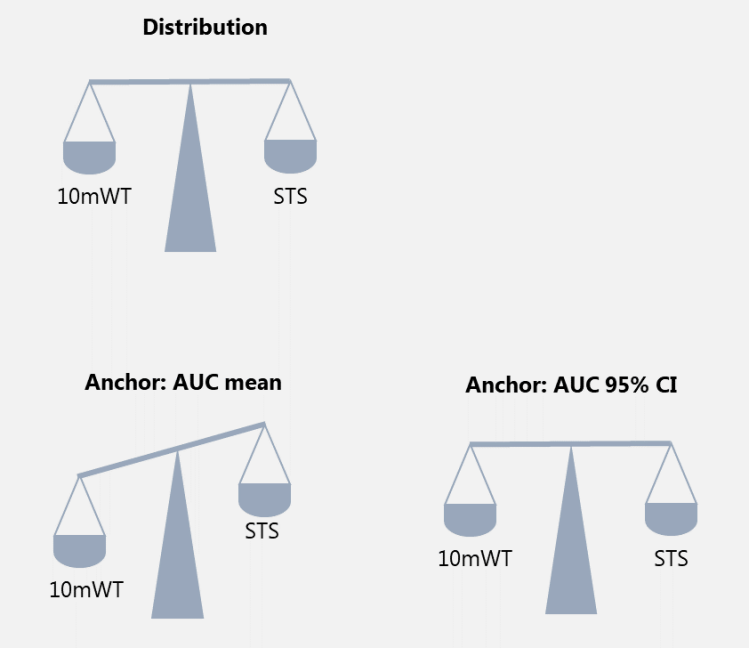


Figure 62: Comparison of responsiveness of 10mWT and STS (Study II)



The heaviest weight scale = the best psychometric property.

Finally, responsiveness will be addressed. Distribution-based responsiveness was unmet and similar for both the 10mWT and the STS after 1 year (Figure 62). These unresponsive findings contrast with previous findings of distribution-based responsiveness in PwDM1 (24,30,83). Possible explanations are different methodology compared to the previous studies (24,83), a larger sample size (24), which enhances the power and increases the chance of detecting a change, lack of DM-type classification (83), and a longer follow-up period (30), which yields larger changes that are more likely detected. However, a subsequent 3-years follow-up study in PwDM1 (82) did similarly not monitor a change in the 10mWT. Concerning anchor-based responsiveness, the 10mWT was responsive (AUC mean, 95% CI; 0.71, 0.52-0.89) based on the best estimate, but did not differ significantly from the STS (0.60, 0.40-0.80) (Figure 62). Anchor-based responsiveness has not previously been documented for these functional mobility tests.

Regarding level of evidence, the present findings concur with the previous findings concerning the 10mWT (no systematic error, absolute reliability, construct validity) and the STS (construct validity), which increases

the credibility of these findings. The certainty for responsiveness of the 10mWT is compromised by divergent findings with the same follow-up time.

10.3 Study III: Physical activity

The finding of a lower physical activity level objectively and subjectively in PwDM1 compared to healthy controls agrees with former objective findings in PwDM1 using a healthy control group as comparator (55,56). Likewise, the findings support a previous finding of decreased subjective physical activity level in PwDM1 (52), although this study was not controlled by a healthy group as comparator. The discrepancy in physical activity level between PwDM1 and healthy controls in the present study is not caused by any difference in demographic data as we adjusted for this confounder in the statistical analyses. Instead, the difference is probably attributed to the DM1 traits such as mental affection, fatigue and muscle wasting. However, apathy did unanticipatedly not predict physical activity in PwDM1, but this result should be interpreted with caution because, despite the range in apathy scores in the PwDM1, clinical apathy (apathy score > 34 (77)) was not present in these participants. In contrast, education (which may partly reflect cognition in some people) was a significant predictor of the physical activity level in the PwDM1 of whom 42% presented with cognitive deficits. This finding do not concur with the present findings in healthy controls but agrees with previous findings in healthy people as to predictors of physical activity and factors associated with self-perception of barriers to physical activity (103,104). Fatigue was present in 59% of the PwDM1 and tended to predict the physical activity level in PwDM1 ($P = 0.06$). Seventy-eight percent of the PwDM1 suffered from muscle impairments but the ankle dorsal flexors were not significant predictors of physical activity in PwDM1. This result may, however, vary if more lower limb muscles are accounted for. The BMI of the PwDM1 in the present study ranged from underweight, normal weight to overweight (IQR is only shown in the demographic table, Table 1). The present study in PwDM1 and a previous study in healthy people (103) found consistently that BMI is not a predictor of physical activity in these groups of people, but BMI has been shown to correlate with physical activity in healthy individuals (105). Marital status did not predict physical activity in PwDM1 in the present study, but marital status has been shown to be a predictor of perceived barriers to physical activity in healthy people (104). Age tended to be a significant predictor of physical activity in the PwDM1 ($P = 0.07$) in the present study. In terms of enhancing the physical activity level in PwDM1, the predictor(s) of physical activity should be addressed. Educational level is not pragmatically changeable but an effort to promote the physical activity level in the less-educated PwDM1 is stressed. Fatigue should likely also be addressed in PwDM1 due to the tendency of fatigue predicting the physical activity level in PwDM1. Paradoxically, exercise has previously been shown effective in reducing fatigue in these people (51). Thus, fatigue and physical activity seems to be circular interacted. Moreover, exercise has generally shown encouraging findings with no harmful effects in PwDM1 (30,42,45,47,51), even though solid evidence remains to be established (25,37,38). The engagement in physical activity by PwDM1 has, however, proven problematic with identified obstacles such as fatigue and physical impairments (52). Cognitive behavioural therapy has been shown to improve the capacity for activity and participation of daily and social activities in PwDM1 based on the DM1-

Active-c questionnaire in a multi-centre RCT (48). Hence, it is plausible that cognitive behavioural therapy may also promote physical activity in PwDM1, but this has not been investigated.

In addition to the lower physical activity level in PwDM1 compared to healthy controls, the PwDM1 adhered poorly to the WHO recommendations on all physical activity intensities. The recommendation of physical activity of ≥ 150 min of moderate intensity per week (106) were overall only adhered to by 50% of the PwDM1 and the proportion was lower for the vigorous intensities. These findings are problematic both from a general- and a DM1-disease specific perspective. As to the first perspective, the physical inactivity finding is troublesome because of the higher risk of life-style diseases with physical inactivity in the general non-specific DM1 population (106). Concerning the latter perspective, physical inactivity is undesirable because it may aggravate the muscle atrophy and further impair the physical functioning in PwDM1. Due to the higher risk of developing diabetes mellitus in PwDM1 compared to the background population (6), the importance of physical activity is stressed.

The strengths of the present study are that physical activity was monitored objectively with duration and intensity of hip movement and was not limited to steps only. Moreover, physical activity was monitored for a representative period of 1 week. The limitations are the risk of misclassifications. First, the accelerometer device did not tolerate water and, hence, swimming was not registered. Secondly, heart rate was not monitored in the participants, which means that activities that did not involve hip movement was not recorded. Heart rate collection was deselected because some PwDM1 suffer from heart arrhythmia, which would violate the heart rate data. A more comprehensive data collection with heart rate monitoring in a subgroup of PwDM1 with normal heart rhythm could be a subject for future investigation. Nevertheless, since the subjective IPAQ addressed all movements and supported the objective findings, the conclusion of decreased physical activity in PwDM1 is likely unaffected by these shortcomings. Thirdly, sedentary behaviour with no acceleration may erroneously be registered as none-wear by the applied Troiano reference (62), but since small movements often appear when being “sedentary”, the risk may be minor. There is a risk that the physical activity level is overestimated in both the PwDM1 and the healthy controls in the present study, because the PwDM1 and the healthy controls who were not compliant with the wear time criteria of the accelerometer may manifest a lower physical activity level. Nonetheless, the impact may be minor because the proportion of non-compliant participants was only up to 12% of the participants. Furthermore, the group who adhered to wear time criteria versus the group who did not were similar as to the established predictors of physical activity in PwDM1 in the present study. The physical activity level may also be overestimated by social desirability bias, but this bias is less likely with the data collection period of 7 days. The social desirability bias is not considered to impact the results of the difference between the PwDM1 and the healthy controls because this bias has most likely influenced both groups equally. Wear time criteria were better fulfilled by the PwDM1 with cohabitants, and the oldest PwDM1 tended to adhere better to wear time criteria. Thus, the objective physical activity data may be less generalizable to younger PwDM1 and to PwDM1 without cohabitants. There was a tendency that PwDM1 were primarily tested during winter and the healthy controls during summer. Thus, it is possible that a seasonal effect influences the results since people are generally less physically active during winter.

10.4 Studies I-III: Generalizability

The findings from Studies I-III are overall considered generalizable to the majority of the non-congenital PwDM1-population because of the large cohort, few dropouts (max. 14% of the participants), heterogenous sample, and the recruitment from different geographic areas. The findings may, however, not be representative for the PwDM1 closer to the age of 60 years, since the decliners were significantly older compared to the non-decliners.

The present studies pooled the juvenile-, adult- and late-onset DM1 phenotypes into a non-congenital sample of PwDM1. This pooling was applied because no consensus seems to exist for the DM1-phenotype classification (11,107–109), except that all agree on using age of onset for classification. Moreover, the congenital phenotype is probably the only phenotype that differs markedly from the other phenotypes. The non-congenital phenotypes may, thus, be difficult to discriminate correctly, which introduces a classification bias. Based on these reflections, we defined our DM1 phenotype sample as non-congenital according to the age of onset. Previous natural history studies (22,23,82) have separated the different non-congenital phenotypes. Two of the studies (23,82) presented a different progression rate, whereas the third study (22) presented an overall comparable progression rate between the DM1 non-congenital phenotypes. Hence, the results are mixed, and we acknowledge that there is a risk that the different non-congenital phenotypes may differ as to progression rate and, thus, impact the responsiveness outcome.

Concerning the psychometric properties of the present outcome measures, the findings offer valuable guidelines for clinicians and researchers. Even though a mean score may be most reliable in theory (64), the present studies evaluated the best score because it is a common clinically relevant outcome, which represents a subject's best performance. Because the assessments were conducted by two assessors, it is likely that the findings are generalizable to other assessors. However, several assessors would improve the generalizability to a broader assessor population. Based on the large, heterogeneous sample, more than one assessor and two different clinical test locations in the present studies, the psychometric findings are overall considered generalizable. Nevertheless, it is the clinician's and researcher's obligation to judge whether the present findings are applicable to their specific individual, assessor and environment (14). For future interventional trials we suggest establishing the psychometric properties in a pilot study prior to study initiation. Nonetheless, the findings offer valuable guidelines for clinicians and researchers who do not have the opportunity to investigate the psychometric properties of these outcome measures prior to application.

11. CONCLUSION

11.1 Conclusions and Recommendations

11.1.1 Studies I-II: Outcome measures

Muscle strength measurements

The feasible and low-cost HHD can be used as a substitute measure for the stationary dynamometry in PwDM1 regarding intra-rater reliability for both groups and single individuals and as to criterion validity for individuals and for the flexor muscles on a group level to determine muscle strength. Moreover, both the stationary dynamometry and the HHD captured muscle strength changes in some muscle groups, almost similarly, after only 1 year despite the slowly declining nature of the disease and may be applicable outcomes for 1-year clinical trials.

Balance measurements

The dynamic balance tests, TUG and step test, are construct valid, intra-rater reliable and feasible tools in PwDM1, and the dynamic TUG was also short-term responsive to balance changes. However, familiarization trials should be implemented for the dynamic balance tests in PwDM1 to eliminate the learning effects. Concerning the static balance tests, only the less feasible m-CTSIB was adequately reliable on a group level and responsive in PwDM1, but the m-CTSIB was associated with a floor effect and is, hence, not recommended in severely affected PwDM1. The remaining static balance tests are not recommended in PwDM1 due to the pronounced ceiling- or floor effects in the heterogeneous cohort of PwDM1.

Functional mobility measurements

The 10mWT and STS are both recommended as satisfactorily construct valid and reliable tools for groups and single individuals to measure functional mobility in PwDM1, but the 10mWT is superior for reliability. None of the functional mobility tests captured a significant change after 1 year, but it is unknown whether these tests are responsive to (larger) changes in functional mobility after an effective intervention in PwDM1. However, the 10mWT was responsive according to the subjective anchor and may be a candidate endpoint for 1-year trials.

All outcome measures

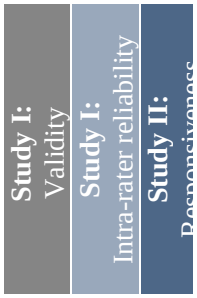
The established SEM-values yield a guideline of measurement error that should be exceeded to represent true changes in a group for research purposes.

The established MDD-values provide a guideline of measurement error that should be exceeded to represent true change for a single individual in the clinic.

The standard deviation of the difference between visit 1 and visit 2 is valuable for conducting a priori sample size calculation for interventional trials (16).

The established MCID-values offer a guideline for researchers and clinicians as reference values for the "average" PwDM1 when interpreting change for a single individual or for a group (the proportion of "responders") (110).

The reliability and responsiveness findings guide researchers and clinicians to select the most reliable and responsive outcome measures so important changes are not concealed by measurement errors or lack of sensitivity. These methodological qualities enable smaller sample sizes.



The findings of reliability, validity and responsiveness are valuable for a priori selection of the most adequate outcome measures for research and in the clinic. In addition, this information is useful for post evaluation for individual follow-up in the clinic as well as for non-significant research findings for which the present outcome measures were applied. A non-significant finding is likely due to no effect or difference, but it can also be due to lack of sufficient validity, reliability or responsiveness of the applied methods used for evaluation.

Although the highest possible level of psychometric evidence of outcome measures is intended, the psychometric properties represent a continuum and the acceptable level of reliability, validity and responsiveness may vary according to the context, e.g. primary- versus secondary outcome, the only test versus part of a test battery, one measurement (inclusion- or exclusion criteria, descriptive purposes) versus repeated measurements (pre-post test, efficacy or natural history purposes), description versus decision-making, stable versus unstable nature of the variable being investigated, and small changes in already trained PwDM1 versus larger effects in sedentary PwDM1.

11.1.2 Study III: Physical activity

Study III: Physical Activity

The PwDM1 demonstrate a physical activity level that is both lower than the WHO recommendations on physical activity and significantly lower than the physical activity level in healthy controls. Promotion of physical activity in PwDM1 is warranted, but assessment of exercise capacity is recommended to plan suitable prescriptions for physical activity in PwDM1.

Study III: Physical Activity

The physical activity level in PwDM1 is only predicted by education, which suggests that physical activity especially should be targeted PwDM1 with lower degrees of education. Because fatigue tended to predict physical activity in PwDM1, fatigue should also be addressed to enhance physical activity in these people. This information provides a better adapted and focussed approach towards a healthier lifestyle in PwDM1.

11.2 Perspectives

11.2.1 Studies I-II: Outcome measures

The present comprehensive findings targeting all the psychometric properties of a broad spectrum of clinically relevant outcome measures in PwDM1 provide the foundation for clinical trials with the aim of improving, stabilizing or delaying the disease progression of the physical function in PwDM1. This will potentially benefit the PwDM1 and the society as to lower economic costs. Additionally, trials failing due to failing psychometric tools or because they are under-powered can be avoided.

To promote a uniform high-quality physical assessment of PwDM1 and to promote comparability of findings worldwide, international consensus on outcome measures in PwDM1 is warranted. Hence, the international OMMYD Initiative provides regular updates, recommendations and guidelines on the full spectrum of outcome measures in PwDM1.

Intra-rater reliability has been documented for the present outcome measures but inter-rater reliability remains to be investigated. Future studies should seek to apply the same set of outcome measures to ensure that findings across studies can be compared. The methodological evidence of outcome measures in PwDM1 is growing, but the entire spectrum of clinically relevant outcome measures in PwDM1 is not fully established. Moreover, even where research exists, solid, large-scale studies or studies with consistent findings are still relevant because it enhances the credibility and strengthens the level of evidence of the methodological findings in PwDM1.

11.2.2 Study III: Physical activity

The present finding of decreased physical activity in PwDM1 is important because it suggests an effort to promote physical activity in these people to improve their physical health, well-being and daily functioning, which also benefits the socioeconomics.

The hurdle and inquiry for future studies are to identify facilitators of physical activity in PwDM1 to enhance the documented lower physical activity level in these people. Moreover, to establish evidence-based exercise

prescriptions as to type of physical activity, duration and intensity that succeed in high compliance and long-term end-of-trial efficacy in PwDM1. Interventions with a social aspect may be subject for investigation. Likewise, engagement of partners, family or a supportive person may also play an important role, since compliance of wearing the physical activity monitor was significantly better for the PwDM1 with cohabitants. Adequate near-contact support may, hence, facilitate a successful adherence to an exercise program or physical activity in general.

12. SUPPLEMENTALS

12.1 Tables

Table 1: Methodological specifications of outcome measure studies in PwDM1

MUSCLE STRENGTH			
MEASUREMENTS	**VALIDITY**	**RELIABILITY**	**RESPONSIVENESS**
Stationary dynamometry	+ (knee extensors) ***Roussel et al. 2019:*** • DM1 n=19 (males) • Knee extensors • Isometric strength • Criterion validity ***Lindeman et al. 1998:*** • NMD n=82 (DM n=33, type unspecified) • Knee extensors • Isokinetic strength • Construct validity (regression, CI unknown)	+ (knee muscles) ***Roussel et al. 2019:*** • DM1 n=19 (males) • Knee extensors • Isometric strength • Absolute and relative reliability ***Tiffreau et al. 2007:*** • NMD n=15 (DM1 n=2) • Knee extensors and flexors • Isokinetic passive motion • Relative reliability	+ (isokinetic strength) ***Lindeman et al. 1995:*** • DM n=25 (type unspecified) • Knee extensors and flexors • Isokinetic strength • 1 yr FU • Significance testing (mostly broad CI) ***Örndahl et al. 1994:*** • DM n=27 (type unspecified) • Knee extensors and flexors • Isokinetic strength • 2 yrs FU • Significance testing (CI unknown)
HHD	+ ***Roussel et al. 2019:*** • DM1 n=19 (males) • Knee extensors • Isometric strength • Criterion validity ***Petitclerc et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=198 • Knee extensors and flexors, hip flexors, ankle dorsal flexors • Isometric strength • Construct validity (regression, narrow CI) ***Hammarén et al. 2014:*** • DM1 n=51 • Knee extensors and flexors, hip flexors, ankle dorsal flexors • Isometric strength • Construct validity (correlation) ***Jimenez-Moreno et al. 2019:*** • DM1 n=51 • Knee extensors, hip flexors, ankle dorsal flexors • Isometric strength • Construct validity (correlation)	++ (ankle dorsal flexors) ***Hébert et al. 2010:*** • DM1 n=46 • Ankle dorsal flexors • Isometric strength • Absolute reliability	+++ ***Roussel et al. 2021:*** • DM1 n=23 • Knee extensors and flexors, hip flexors and extensors, ankle dorsal flexors • Isometric strength • 3 yrs FU • Significance testing (narrow to broad CI) ***Gagnon et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=100 • Knee extensors and flexors, hip flexors, ankle dorsal flexors • Isometric strength • 9 yrs FU • Significance testing (mostly broad CI) ***Kierkegaard et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=113 • Knee extensors and flexors, hip flexors, ankle dorsal flexors • Isometric strength • 9 yrs FU • Significance testing (narrow to broad IQR) and anchor-based approach

			Sedehizadeh et al. 2016: • DM1 n=38 • Ankle dorsal flexors • Isometric strength • 1½ yr FU • Significance testing (CI unknown) ***Hammarén et al. 2015:*** • DM1 n=43 • Knee extensors and flexors, hip flexors, ankle dorsal flexors • Isometric strength • 5 yrs FU • Significance testing (narrow CI)

BALANCE

MEASUREMENTS	VALIDITY	RELIABILITY	RESPONSIVENESS
Step test	++ ***Hammarén et al. 2014:*** • DM1 n=51 • Construct validity (correlation)	+ ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	++ ***Hammarén et al. 2015:*** • DM1 n=43 • 5 yrs FU • Significance testing (broad CI)
TUG	+++ ***Petitclerc et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=198 • Construct validity (regression, narrow CI) ***Hammarén et al. 2014:*** • DM1 n=51 • Construct validity (correlation)	+ (intra-rater) ***Kierkegaard et al. 2017:*** • DM1 n=70 • Intra-session reliability ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	+++ ***Roussel et al. 2021:*** • DM1 n=23 • 3 yrs FU • Significance testing (narrow CI) ***Kierkegaard et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=113 • 9 yrs FU • Significance testing (moderate IQR) and anchor-based approach ***Hammarén et al. 2015:*** • DM1 n=43 • 5 yrs FU • Significance testing (broad CI)
Feet-together stance	+ ***Hammarén et al. 2012:*** • DM1 n=10 • Ceiling effect	+ ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	÷
Tandem stance	÷	+ ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	÷
One-leg stance	÷	+ ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	÷

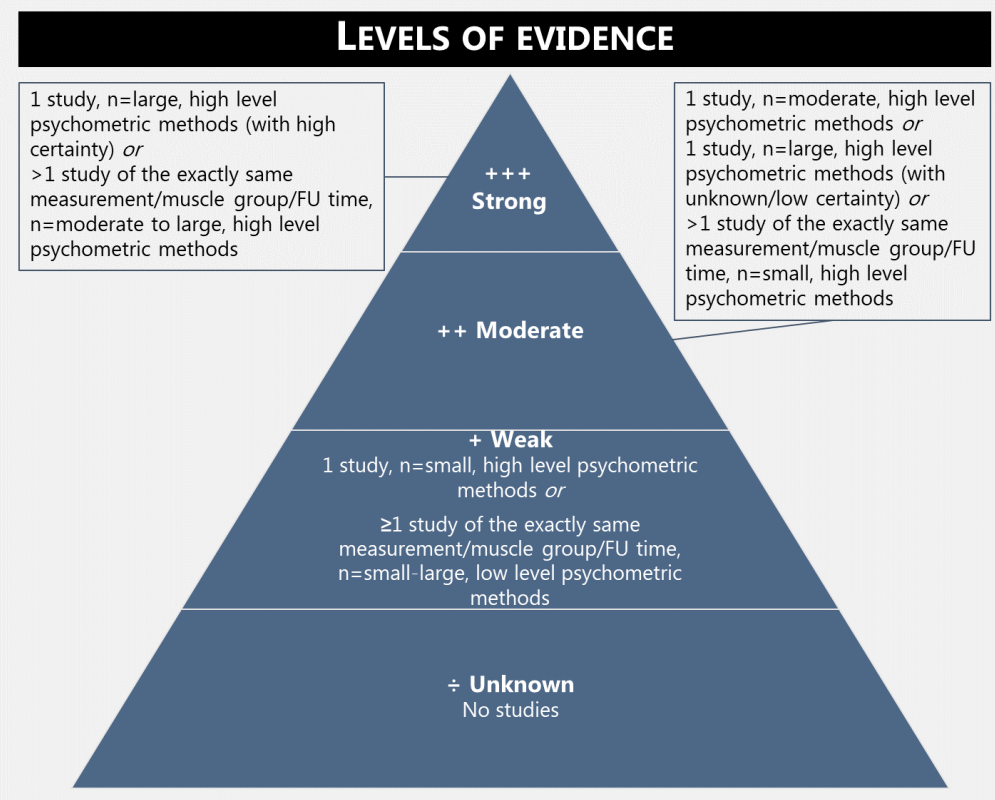
m-CTSIB	÷ ***Pucillo et al. 2018:*** • DM1 n=22 • Limits of stability • NeuroCom SMART Balance Master • Construct validity (correlation)	÷	÷

FUNCTIONAL MOBILITY

MEASUREMENTS	VALIDITY	RELIABILITY	RESPONSIVENESS
10mWT	**+++** ***Jimenez-Moreno et al. 2019:*** • DM1 n=113 • Construct validity (correlation) ***Pucillo et al. 2018:*** • DM1 n=22 • Construct validity (correlation) ***Petitclerc et al. 2018:*** *(part of a larger research project, Quebec Canada)* • DM1 n=198 • Construct validity (regression, narrow CI) ***Hammarén et al. 2014:*** • DM1 n=51 • Construct validity (correlation)	**+** (intra-rater) ***Jimenez-Moreno et al. 2019:*** • DM1 n=113 • Intra-session reliability ***Kierkegaard et al. 2017:*** • DM1 n=70 • Intra-session reliability ***Hammarén et al. 2012:*** • DM1 n=10 • Absolute and relative reliability	**+++** ***Roussel et al. 2021:*** • DM1 n=23 • 3 yrs FU • Significance testing (narrow CI) ***Jimenez-Moreno et al. 2019:*** • DM1 n=98 • 1 yr FU • Significance testing (narrow CI) ***Hammarén et al. 2015:*** • DM1 n=43 • 5 yrs FU • Significance testing (broad CI)
STS	**+** (30-second STS) ***Jimenez-Moreno et al. 2019:*** • DM1 n=113 • 30-second STS • Construct validity (correlation)	÷ (intra-rater) ***Jimenez-Moreno et al. 2019:*** • DM1 n=113 • 30-second STS • Intra-session reliability ***Kierkegaard et al. 2017:*** • DM1 n=70 • 10-times STS • Intra-session reliability	**+++** (30-second STS) ***Jimenez-Moreno et al. 2019:*** • DM1 n=98 • 30-second STS • 1 yr FU • Significance testing (narrow CI) ***Nitz et al. 1999:*** • DM n=36 (type unspecified) • 1-time STS • 1 yr FU • Significance testing (CI unknown)

NMD = Neuromuscular diseases

FU = Follow-up



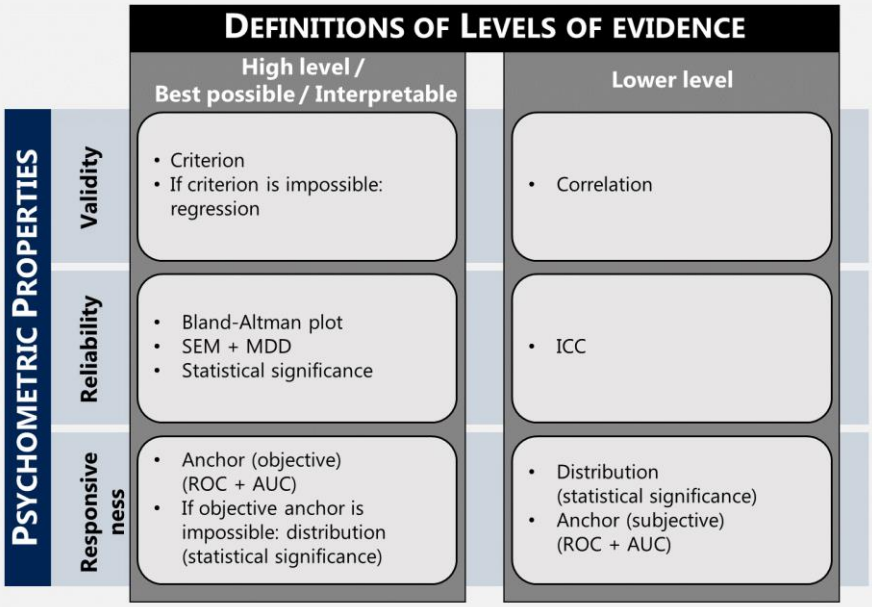


Table 2: Missing values for demographic data for PwDM1 (Studies I-III)

	STUDY I		STUDY II	STUDY III
	VALIDITY	RELIABILITY	RESPONSIVENESS	PWDM1
AES-S[1]	n=15 • Test not implemented (n=8) • Incomplete test (n=2) • Drop out before test (n=4) • Test not conducted (n=1)	n=10 • Test not implemented (n=8) • Incomplete test (n=2)	n=8 • Test not implemented (n=7) • Incomplete test (n=1)	n=11 • Test not implemented (n=8) • Incomplete test (n=1) • Drop out before test (n=1) • Not possible to attend test (n=1)
Stroop Word	n=17 • Test not implemented (n=8) • Colour blind (n=2) • Invalid score (n=1) • Reading disability (n=1) • Drop out before test (n=4) • Test not conducted (n=1)	n=12 • Test not implemented (n=8) • Colour blind (n=2) • Invalid score (n=1) • Reading disability (n=1)	n=11 • Test not implemented (n=7) • Colour blind (n=2) • Invalid score (n=1) • Reading disability (n=1)	n=14 • Test not implemented (n=8) • Colour blind (n=2) • Invalid score (n=1) • Reading disability (n=1) • Drop out before test (n=1) • Not possible to attend test (n=1)
Stroop Colour	n=16 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Drop out before test (n=4) • Test not conducted (n=1)	n=11 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1)	n=10 • Test not implemented (n=7) • Colour blind (n=2) • Reading disability (n=1)	n=13 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Drop out before test (n=1) • Not possible to attend test (n=1)
Stroop Colour-Word	n=17 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Drop out before test (n=4) • Test not conducted (n=1) • Incomplete test (n=1)	n=12 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Declined to finish test (n=1)	n=10 • Test not implemented (n=7) • Colour blind (n=2) • Reading disability (n=1)	n=14 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Declined to finish test (n=1) • Drop out before test (n=1) • Not possible to attend test (n=1)
Stroop Interference	n=16 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Drop out before test (n=4)	n=11 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1)	n=10 • Test not implemented (n=7) • Colour blind (n=2) • Reading disability (n=1)	n=13 • Test not implemented (n=8) • Colour blind (n=2) • Reading disability (n=1) • Drop out before test (n=1)

	• Test not conducted (n=1)			• Not possible to attend test (n=1)
FSS-7[2]	NA	NA	NA	n=1 • Incomplete test (n=1)

[1]AES-S=Apathy Evaluation Scale (Self-rated)

[2]FSS-7=Fatigue Severity Scale (7 items)

Table 3: Missing values for outcome measures (Studies I-II)

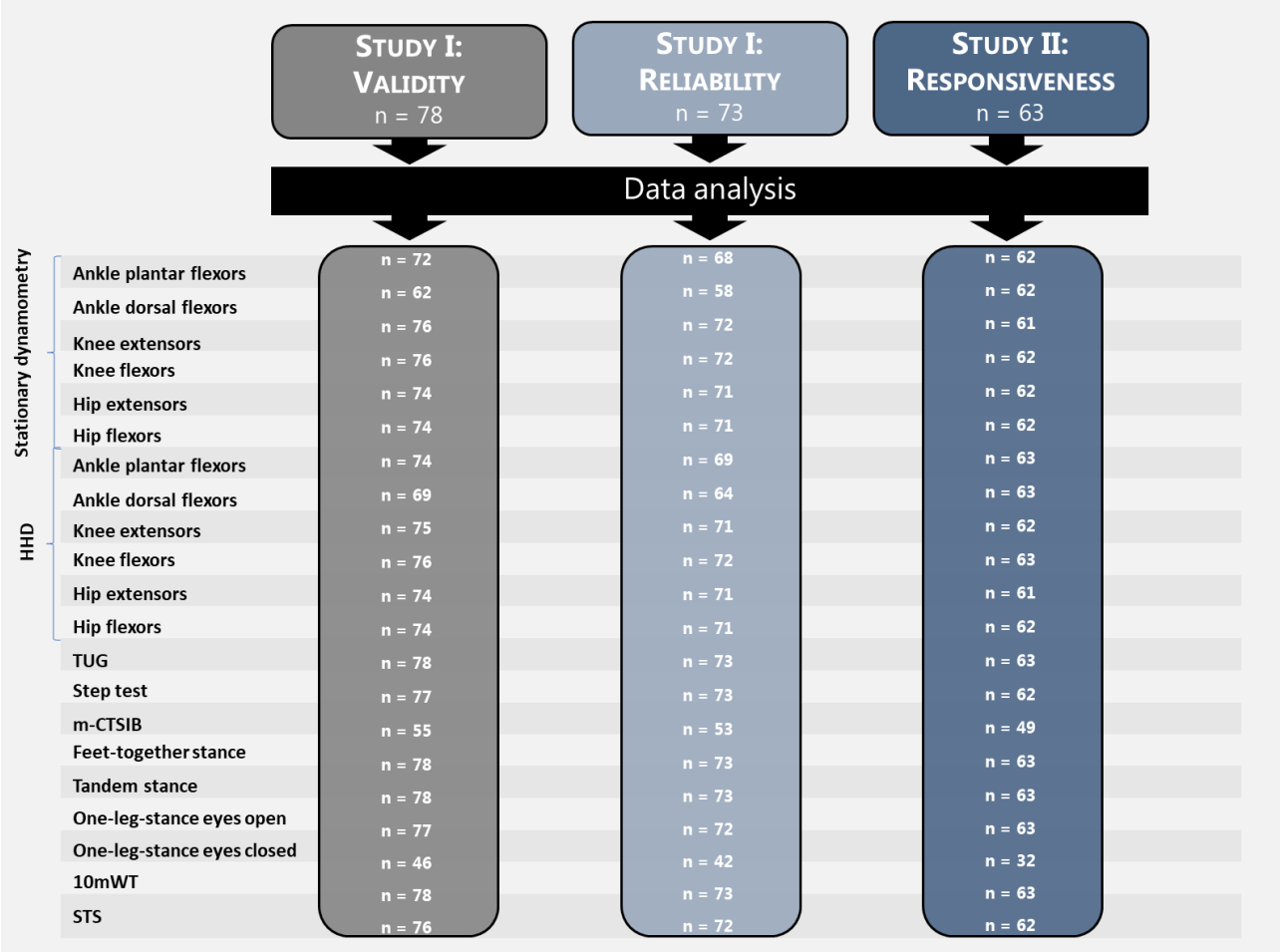
	STUDY I		STUDY II
	VALIDITY	RELIABILITY	RESPONSIVENESS
Stationary dynamometry			
Ankle plantar flexors	n=6 • Inability to activate test (n=4) • AFO (n=1) • Incomplete test (n=1)	n=5 • AFO at visit 1 (n=1) • Missing value for first trial (n=1) • Inability to activate test at visit 1 but not visit 2 (n=4) • Inability to activate test at visit 2 but not visit 1 (n=2) • Inability to activate test at visits 1+2 (n=3) *(thus, the intersection is 5 missing values)*	n=1 • Technical issue (n=1)
Ankle dorsal flexors	n=16 • Inability to activate test (n=14) • AFO (n=1) • Incomplete test (n=1)	n=15 • Inability to activate test (n=15)	n=1 • Technical issue (n=1)
Knee extensors	n=2 • Technical issue (n=1) • Incomplete test (n=1)	n=1 • Technical issue (n=1)	n=2 • Technical issue (n=1) • Knee pain (n=1)
Knee flexors	n=2 • Technical issue (n=1) • Incomplete test (n=1)	n=1 • Technical issue (n=1)	n=1 • Technical issue (n=1)
Hip extensors	n=4 • Technical issue (n=2) • Incomplete test (n=2)	n=2 • Technical issue (n=1) • Incomplete test (n=1)	n=1 • Technical issue (n=1)
Hip flexors	n=4 • Technical issue (n=2) • Incomplete test (n=2)	n=2 • Technical issue (n=1) • Incomplete test (n=1)	n=1 • Technical issue (n=1)
HHD			
Ankle plantar flexors	n=4 • MRC[1]<3 (n=3) • Incomplete test (n=1)	n=4 • MRC<3 (n=3) • Incomplete test (n=1)	n=0
Ankle dorsal flexors	n=9 • MRC<3 (n=8) • Incomplete test (n=1)	n=9 • MRC<3 (n=8) • Incomplete test (n=1)	n=0

Knee extensors	n=3 • Rater unable to hold position (n=1) • Incomplete test (n=2)	n=2 • Rater unable to hold position (n=1) • Incomplete test (n=1)	n=1 • Knee pain (n=1)
Knee flexors	n=2 • Incomplete test (n=2)	n=1 • Incomplete test (n=1)	n=0
Hip extensors	n=4 • Incomplete test (n=1) • Rater unable to hold position (n=1) • Patient compensation (n=2)	n=2 • Incomplete test (n=1) • Rater unable to hold position (n=1)	n=2 • Patient compensation (n=1) • Rater unable to hold position (n=1)
Hip flexors	n=4 • Rater unable to hold position (n=2) • Incomplete test (n=2)	n=2 • Rater unable to hold position (n=1) • Incomplete test (n=1)	n=1 • Rater unable to hold position (n=1)
TUG	n=0	n=0	n=0
Step test	n=1 • Uncomfortable with performing test (n=1)	n=0	n=1 • Knee pain (n=1)
m-CTSIB	n=23 • Technical issue (n=1) • Inability to complete test (n=20) • Not values for all trials (n=2)	n=20 • Technical issue (n=1) • Inability to complete test at visits 1+2 (n=14) • Inability to complete test at visit 1 but not visit 2 (n=4) • Inability to complete test at visit 2 but not visit 1 (n=7) *(thus, the intersection is 20 missing values)*	n=14 • Technical issue (n=2) • Inability to complete test (n=10) • Inability to initiate test (n=2)
Feet-together stance	n=0	n=0	n=0
Tandem stance	n=0	n=0	n=0
One-leg-stance eyes open	n=1 • Inability to initiate test (n=1)	n=1 • Inability to initiate test (n=1)	n=0
One-leg-stance eyes closed	n=32 • NA (inability to initiate one-leg-stance eyes open test) (n=1) • Not performed (n=1) • Inability to stand ≥30 s in one-leg-stance eyes open test (n=30)	n=31 • NA (inability to initiate one-leg-stance eyes open test) (n=1) • Inability to stand ≥30 s in one-leg-stance eyes open test (n=30)	n=31 • Inability to stand ≥30 s in one-leg-stance eyes open test (n=30) • Inability to initiate test (n=1)
10mWT	n=0	n=0	n=0
STS	n=2 • Incomplete test (n=2)	n=1 • Incomplete test (n=1)	n=1 • Incomplete test (n=1)

[1]*MRC<3 = ROM (Range Of Motion) against gravity impossible.*

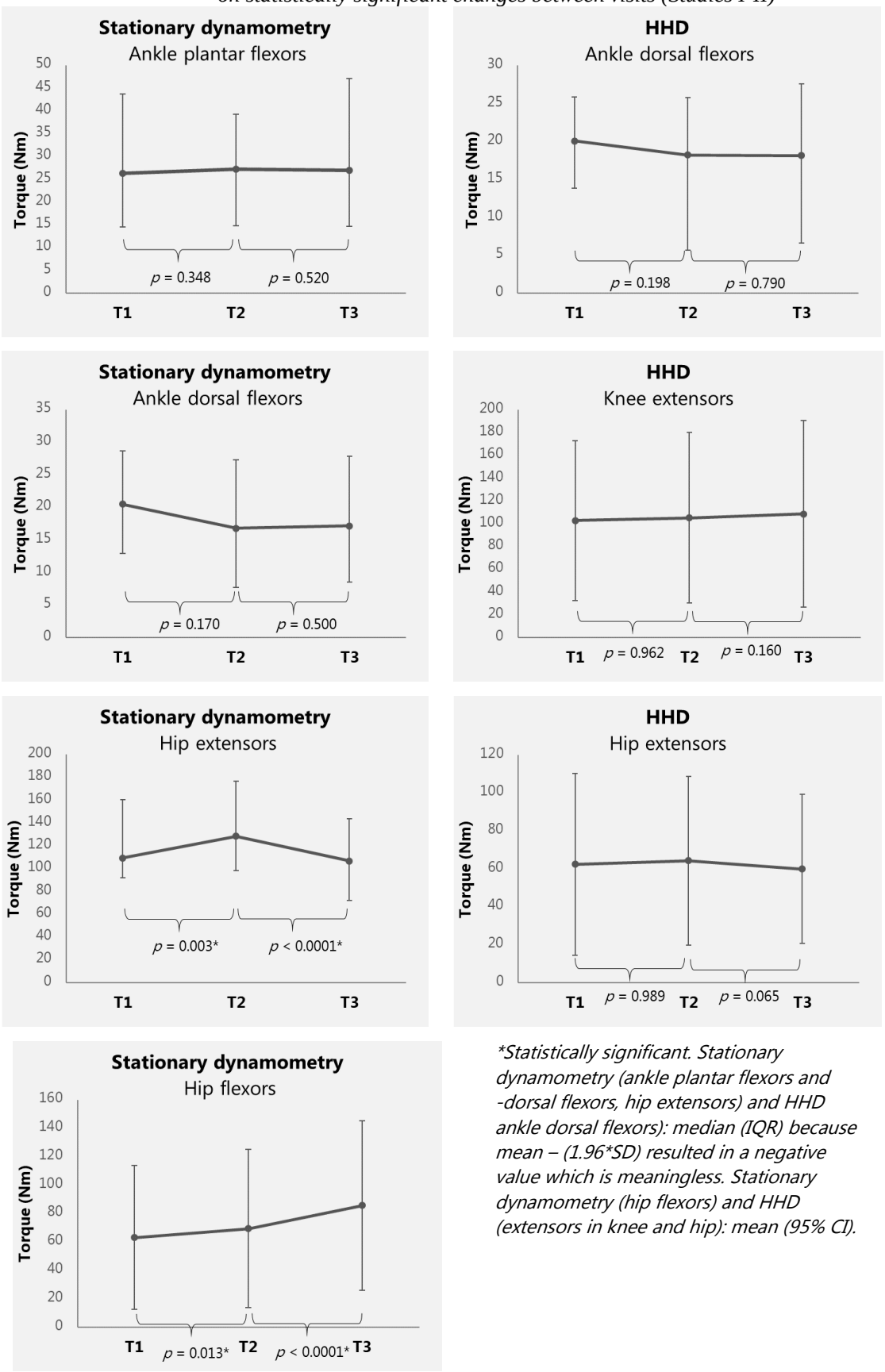
12.2 Figures

Figure 1: Number of PwDM1 included in data analyses (Studies I-II)



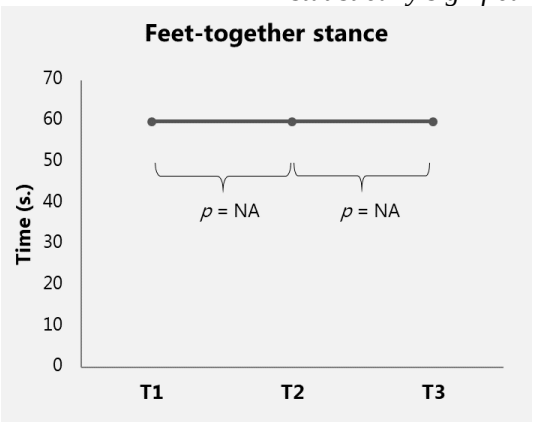
(57,58)

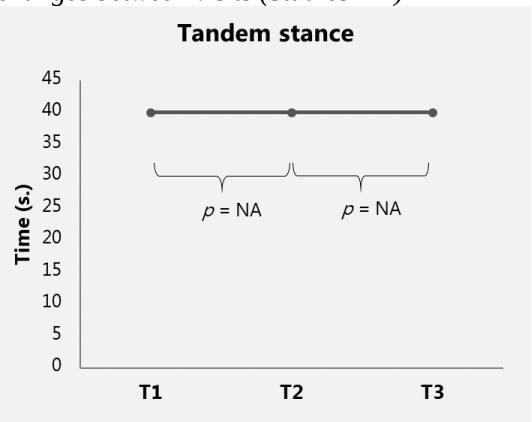
Figure 2: Reliability (T1-T2) and responsiveness (T2-T3) of muscle strength measurements in PwDM1 based on statistically significant changes between visits (Studies I-II)

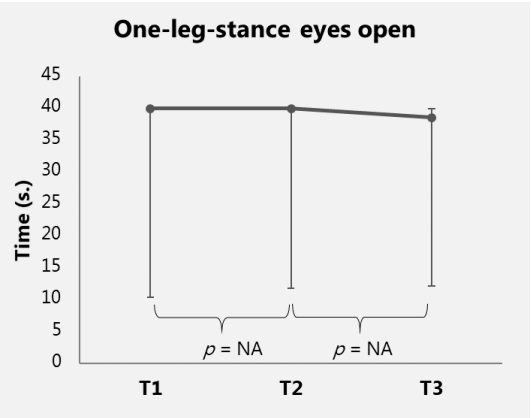


**Statistically significant. Stationary dynamometry (ankle plantar flexors and -dorsal flexors, hip extensors) and HHD ankle dorsal flexors): median (IQR) because mean – (1.96*SD) resulted in a negative value which is meaningless. Stationary dynamometry (hip flexors) and HHD (extensors in knee and hip): mean (95% CI).*

Figure 3: Reliability (T1-T2) and responsiveness (T2-T3) of balance measurements in PwDM1 based on statistically significant changes between visits (Studies I-II)







Feet-together stance, tandem stance, and one-leg-stance eyes open: median (IQR) because of null-inflation (no true variance across participants).

13. REFERENCES

1. Schara U. Congenital and childhood-onset myotonic dystrophy: importance of long-term data in natural history. Dev Med Child Neurol. 2016;58(7):652–652.

2. Hammarén E, Kjellby-Wendt G, Kowalski J, Lindberg C. Factors of importance for dynamic balance impairment and frequency of falls in individuals with myotonic dystrophy type 1 – A cross-sectional study – Including reference values of Timed Up & Go, 10m walk and step test. Neuromuscul Disord. 2014;24(3):207–15.

3. Kierkegaard M, Petitclerc E, Hébert LJ, Gagnon C. Is one trial enough for repeated testing? Same-day assessments of walking, mobility and fine hand use in people with myotonic dystrophy type 1. Neuromuscul Disord NMD. 2017;27(2):153–8.

4. Turner C, Hilton-Jones D. Myotonic dystrophy: diagnosis, management and new therapies. Curr Opin Neurol. 2014;27(5):599–606.

5. Bouchard J-P, Cossette L, Bassez G, Puymirat J. Natural history of skeletal muscle involvement in myotonic dystrophy type 1: a retrospective study in 204 cases. J Neurol. 2015;262(2):285–93.

6. Kaminsky P, Poussel M, Pruna L, Deibener J, Chenuel B, Brembilla-Perrot B. Organ dysfunction and muscular disability in myotonic dystrophy type 1. Medicine (Baltimore). 2011;90(4):262–8.

7. Okkersen K, Buskes M, Groenewoud J, Kessels RPC, Knoop H, van Engelen B et al. The cognitive profile of myotonic dystrophy type 1: A systematic review and meta-analysis. Cortex J Devoted Study Nerv Syst Behav. 2017;95:143–55.

8. Baldanzi S, Bevilacqua F, Lorio R, Volpi L, Simoncini C, Petrucci A et al. Disease awareness in myotonic dystrophy type 1: an observational cross-sectional study. Orphanet Journal of Rare Diseases. 2016(11):34.

9. Rubinsztein JS, Rubinsztein DC, Goodburn S, Holland AJ. Apathy and hypersomnia are common features of myotonic dystrophy. J Neurol Neurosurg Psychiatry. 1998;64(4):510–5.

10. Kalkman JS, Schillings ML, van der Werf SP, Padberg GW, Zwarts MJ, van Engelen BGM et al. Experienced fatigue in facioscapulohumeral dystrophy, myotonic dystrophy, and HMSN-I. J Neurol Neurosurg Psychiatry. 2005;76(10):1406–9.

11. Smith CA, Gutmann L. Myotonic Dystrophy Type 1 Management and Therapeutics. Curr Treat Options Neurol. 2016;18(52). DOI 10.1007/s11940-016-0434-1

12. Gonzalez-Alegre P. Recent advances in molecular therapies for neurological disease: triplet repeat disorders. Hum Mol Genet. 2019;28(R1):R80–7.

13. Nguyen C-TE, Campbell C. Myotonic dystrophy type 1. Can Med Assoc J. 2016;188(14):1033–1033.

14. Portney L, Watkins M. Reliability of Measurements. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

15. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

16. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med Auckl NZ. 1998;26(4):217–38.

17. Bland J, Altman D. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999(8):135–60.

18. Stratford P, Goldsmith C. Use of the Standard Error as a Reliability Index of Interest: An Applied Example Using Elbow Flexor Strength Data. Phys Ther. 1997(77):745–50.

19. Portney L, Watkins M. Validity of Measurements. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

20. Portney L, Watkins M. Statistical Measures of Validity. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

21. Petitclerc É, Hébert LJ, Mathieu J, Desrosiers J, Gagnon C. Relationships between Lower Limb Muscle Strength Impairments and Physical Limitations in DM1. J Neuromuscul Dis. 2018;5(2):215–24.

22. Gagnon C, Petitclerc É, Kierkegaard M, Mathieu J, Duchesne É, Hébert LJ. A 9-year follow-up study of quantitative muscle strength changes in myotonic dystrophy type 1. J Neurol. 2018;265(7):1698–705.

23. Kierkegaard M, Petitclerc É, Hébert LJ, Mathieu J, Gagnon C. Responsiveness of performance-based outcome measures for mobility, balance, muscle strength and manual dexterity in adults with myotonic dystrophy type 1. J Rehabil Med. 2018;50(3):269–77.

24. Jimenez-Moreno AC, Nikolenko N, Kierkegaard M, Blain AP, Newman J, Massey C et al. Analysis of the functional capacity outcome measures for myotonic dystrophy. Ann Clin Transl Neurol. 2019;6(8):1487–97.

25. Roussel M-P, Hébert LJ, Duchesne E. Intra-Rater Reliability and Concurrent Validity of Quantified Muscle Testing for Maximal Knee Extensors Strength in Men with Myotonic Dystrophy Type 1. J Neuromuscul Dis. 2019;6(2):233–40.

26. Lindeman E, Leffers P, Spaans F, Drukker J, Reulen J. Deterioration of motor function in myotonic dystrophy and hereditary motor and sensory neuropathy. Scand J Rehab Med. 1995;27:59–64.

27. Örndahl G, Grimby G, Grimby A, Johansson G, Wilhelmsen L. Functional deterioration and selenium-vitamin E treatment in myotonic dystrophy. A placebo-controlled study. J Intern Med. 1994;235(3):205–10.

28. Hébert LJ, Remec J-F, Saulnier J, Vial C, Puymirat J. The use of muscle strength assessed with handheld dynamometers as a non-invasive biological marker in myotonic dystrophy type 1 patients: a multicenter study. BMC Musculoskelet Disord. 2010;11(1):1.

29. Hammarén E, Ohlsson JA, Lindberg C, Kjellby-Wendt G. Reliability of static and dynamic balance tests in subjects with myotonic dystrophy type 1. Adv Physiother. 2012;14(2):48–54.

30. Hammarén E, Kjellby-Wendt G, Lindberg C. Muscle force, balance and falls in muscular impaired individuals with myotonic dystrophy type 1: A five-year prospective cohort study. Neuromuscul Disord. 2015;25(2):141–8.

31. Pucillo EM, Mcintyre MM, Pautler M, Hung M, Bounsanga J, Voss MW et al. Modified dynamic gait index and limits of stability in myotonic dystrophy type 1. Muscle Nerve. 2018;58(5):694–9.

32. World Health Organization. Physical activity. 2021. Link: https://www.who.int/health-topics/physical-activity#tab=tab_1

33. McDonald CM. Physical activity, health impairments, and disability in neuromuscular disease. Am J Phys Med Rehabil. 2002;81(11 Suppl):S108-120.

34. Vignos P. Physical models of rehabilitation in neuromuscular disease. Muscle Nerve. 1983(6):323–38.

35. Brady LI, MacNeil LG, Tarnopolsky MA. Impact of Habitual Exercise on the Strength of Individuals with Myotonic Dystrophy Type 1: Am J Phys Med Rehabil. 2014;93(9):739–50.

36. Voet NB, van der Kooi EL, van Engelen BG, Geurts AC. Strength training and aerobic exercise training for muscle disease. Cochrane Database Syst Rev. 2019;12:CD003907.

37. Voet NB, van der Kooi EL, Riphagen II, Lindeman E, van Engelen BG, Geurts AC. Strength training and aerobic exercise training for muscle disease. Cochrane Database of Systematic Reviews 2019, Issue 12. Art. No.: CD003907. DOI: 10.1002/14651858.CD003907.pub5.

38. Gianola S, Pecoraro V, Lambiase S, Gatti R, Banfi G, Moja L. Efficacy of Muscle Exercise in Patients with Muscular Dystrophy: A Systematic Review Showing a Missed Opportunity to Improve Outcomes. PLoS ONE. 2013;8(6):e65414.

39. Rimmer JH, Chen M-D, McCubbin JA, Drum C, Peterson J. Exercise intervention research on persons with disabilities: what we know and where we need to go. Am J Phys Med Rehabil. 2010;89(3):249–63.

40. Lindeman E, Spaans F, Reulen J, Leffers P, Drukker J. Progressive resistance training in neuromuscular patients. Effects on force and surface EMG. J Electromyogr Kinesiol Off J Int Soc Electrophysiol Kinesiol. 1999;9(6):379–84.

41. Tollbäck A, Eriksson S, Wredenberg A, Jenner G, Vargas R, Borg K et al. Effects of high resistance training in patients with myotonic dystrophy. Scand J Rehabil Med. 1999;31(1):9–16.

42. Orngreen MC, Olsen DB, Vissing J. Aerobic training in patients with myotonic dystrophy type 1. Ann Neurol. 2005;57(5):754–7.

43. Cup EH, Pieterse AJ, Ten Broek-Pastoor JM, Munneke M, van Engelen BG, Hendricks HT et al. Exercise therapy and other types of physical therapy for patients with neuromuscular diseases: a systematic review. Arch Phys Med Rehabil. 2007;88(11):1452–64.

44. Kierkegaard M, Harms-Ringdahl K, Holmqvist LW, Tollbäck A. Functioning and disability in adults with myotonic dystrophy type 1. Disabil Rehabil. 2011;33(19–20):1826–36.

45. Aldehag A, Jonsson H, Lindblad J, Kottorp A, Ansved T, Kierkegaard M. Effects of hand-training in persons with myotonic dystrophy type 1 – a randomised controlled cross-over pilot study. Disabil Rehabil. 2013;35(21):1798–807.

46. Sheikh AM, Vissing J. Exercise therapy for muscle and lower motor neuron diseases. Acta Myol Myopathies Cardiomyopathies Off J Mediterr Soc Myol. 2019;38(4):215–32.

47. Missaoui B, Rakotovao E, Bendaya S, Mane M, Pichon B, Faucher M et al. Posture and gait abilities in patients with myotonic dystrophy (Steinert disease). Evaluation on the short-term of a rehabilitation program. Ann Phys Rehabil Med. 2010;53(6–7):387–98.

48. Okkersen K, Jimenez-Moreno C, Wenninger S, Daidj F, Glennon J, Cumming S et al. Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial. Lancet Neurol. 2018;17(8):671–80.

49. Lessard I, Gaboury S, Gagnon C, Bouchard K, Chapron K, Lavoie M et al. Effects and Acceptability of an Individualized Home-Based 10-Week Training Program in Adults with Myotonic Dystrophy Type 1. J Neuromuscul Dis. 2021;8(1):137–49.

50. Roussel M-P, Hébert LJ, Duchesne E. Strength-training effectively alleviates skeletal muscle impairments in myotonic dystrophy type 1. Neuromuscul Disord NMD. 2020;30(4):283–93.

51. Jimenez-Moreno A, Okkersen K, Maas D, Knoop H, Lochmuller H, Gorman G et al. Implementing a Graded Exercise Therapy to Severely Fatigued DM1 Patients: Experience from the OPTIMISTIC Study. 2017;International myotonic dystrophy consortium meeting(sept).

52. Gagnon C, Chouinard M-C, Laberge L, Brisson D, Gaudet D, Lavoie M et al. Prevalence of Lifestyle Risk Factors in Myotonic Dystrophy Type 1. Can J Neurol Sci. 2013;40(01):42–7.

53. Eichinger K, Dekdebrun J, Dilek N, Pandya S, Chen D. Physical Activity Profile and Barriers to Physical Activity in Individuals with Myotonic Dystrophy (P5.084). Neurology. 2014;82(10):P5.084.

54. Phillips M, Marr L. P3-12 Participation in physical activity by people with myotonic dystrophy. Med Genet. 2009;440.

55. Ayvat F, Ayvat E, Kılınç OG, Kılınç M, Yıldırım SA, Tan E. Objective and Subjective Assessment of Physical Activity in Adults with Muscle Diseases. Turk J Neurol Turk Noroloji Derg. 2019;25(3):117–22.

56. Wiles CM. Falls and stumbles in myotonic dystrophy. J Neurol Neurosurg Psychiatry. 2005;77(3):393–6.

57. Knak KL, Sheikh AM, Andersen H, Witting N, Vissing J. Intrarater reliability and validity of outcome measures in myotonic dystrophy type 1. Neurology. 2020;94(24):e2508–20.

58. Knak KL, Sheikh AM, Witting N, Vissing J. Responsiveness of outcome measures in myotonic dystrophy type 1. Ann Clin Transl Neurol. 2020;7(8):1382–91.

59. Knak KL, Sheikh AM, Witting N, Vissing J. Physical activity in myotonic dystrophy type 1. J Neurol. 2020;267(6):1679–86.

60. Biodex. Biosway Portable Balance System Operation Manual 950-460 950-461. 2019. Link: http://www.biodex.com/sites/default/files/950460man_10202revd.pdf

61. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998;30(5):777–81.

62. Troiano RP. Large-scale applications of accelerometers: new frontiers and new questions. Med Sci Sports Exerc. 2007;39(9):1501.

63. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.

64. Portney L, Watkins M. Statistical Measures of Reliability. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

65. Wikipedia. Bland–Altman plot. 2019. Link: https://en.wikipedia.org/wiki/Bland%E2%80%93Altman_plot

66. Schwenk M, Gogulla S, Englert S, Czempik A, Hauer K. Test–retest reliability and minimal detectable change of repeated sit-to-stand analysis using one body fixed sensor in geriatric patients. Physiol Meas. 2012;33(11):1931.

67. Faria CDCM, Teixeira-Salmela LF, Neto MG, Rodrigues-de-Paula F. Performance-based tests in subjects with stroke: outcome scores, reliability and measurement errors. Clin Rehabil. 2012;26(5):460–9.

68. Huang S-L, Hsieh C-L, Wu R-M, Tai C-H, Lin C-H, Lu W-S. Minimal detectable change of the timed “up & go” test and the dynamic gait index in people with Parkinson disease. Phys Ther. 2011;91(1):114–21.

69. Otten R, Whiteley R, Mitchell T. Effect of subject restraint and resistance pad placement on isokinetic knee flexor and extensor strength: implications for testing and rehabilitation. Sports Health. 2013;5(2):137–42.

70. Portney L, Watkins M. Concepts of Measurement. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

71. Norman GR, Sloan JA, Wyrwich KW. Interpretation of Changes in Health-Related Quality of Life: The Remarkable Universality of Half a Standard Deviation. Med Care. 2003;41(5):582–92.

72. Portney L, Watkins M. Correlation. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.

73. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

74. Laerd Statistics. ROC Curve. Link: https://statistics.laerd.com/premium/spss/blr/binomial-logistic-regression-in-spss-12.php#roc

75. Altman D. Some common problems in medical research. In: Practical Statistics for Medical Research. London; 1991.

76. Valentin G, Maribo T. Hand-held dynamometry fixated with a tripod is reliable for assessment of back extensor strength in women with osteoporosis. Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA. 2014;25(8):2143–9.

77. Shirley Ryan Abilitylab. Apathy Evaluation Scale (Self, Informant, and Clinician Versions). 2021. Link: https://www.sralab.org/rehabilitation-measures/apathy-evaluation-scale-includes-three-forms-self-informant-and-clinician

78. Gallais B, Gagnon C, Côté I, Forgues G, Laberge L. Reliability of the Apathy Evaluation Scale in Myotonic Dystrophy Type 1. J Neuromuscul Dis. 2018;5(1):39–46.

79. Lencioni T, Piscosquito G, Rabuffetti M, Bovi G, Di Sipio E, Diverio M et al. Responsiveness of gait analysis parameters in a cohort of 71 CMT subjects. Neuromuscul Disord NMD. 2017;27(11):1029–37.

80. Peric S, Vujnic M, Dobricic V, Marjanovic A, Basta I, Novakovic I et al. Five-year study of quality of life in myotonic dystrophy. Acta Neurol Scand. 2016;134(5):346–51.

81. Gagnon C, Heatwole C, Hébert LJ, Hogrel J-Y, Laberge L, Leone M et al. Report of the third outcome measures in myotonic dystrophy type 1 (OMMYD-3) international workshop Paris, France, 2015. J Neuromuscul Dis. 2018;5(4):523–37.

82. Roussel M-P, Fiset M-M, Gauthier L, Lavoie C, McNicoll É, Pouliot L et al. Assessment of muscular strength and functional capacity in the juvenile and adult myotonic dystrophy type 1 population: a 3-year follow-up study. J Neurol. 2021. https://doi.org/10.1007/s00415-021-10533-6

83. Nitz JC, Burns YR, Jackson RV. A longitudinal physical profile assessment of skeletal muscle manifestations in myotonic dystrophy. Clin Rehabil. 1999;13(1):64–73.

84. Sedehizadeh S, Brook JD, Maddison P. Body composition and clinical outcome measures in patients with myotonic dystrophy type 1. Neuromuscul Disord NMD. 2017;27(3):286–9.

85. Davis PR, McKay MJ, Baldwin JN, Burns J, Pareyson D, Rose KJ. Repeatability, consistency, and accuracy of hand-held dynamometry with and without fixation for measuring ankle plantarflexion strength in healthy adolescents and adults. Muscle Nerve. 2017;56(5):896–900.

86. Martins J, da Silva JR, da Silva MRB, Bevilaqua-Grossi D. Reliability and Validity of the Belt-Stabilized Handheld Dynamometer in Hip- and Knee-Strength Tests. J Athl Train. 2017;52(9):809–19.

87. Thorborg K, Bandholm T, Hölmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. Knee Surg Sports Traumatol Arthrosc Off J ESSKA. 2013;21(3):550–5.

88. Wollin M, Purdam C, Drew MK. Reliability of externally fixed dynamometry hamstring strength testing in elite youth football players. J Sci Med Sport. 2016;19(1):93–6.

89. Buckinx F, Croisier J-L, Reginster J-Y, Dardenne N, Beaudart C, Slomian J, m.fl. Reliability of muscle strength measures obtained with a hand-held dynamometer in an elderly population. Clin Physiol Funct Imaging. 2017;37(3):332–40.

90. Kimura M, Ogata Y, Akebi T, Ochi M, Hachisuka K, Saeki S. Development and reliability of a hand-held dynamometer device to measure the ankle dorsiflexor muscle strength. J Rehabil Assist Technol Eng. 2018;5:2055668318802558.

91. Mañago MM, Hebert JR, Schenkman M. Psychometric Properties of a Clinical Strength Assessment Protocol in People with Multiple Sclerosis. Int J MS Care. 2017;19(5):253–62.

92. van der Plas E, Gutmann L, Thedens D, Shields RK, Langbehn K, Guo Z et al. Quantitative muscle MRI as a sensitive marker of early muscle pathology in myotonic dystrophy type 1. Muscle Nerve. 2021;63(4):553–62.

93. Shumway-Cook A, Woollacott M. Normal Postural Control. In: Motor Control - Translating Research into Clinical Practice. 3rd Edition. Lippincott Williams & Wilkins; 2007.

94. Scaglioni-Solano P, Aragón-Vargas LF. Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and sensory integration in highly functional older adults. Int J Rehabil Res Int Z Rehabil Rev Int Rech Readaptation. 2014;37(2):138–43.

95. Riemann BL, Lininger M, Kirkland MK, Petrizzo J. Age related changes in balance performance during self-selected and narrow stance testing. Arch Gerontol Geriatr. 2018;75:65–9.

96. Wrisley DM, Whitney SL. The effect of foot position on the modified clinical test of sensory interaction and balance. Arch Phys Med Rehabil. 2004;85(2):335–8.

97. Kristensen MT, Bloch ML, Jønsson LR, Jakobsen TL. Interrater reliability of the standardized Timed Up and Go Test when used in hospitalized and community-dwelling older individuals. Physiother Res Int J Res Clin Phys Ther. 2019;24(2):e1769.

98. Marques A, Cruz J, Quina S, Regêncio M, Jácome C. Reliability, Agreement and Minimal Detectable Change of the Timed Up & Go and the 10-Meter Walk Tests in Older Patients with COPD. COPD. 2016;13(3):279–87.

99. Choi YM, Dobson F, Martin J, Bennell KL, Hinman RS. Interrater and intrarater reliability of common clinical standing balance tests for people with hip osteoarthritis. Phys Ther. 2014;94(5):696–704.

100. Andersen CW, Kristensen MT. Performance Stability and Interrater Reliability of Culturally Adapted 10-Meter Walking Test for Danes with Neurological Disorders. J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc. 2019;28(9):2459–67.

101. Newcomer K, Krug H, Mahowald M. Validity and reliability of the timed-stands test for patients with rheumatoid arthritis and other chronic diseases. J Rheumatol. 1993;20(1):21-7.(1):21–7.

102. Nilsagård Y, Andreasson M, Carling A, Vesterlin H. Examining the validity and sensitivity to change of the 5 and 10 sit-to-stand tests in people with multiple sclerosis. Physiother Res Int. 2017;22(4):e1681.

103. Ransdell LB, Wells CL. Physical activity in urban white, African-American, and Mexican-American women. Med Sci Sports Exerc. 1998;30(11):1608–15.

104. Herazo-Beltrán Y, Pinillos Y, Vidarte J, Crissien E, Suarez D, García R. Predictors of perceived barriers to physical activity in the general adult population: a cross-sectional study. Braz J Phys Ther. 2017;21(1):44–50.

105. Ofori EK, Angmorterh SK. Relationship between physical activity, body mass index (BMI) and lipid profile of students in Ghana. Pan Afr Med J. 2019. Link: http://www.panafrican-med-journal.com/content/article/33/30/full/

106. World Health Organization. Physical Activity. 2018. Link: http://www.who.int/ncds/prevention/physical-activity/gappa 17-09-2018

107. De Antonio M, Dogan C, Hamroun D, Mati M, Zerrouki S, Eymard B et al. Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. Rev Neurol (Paris). 2016;172(10):572–80.

108. Petitclerc É, Hébert LJ, Desrosiers J, Gagnon C. Lower limb muscle impairment in myotonic dystrophy type 1: the need for better guidelines. Muscle Nerve. 2015;51(4):473–8.

109. Gagnon C, Mathieu J, Noreau L. Life habits in myotonic dystrophy type 1. J Rehabil Med. 2007;39(7):560–6.

110. Jakobsson M, Brisby H, Gutke A, Lundberg M, Smeets R. One-minute stair climbing, 50-foot walk, and timed up-and-go were responsive measures for patients with chronic low back pain undergoing lumbar fusion surgery. BMC Musculoskelet Disord. 2019;20(1):137.

14. PAPERS

I

Intra-rater Reliability and Validity of Outcome Measures in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Andersen H, Witting N, and Vissing J

Neurology 2020; 94: e2508-e2520. doi:10.1212/WNL.0000000000009625

II

Responsiveness of Outcome Measures in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Witting N, and Vissing J

Annals of Clinical and Translational Neurology 2020; 7(8): 1382-1391

III

Physical Activity in Myotonic Dystrophy Type 1

Knak KL, Sheikh AM, Witting N, and Vissing J

Journal of Neurology 2020; 267: 1679-1686

PAPER I

Intra-rater Reliability and Validity of Outcome Measures in Myotonic Dystrophy Type I

Knak KL, Sheikh AM, Andersen H, Witting N, and Vissing J

Published in *Neurology*, 2020

ARTICLE

Intrarater reliability and validity of outcome measures in myotonic dystrophy type 1

Kirsten L. Knak, PT, MSc, Aisha M. Sheikh, PT, MSc, Henning Andersen, MD, PhD, Nanna Witting, MD, PhD, and John Vissing, MD, PhD

Correspondence
Ms. Knak
kirsten.lykke.knak@regionh.dk

Neurology® 2020;94:e2508-e2520. doi:10.1212/WNL.0000000000009625

Abstract

Objective

To investigate intrarater reliability and concurrent and construct validity of muscle strength, balance, and functional mobility measures in individuals with noncongenital myotonic dystrophy type 1 (DM1).

Methods

Seventy-eight adults with noncongenital DM1 participated in visit 1, and 73 of the them participated in visit 2 separated by 1 to 2 weeks. The assessments consisted of muscle strength tests with handheld dynamometry (HHD) and stationary dynamometry in the lower limb. The balance tests consisted of the step test, Timed Up and Go test, feet-together stance, tandem stance, 1-leg stance, and modified Clinical Test of Sensory Integration and Balance on a balance platform. The functional mobility tests consisted of the 10-m walk test (10mWT) and 10-times Sit-to-Stand test.

Results

The HHD and stationary dynamometry had sufficient intrarater reliability for most muscle groups on a group ($SEM_{\%}$ ≤15%) and individual (minimal detectable difference [$MDD_{95\%}$] ≤30%) level, but the HHD was most reliable. Stationary dynamometry measured a higher torque than HHD for all extensor muscles, but for single individuals, none of the devices were favored. Overall, intrarater reliability and validity were sufficient only for the dynamic balance tests, not the static balance tests. Both functional mobility tests were sufficiently reliable and valid, but the 10mWT was most reliable.

Conclusion

Overall, HHD is recommended as a reliable and valid tool for single individuals and for flexor muscles on a group level. For balance assessments, the dynamic balance tests are recommended as the most valid and reliable balance tests. Both functional mobility tests are recommended for valid and reliable outcomes, but the 10mWT was superior for reliability.

From the Department of Neurology (K.L.K., A.M.S., N.W., J.V.), Rigshospitalet, Copenhagen; and Department of Neurology (H.A.), Aarhus University Hospital, Denmark.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

CI = confidence interval; **DM1** = myotonic dystrophy type 1; **HHD** = handheld dynamometry; **MCID** = minimal clinically important difference; **mCTSIB** = modified Clinical Test of Sensory Integration and Balance; **MDD** = minimal detectable difference; **STS** = Sit-to-Stand test; **10mWT** = 10-m walk test; **TUG** = Timed Up and Go test.

Myotonic dystrophy type 1 (DM1) has a prevalence of 10 in 100,000 in Europe and affects primarily distal limb muscles.[1] There is no cure for DM1,[2] but disease-modifying drugs are under development.[2] Reliable and valid outcome measures are therefore needed to establish meaningful efficacy of these new treatments. Intrarater reliability is the ability of a method to measure a repeated variable consistently free from error with the same rater at retest,[3] and validity is the ability of a method to measure what it is designed to measure.[4] Intrarater reliability in DM1 has been investigated for some balance tests,[5] functional mobility tests,[5] and muscle strength tests.[6,7] However, a larger cohort is needed to verify the findings of the small studies, and no studies exist for the modified Clinical Test of Sensory Integration and Balance (mCTSIB) on a balance platform, 10-times Sit-to-Stand test (STS), and muscle strength dynamometry of isometric torque in most of the muscle groups in the lower limb in DM1. Studies of validity in DM1 are restricted to a few studies of isometric knee muscle strength,[7] 10-m walk test (10mWT),[8] and dynamic balance tests,[8] but studies of most muscle groups in the lower limb, STS, and static balance tests are needed in DM1.

The objective was therefore to investigate intrarater reliability and concurrent and construct validity of balance, functional mobility, and muscle strength measures in a large cohort of individuals with DM1.

Methods

Subjects

Seventy-eight adult patients with noncongenital DM1 were recruited from November 2017 to September 2018 at Rigshospitalet and Aarhus University Hospital (a flowchart is available from Dryad, appendix A, doi.org/10.5061/dryad.h18931zfq). Thus, the general recommendation of a sample size of at least 50 participants for comparison-[9] and reliability studies[10] was fulfilled. The inclusion criteria were as follows: (1) genetically verified DM1 with CTG repeats >80, (2) age of 18 to 60 years, (3) ability to rise from a chair without using arms, (4) ability to perform 10-m walk with or without walking aids, and (5) living around Copenhagen or Aarhus. Exclusion criteria were (1) onset of disease before 1 year of age, (2) inability to understand instructions, (3) other disorders or medicine that could confound interpretation of outcome measures, (4) drug or alcohol abuse within 3 months, and (5) pregnancy.

Study design

The cohort study was conducted at the 2 participating Departments of Neurology. Two visits, separated by 1 to 2 weeks, were performed at the same time of the day to avoid circadian variation in performance. At retest, participants and assessors were blinded to the test results at visit 1 to eliminate motivational or tester bias. Assessments were divided into 4 ordered categories with assessments associated with the highest fall risk executed first: (1) dynamic balance, (2) static balance, (3) functional mobility, and (4) muscle strength. Excel software block-randomized the tests within each category (balance, functional mobility, and muscle strength) into 6 different combinations of order to avoid systematic fatigue bias in some of the tests. The same order and procedures were repeated at retest. Calibration of equipment was validated every 2 weeks. Two physiotherapists conducted the assessments, and the same assessor retested each participant for intrarater reliability. Subjects refrained from exhausting or unaccustomed physical activities the day before testing and were instructed to wear closed, flat, comfortable shoes. The participants received short verbal and visual instructions to minimize influence of a possible cognitive impairment in DM1. The dominant leg, defined as the preferred leg for kicking a ball, was tested. Subjects used the same walking aids at both visits.

Clinical measurements

A detailed protocol is available from Dryad (appendix B, doi.org/10.5061/dryad.h18931zfq). The dynamic balance measurements included Timed Up and Go test (TUG) and step test. The static balance measurements included mCTSIB on a balance platform, feet-together stance, tandem stance, and 1-leg stance with eyes open and closed. Functional mobility measurements consisted of 10mWT (maximum pace) and 10-times STS. For all balance and functional mobility measurements, 2 trials were performed without practice trials.

Muscle strength was assessed with maximal voluntary isometric contraction in the dominant leg (flexor and extensor muscles at ankle, knee, and hip) with handheld dynamometry (HHD) (microFET2; Hoggan Scientific, LLC, Salt Lake City, UT) and stationary dynamometry (Biodex System 3 and 4 PRO; Biodex Medical Systems, Upton, NY). For all muscle strength measurements, 2 practice trials and 3 recorded trials were conducted with standardized encouragement. For comparison of muscle strength across individuals, muscle strength was measured in torque (Newton-meter). For HHD, Newton-meter was calculated by multiplying the lever arm (the distance from the joint center to point of HHD application) by Newton. For psychometric comparability between HHD and stationary dynamometry, muscle specific conditions (velocity and muscle length) and methodologic conditions (number of trials, strength unit, duration, encouragement, etc) were similar for both devices.

Statistical analysis

The statistical significance level was defined as $p \leq 0.05$. If the model assumptions for the analyses were improved with logarithmetic transformation, the analyses were conducted on log-transformed data and afterward back-transformed into ratio (percentage) with the equivalent antilog.

Intrarater reliability

Systematic error was analyzed with a parametric 2-tailed paired-samples *t* test. Intrarater reliability can be divided into relative and absolute reliability. Relative reliability describes relative percentage of reliability (range 0.00–1.00[11]) and was analyzed with the intraclass correlation (2-way random-effects model, absolute agreement, 95% confidence interval [CI], single measures). The higher the intraclass correlation was, the higher the relative reliability was. Absolute reliability describes the agreement (Bland-Altman plots) between repeated measurements or the absolute value of measurement error (e.g., Nm) (SEM and minimal detectable difference [MDD]). The higher the agreement was between tests and the lower the absolute measurement error was, the higher the degree of absolute reliability was. Agreement between visits was visualized by Bland-Altman plots using the following equation[12]:

$$S(x, y) = \left(\frac{S_1 + S_2}{2}, S_1 - S_2\right)$$

where S_1 = visit 1 and S_2 = visit 2. The smallest change beyond measurement error that represents a real change was investigated by the SEM for a group and by MDD for a single individual[13] by applying the following equations[14,15]:

$$SEM = \frac{SD_{diff}}{\sqrt{2}}$$

$$SEM\% = \frac{SEM}{mean} \times 100$$

$$MDD_{95} = SEM \times 1.96 \times \sqrt{2}$$

$$MDD_{95\%} = \frac{MDD}{mean} \times 100$$

where SD_{diff} is the SD of the difference score (visit 1 – visit 2) and mean = [(visit 1 + visit 2)/2].

The SD of the difference between visits is useful for a priori sample size calculation for efficacy trials,[16] and the $MDD_{95\%}$ is beneficial for clinicians to evaluate whether a change in an individual's condition is genuine. Low absolute measurement error/high degree of intrarater reliability is especially important in contexts in which change is of interest (e.g., follow-up in the clinic, before and after an intervention in clinical trials), so genuine clinically meaningful changes can be captured.

For a group, an $SEM_{\%}$ ≤15% has previously been accepted to capture smaller changes in people with stroke,[17] which was also applied in this study. The measurement error is inherently larger for single individuals; therefore, the previously applied guideline of an $MDD_{95\%}$ ≤30% in people with Parkinson disease[18] was applied for single individuals in the present study.

Concurrent and construct validity

Concurrent criterion-related validity was used to investigate agreement between HHD and stationary dynamometry using a paired-samples *t* test and Bland-Altman plots (stationary dynamometry – HHD). Construct validity was used for balance and functional mobility tests to investigate the ability of the tests to reflect the underlying theoretical concept.[4] Thus, the ability of the tests to predict other tests with similar aspects (e.g., balance) was calculated with simple linear regression and adjusted coefficient of determination (R^2). The association between tests with similar aspects was calculated with the Pearson product-moment correlation coefficient (r). A ceiling and floor effect was defined as >15% of the participants achieving the highest or lowest possible score, respectively.[19] Because values above or beyond the maximum or minimum score cannot be detected, ceiling and floor effects introduce measurement inaccuracy.

A higher degree of validity might especially be requested for purposes of enrollment in clinical trials based on inclusion and exclusion criteria, cross-sectional studies, and descriptive studies in which performance from 1 visit is investigated.

Other analyses

Minimal clinically important difference (MCID) is the smallest difference considered important to a patient.[20] MCID was calculated as both absolute and percentage value[21]:

$$MCID = \frac{SD_{baseline}}{2}$$

$$MCID\% = \frac{MCID}{baseline} \times 100$$

A previous study has suggested that MDD < MCID is acceptable while MDD > MCID is troublesome.[22] Therefore, *an existing arbitrary guideline* was applied to estimate the MCID.[21]

Differences in age, body mass index, and Muscular Impairment Rating Scale score between participants who completed the study and dropouts and individuals who declined to participate were investigated by unpaired *t* tests. Difference in sex was investigated by χ^2 test of homogeneity.

Standard protocol approvals, registrations, and patient consents

Approval from the Regional Committee on Health Research Ethics in Denmark (H-17017556) and written informed consent were obtained.

Data availability

Data not published in the article are available from Dryad (doi.org/10.5061/dryad.h18931zfq).

Results

Table 1 provides demographic data. The results were for the best value, including outliers, and based on the original, untransformed data unless otherwise stated. Inference statistic was not calculated for feet-together stance, tandem stance, and 1-leg stance eyes open due to null inflation (i.e., no true variation between participants). Individual differences for intrarater reliability are shown in figure 1. Relative reliability and MCID are shown in tables 2 and 3.

The participants who completed the study (visits 1 + 2) were similar in age to the dropouts ($p = 0.081$) but were younger (40 ± 10 years [mean ± SD]) than individuals who declined to participate (47 ± 9 years) ($p = 0.004$). There was no difference regarding sex in dropouts ($p = 0.613$) and individuals who declined to participate ($p = 1.00$). Body mass index was lower for participants who completed the study (median 23.6 kg/m^2, interquartile range 20.7–27.5 kg/m^2) compared to dropouts (32.8 kg/m^2, interquartile range 28.7–36.9 kg/m^2), but disease severity (Muscular Impairment Rating Scale score) was similar between the groups ($p = 0.48$).

Muscle strength measurements

Group level

Compared to stationary dynamometry, HHD showed no systematic errors and lower or similar absolute measurement error for groups and therefore was more reliable (table 2).

Compared to HHD on a group level regarding validity, stationary dynamometry measured a statistically significantly higher torque with a difference of 12.53 Nm (95% CI 8.65–16.41) in ankle plantar flexors, 32.93 Nm (95% CI 25.04–40.82) in knee extensors, and 63.05 Nm (95% CI 51.12–74.98 Nm) in hip extensors but a statistically significantly lower torque with a difference of −8.71 Nm (95% CI −11.88 to −5.53) in knee flexors. There was no statistically significant difference in ankle dorsal flexors and hip flexors. Consequently, on a group level, stationary dynamometry seemed to better capture force in the extensor muscles.

Individual level

HHD had higher absolute reliability than stationary dynamometry because HHD showed lower absolute measurement error for single individuals (table 2).

On an individual level regarding validity, muscle strength differed noticeably between HHD and stationary dynamometry, but none of the dynamometers were superior to the other within the 95% expected normal range (figure 2). Thus, to measure peak torque, both devices can be used but not interchangeably.

Balance measurements

Group level

The dynamic balance measures (TUG, step test) showed the lowest absolute measurement error on a group level but a highly statistically significant learning effect (table 3). This contrasted with the static balance measures (mCTSIB, 1-leg stance eyes closed) but especially with the 1-leg stance eyes closed test, which showed very poor reliability (table 3). Therefore, the dynamic balance measures showed the best reliability.

Validity was not investigated for 1-leg stance eyes closed because of the very poor reliability because this will inherently result in poor validity.[9] A ceiling effect was detected with 92% of participants obtaining maximum score in the feet-together stance, 55% on the tandem stance, and 30% on the 1-leg stance eyes open, and a floor effect was detected in participants obtaining a minimum score on the mCTSIB (37%) and 1-leg stance eyes closed (42%). For 1-leg stance eyes closed, a surrogate measure was used because 42% were unable to stand ≥30 seconds in 1-leg stance eyes open and therefore did not qualify for the 1-leg stance eyes closed test.[5] Thus, poor validity was shown for the static balance tests.

A poorer performance of balance (predictor variable) predicted a poorer performance of balance and muscle strength (outcome variables), and a better performance of balance predicted a better performance of muscle strength. A 1-second increase in TUG predicted a statistically significant reduction of 1.66 steps (95% CI −2.35 to −0.98) in the step test and accounted for 22.6% of the variation in the step test. A doubling of time in TUG predicted a statistically significant decrease of 37.9% (95% CI −58.8% to −6.4%) in lower extremity muscle strength and accounted for 7.0% of the variation in lower extremity muscle strength. An increase of 1 step in the step test predicted a statistically significant increase of 5.5% (95% CI 3.2%–7.7%) in ankle muscle strength and accounted for 28.6% of the variation in ankle muscle strength. Predictions are presented in table 4. The correlations between balance tests and tests assessing similar aspects ranged from little ($\tau_b = -0.220$) to moderate ($r = 0.546$) (table 5). Therefore, the validity seemed reasonable for the dynamic balance tests.

Individual level

The findings of absolute intrarater reliability and construct validity for single individuals were similar to the findings on a group level.

Functional mobility measurements

Group level

The 10mWT was more absolutely reliable than STS on a group level (table 3).

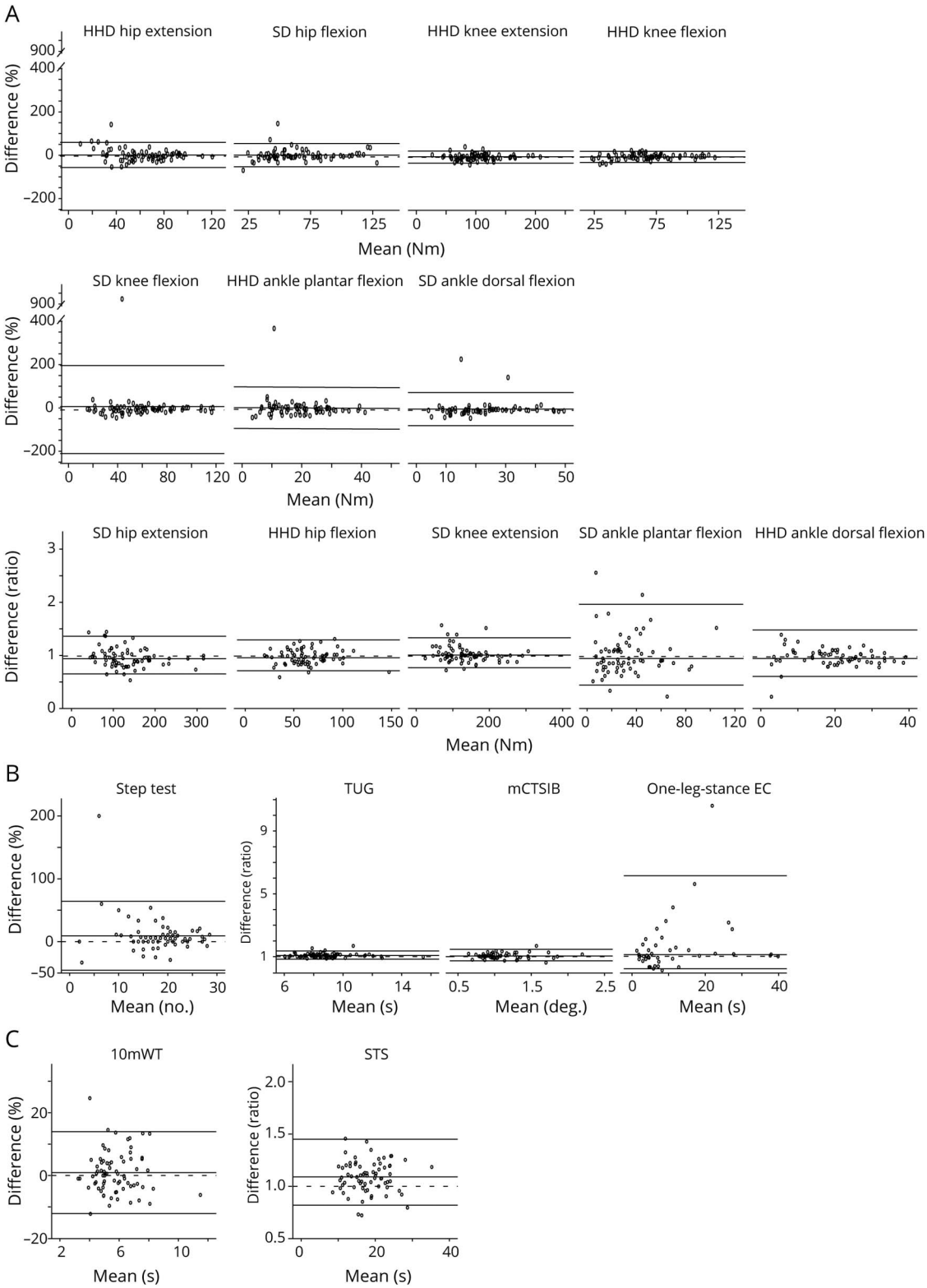
For validity, a poorer performance of functional mobility predicted a poorer performance of other tests of functional

Table 1 Demographic data and intrarater consistency

Demographic data (visit 1, n = 78)	
Sex, n	78
Women	39
Men	39
Age, mean (SD), y	40 (10)
BMI, median (2.5–97.5 percentile), kg/m^2	24 (16–37)
MIRS score,[a] n	
Grade 1	0
Grade 2	17
Grade 3	3
Grade 4	51
Grade 5	7
Walking aid, n	
Insoles	1
AFO	8
3-Wheeled scooter	1
Cane	1
Walker	2
AES-S score,[b] median (2.5–97.5 percentile)	12 (2–26)
Apathy, n	0
Stroop score,[c] median (2.5–97.5 percentile)	
Word score	32 (17–54)
Cognitive impairment, n	21
Color score	34 (12–61)
Cognitive impairment, n	13
Color-word score	38 (25–58)
Cognitive impairment, n	2
Interference score	50 (46–54)
Cognitive impairment, n	0
Intrarater consistency (visits 1 + 2, n = 73)	
Time between test and retest, median (2.5–97.5 percentile), d	7 (6–39)
Time of day difference between test and retest, median (2.5–97.5 percentile), h	0.5 (0–4)

Abbreviations: AES-S = Apathy Evaluation Scale; AFO = ankle-foot orthosis; BMI = body mass index; HHD = handheld dynamometry; MIRS = Muscular Impairment Rating Scale.
The demographic data are presented for the individuals who participated in visit 1. The data did not change noticeably for those who continued to visit 2, and the data are therefore not shown. Missing values for Stroop color and Stroop inferences (n = 16) were due to test not being implemented (n = 8), withdrawal before the assessment (n = 4), test not conducted (n = 1), participant was color blind (n = 2), and participant was unable to read (n = 1). Missing values for Stroop color/word (n = 17) were due to test not being implemented (n = 8), withdrawal before the assessment (n = 4), test not conducted (n = 1), incomplete test (n = 1), participant was color blind (n = 2), and participant was unable to read (n = 1).
[a] Grade 1 = no muscular impairment, grade 2 = minimal signs, grade 3 = distal weakness, grade 4 = mild to moderate proximal weakness, grade 5 = severe proximal weakness.[43]
[b] AES-S is self-rated. A score >34 = apathy. Missing values (n = 15) due to test was not implemented (n = 8), withdrawal before the assessment (n = 4), test was not conducted (n = 1), and incomplete tests (n = 2).
[c] Verbal Stroop color and word test (adult version). Higher score means better cognitive performance. The 95% confidence interval for Stroop was 30.4 to 69.91. Missing values for Stroop word (n = 17) were due to test not being implemented (n = 8), withdrawal before the assessment (n = 4), test not conducted (n = 1), invalid score (n = 1), participant was color blind (n = 2), and participant was unable to read (n = 1).

Figure 1 Bland-Altman plots of intrarater reliability of muscle strength, balance, and functional mobility measures



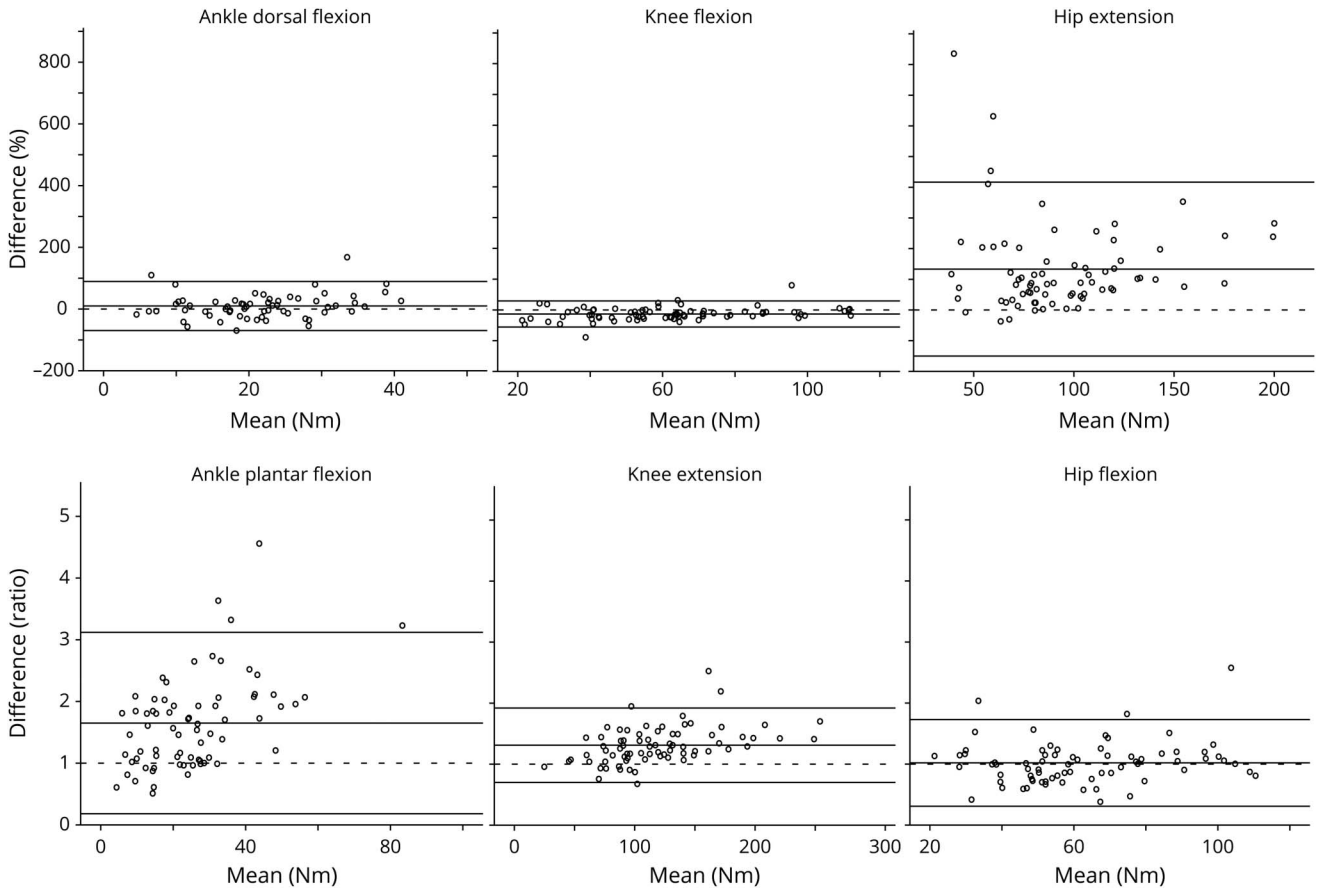
(A) Muscle strength measures, (B) balance measures, and (C) functional mobility measures. The x-axis: (visit 1 + visit 2)/2; y-axis: visit 1 − visit 2; difference (percent): original data; and difference (ratio): $\log_{10}$ arithmetic transformed data. A thick circle represents >1 participant. The closer the circles (participants) are to zero for untransformed difference (percent) and to 1.00 for $\log_{10}$ arithmetic back-transformed difference (ratio), the higher the agreement is between visits 1 and 2, indicating less absolute measurement error. Data on knee flexion were for both devices expressed on the difference (percent) scale. EC = eyes closed; HHD = handheld dynamometry; mCTSIB = modified Clinical Test of Sensory Integration and Balance; SD = stationary dynamometry; STS = Sit-to-Stand test; 10mWT = 10-m walk test; TUG = Timed Up and Go test.

Table 2 Intrarater reliability and MCID of muscle strength measures

	No.[f]	V1 mean (SD)	V2 mean (SD)	Difference, mean (SD)	*p* Value	SEM ($SEM_{\%}$)	MDD_{95} ($MDD_{95\%}$) [$MDD_{95\%,\ without\ outliers}$][g]	ICC (95% CI)	MCID ($MCID_{\%}$)
Stationary dynamometry, Nm									
Ankle plantar flexion	68	26.30 (5.69–90.08)[a]	27.45 (3.64–96.83)[a]	1.04 (0.95, 1.14); 4 (−5, 14)[b]	0.348	±1.30 (±30%)[c]	±2.07 (±107%)[c] [±71%]	0.79[d] (0.67–0.86)	1.43 (43%)[e]
Ankle dorsal flexion	58	20.50 (5.05–49.50)[a]	20.75 (4.15–47.96)[a]	0.88 (4.81)	0.170	±3.40 (±15%)	±9.43 (±41%) [±17%]	0.91[d] (0.86–0.95)	1.32 (32%)[e]
Knee extension	72	136.91 (59.93)	135.31 (62.65)	0.98 (0.94, 1.01); −2 (−6, 1)[b]	0.452	±1.10 (±10%)[c]	±1.32 (±32%)[c] [±24%]	0.96[d] (0.93–0.97)	29.97 (22%)
Knee flexion	72	57.32 (25.94)	59.66 (27.09)	2.33 (11.12)	0.080	±7.87 (±13%)	±21.80 (±37%) [±25%]	0.91[d] (0.86–0.94)	12.98 (23%)
Hip extension	71	128.51 (58.12)	137.31 (65.48)	1.05 (1.01, 1.10); 5 (1, 10)[b]	0.003	±1.14 (±14%)[c]	±1.45 (±45%)[c] [±38%]	0.92[d] (0.86–0.95)	29.06 (23%)
Hip flexion	71	63.11 (25.82)	67.18 (28.39)	4.07 (13.52)	0.013	±9.56 (±15%)	±26.50 (±41%) [±25%]	0.87[d] (0.79–0.92)	12.91 (20%)
HHD, Nm									
Ankle plantar flexion	69	19.02 (9.31)	19.36 (9.10)	0.34 (3.60)	0.436	±2.55 (±13%)	±7.06 (±37%) [±31%]	0.92[d] (0.88–0.95)	4.66 (25%)
Ankle dorsal flexion	64	19.48 (9.18)	19.88 (9.74)	0.40 (2.43)	0.198	±1.72 (±9%)	±4.77 (±24%) [NA]	0.97[d] (0.95–0.98)	4.59 (24%)
Knee extension	71	102.81 (35.83)	102.73 (37.31)	−0.08 (13.49)	0.962	±9.54 (±9%)	±26.44 (±26%) [±22%]	0.93[d] (0.89–0.96)	17.92 (17%)
Knee flexion	72	66.01 (24.14)	67.22 (25.78)	1.21 (7.28)	0.165	±5.15 (±8%)	±14.27 (±21%) [NA]	0.96[d] (0.93–0.97)	12.06 (18%)
Hip extension	70	62.36 (24.36)	62.34 (23.84)	−0.02 (11.84)	0.989	±8.37 (±13%)	±23.21 (±37%) [NA]	0.88[d] (0.82–0.92)	12.18 (20%)
Hip flexion	70	63.07 (22.95)	64.86 (24.62)	1.03 (0.99, 1.07); 3 (−1, 7)[b]	0.173	±1.11 (±11%)[c]	±1.34 (±34%)[c] [±31%]	0.89[d] (0.84–0.93)	11.48 (18%)

Abbreviations: CI = confidence interval; HHD = handheld dynamometry; ICC = intraclass correlation; MCID = minimal clinically important difference; MDD = minimal detectable difference; V1 = visit 1; V2 = visit 2.
[a] Median (2.5–97.5 percentiles) were presented because mean ± SD crossed zero, which is nonsense for descriptive values for V1 + V2. It was not because of abnormal distribution.
[b] $Antilog_{10}$: ratio geometric mean (95% CI); percent geometric mean (95% CI). Therefore, the geometric mean does not equal V2 − V1.
[c] $Antilog_{10}$: ratio SEM/MDD (95% CI); percent SEM/MDD (95% CI).
[d] $p < 0.05$.
[e] $Antilog_{10}$: ratio MCID; percent MCID.
[f] Stationary dynamometry ankle plantar flexion: missing values (n = 5) due to inability to activate the test at V1 but not at V2 (n = 4), inability to activate the test at V2 but not at V1 (n = 2), inability to activate the test at V1 + V2 (n = 3), ankle-foot orthosis at V1 (n = 1), and missing first value (n = 1). Thus, the intersection was n = 68. HHD ankle plantar flexion: missing values (n = 4) due to Medical Research Council score <3 (n = 3) and participant stopped before HHD assessment (n = 1). Stationary dynamometry ankle dorsal flexion: missing values (n = 15) due to inability to activate test (n = 15). HHD ankle dorsal flexion: missing values (n = 9) due to Medical Research Council score <3 (n = 8) and participant stopped before HHD (n = 1). Stationary dynamometry knee extension: missing value (n = 1) due to technical issues (n = 1). HHD ankle-knee extension: missing values (n = 2) due to tester unable to hold position (n = 1) and participant stopped before HHD (n = 1). Stationary dynamometry knee flexion: missing value (n = 1) due to technical issues (n = 1). HHD ankle-knee flexion: missing values (n = 1) due to participant stopped before HHD (n = 1). Stationary dynamometry hip flexion: missing values (n = 2) due to participant stopped before the test (n = 1) and technical issues (n = 1). HHD hip flexion: missing values (n = 2) due to tester unable to hold position (n = 1) and participant stopped before HHD (n = 1). Stationary dynamometry hip extension: missing values (n = 2) due to participant stopped before the test (n = 1) and technical issues (n = 1). HHD hip extension: missing values (n = 2) due to tester unable to hold position (n = 1) and participant stopped before HHD (n = 1).
[g] Outliers: stationary dynamometry: ankle plantar flexion (n = 4), ankle dorsal flexion (n = 5), knee extension (n = 6), knee flexion (n = 1), hip extension (n = 3), and hip flexion (n = 8). HHD: ankle plantar flexion (n = 2), ankle dorsal flexion (n = 0), knee extension (n = 3), knee flexion (n = 0), hip extension (n = 0), and hip flexion (n = 3).

Figure 2 Bland-Altman plots of validity of muscle strength measures



x-axis: (stationary dynamometry + handheld dynamometry [HHD])/2; y-axis: stationary dynamometry – HHD; difference (percent): original data; and difference (ratio): $\log_{10}$ arithmetic transformed data. Dotted line indicates no difference between stationary dynamometry and HHD. Three undotted lines indicate mean difference ±2 SD. Each circle represents 1 participant. The closer the circles are to 0 for difference (percent) and to 1 for difference (ratio), the higher the agreement is between stationary dynamometry and HHD, indicating higher validity.

mobility and muscle strength. A 1-second increase in STS predicted a statistically significant increase of 0.168 seconds (95% CI 0.127–0.209) in the 10mWT and accounted for 46.5% of the variation in the 10mWT. In addition, an increase of 1 second in STS predicted a statistically significant reduction of 2.5% (95% CI −4.1% to −0.8%) in proximal lower extremity muscle strength and accounted for 10.3% of the variation in proximal lower extremity muscle strength. A 1-second increase in 10mWT predicted a statistically significant reduction of −13.9% (95% CI −19.8% to −7.5%) in lower extremity muscle strength and accounted for 22.4% of the variation in lower extremity muscle strength. Table 5 provides the predictions. The correlations between functional mobility tests and tests assessing similar aspects ranged from fair (r = −0.340) to good (r = 0.687) (table 5). Overall, the validity seemed fair for the functional mobility measurements.

Individual level

The absolute reliability and construct validity findings on a group level also applied to single individual level.

Discussion

The main findings were that HHD was more reliable than stationary dynamometry for both groups and single individuals, but for validity, stationary dynamometry was superior for the extensor muscles on a group level but not on an individual level. The dynamic balance tests were more reliable and valid than the static balance tests, and both of the functional mobility tests were valid, but the 10mWT was most reliable for groups and single individuals.

Unexpectedly, HHD showed lower or similar measurement error on both group and individual levels in the lower limb compared to stationary dynamometry, which is generally considered to be the gold standard. This new finding can be explained partly by the standardized, fixed position of the assessor for the HHD assessment. In addition, participant cooperation is more easily perceived by an assessor than by a machine.

The $SEM_{\%}$ was ≤15% and thus acceptable for all muscles with both HHD and stationary dynamometry, except for ankle

Table 3 Intrarater reliability and MCID of balance and functional mobility measures

	No.[h]	V1 mean (SD)	V2 mean (SD)	Difference mean (SD)	*p* Value	SEM ($SEM_{\%}$)	MDD_{95} ($MDD_{95\%}$) [$MDD_{95\%,\ without\ outliers}$][i]	ICC (95% CI)	MCID ($MCID_{\%}$)
TUG, s	73	8.44 (1.84)	8.08 (1.65)	0.94 (0.91, 0.96); −6 (−9, −4)[a]	<0.0005	±1.09 (±9%)[b]	±1.26 (±26%)[b] [±19%]	0.68[c] (0.54–0.79)	0.92 (11%)
Step test, n	73	17.63 (5.73)	18.63 (5.72)	1.00 (2.40)	0.001	±1.70 (±9%)	±4.70 (±26%) [±19%]	0.90[c] (0.82–0.94)	2.87 (16%)
mCTSIB, degrees	53	1.00 (0.27)	0.96 (0.26)	0.96 (0.91, 1.01); −4 (−9, 1)[a]	0.096	±1.13 (±13%)[b]	±1.42 (±42%)[b] [±34%]	0.81[c] (0.70–0.89)	0.14 (14%)
Feet-together stance, s	73	60.00 (15.81–60.00)[d]	60.00 (13.70–60.00)[d]	e	e	e	e	e	e
Tandem stance, s	73	40.00 (1.60–40.00)[d]	40.00 (3.14–40.00)[d]	e	e	e	e	e	e
1-Leg stance eyes open, s	72	40.00 (0.85–40.00)[d]	40.00 (1.15–40.00)[d]	e	e	e	e	e	e
1-Leg stance eyes closed, s	42	5.78 (1.53–40.00)[f]	5.75 (1.24–39.45)[f]	−0.91 (0.69, 1.20); −9 (−31,20)[a]	0.50	±1.86 (±86%)[b]	±5.58 (±458%)[b] [±256%]	0.62[c] (0.40–0.77)	1.70 (70%)[g]
STS, s	72	17.51 (5.58)	16.26 (4.86)	0.92 (0.89, 0.95); −8 (−11, −5)[a]	<0.0005	±1.11 (±11%)[b]	±1.34 (±34%)[b] [±27%]	0.85[c] (0.72–0.91)	2.79 (16%)
10mWT, s	73	5.52 (1.37)	5.55 (1.34)	0.04 (0.37)	0.384	±0.26 (±4%)	±0.72 (±12%) [±12%]	0.96[c] (0.94–0.98)	0.69 (13%)

Abbreviations: CI = confidence interval; ICC = intraclass correlation; MCID = minimal clinically important difference; mCTSIB = modified Clinical Test of Sensory Integration and Balance; MDD = minimal detectable difference; STS = Sit-to-Stand test; 10mWT = 10-m walk test; TUG = Timed Up and Go test; V1 = visit 1; V2 = visit 2.
[a] $Antilog_{10}$: ratio geometric mean (95% CI); percent geometric mean (95% CI). Therefore, the geometric mean does not equal V2 − V1.
[b] $Antilog_{10}$: ratio MDD (95% CI); percent MDD (95% CI).
[c] $p < 0.05$.
[d] Median (2.5–97.5 percentiles).
[e] Impossible to estimate due to null inflation (i.e., no true variation across participants).
[f] Median (2.5–97.5 percentiles) were presented because mean ± SD crossed zero, which is nonsense for descriptive values for V1 + V2. It was not because of abnormal distribution.
[g] $Antilog_{10}$: ratio MCID; percent MCID.
[h] TUG and ST: no missing values (n = 0). mCTSIB: missing values (n = 20) due to technical issues (n = 1), inability of participants to complete the mCTSIB test at V1 + V2 (n = 14), incomplete test at V1 but not V2 (n = 4), and incomplete test at V2 but not V1 (n = 7). Thus, the intersection is 53. Feet-together stance and tandem stance: no missing values (n = 0). One-leg stance eyes open: missing value (n = 1) due to inability to initiate the test (n = 1). One-leg stance eyes closed: missing values (n = 31) due to inability to initiate the 1-leg stance eyes open (n = 1) and inability to stand ≥30 seconds in 1-leg stance eyes open (n = 30), which qualified for the 1-leg stance eyes closed according to Hammarén et al.[5] STS: missing value (n = 1) due to incomplete test (n = 1). 10mWT: no missing values (n = 0).
[i] Outliers: TUG (n = 3), ST (n = 6), mCTSIB (n = 2), 1-leg stance eyes closed (n = 4), STS (n = 3), and 10mWT (n = 4).

plantar flexors with stationary dynamometry ($SEM_{\%}$ = 30%). The findings of SEM for HHD and stationary dynamometry were overall comparable to previous findings in patients with DM1[6,7] and healthy controls.[22–26]

Concerning concurrent validity, stationary dynamometry measured a higher torque than HHD only in the generally stronger extensor muscles on a group level, which has been shown before in DM1.[7] This might be attributed to the fixed, optimal positioning of the assessor using HHD. A similar tendency with higher muscle strength detected by devices using more fixation has been shown in healthy individuals.[25] The lower torque assessed by HHD was likely attributed to the lower strength of the assessor vs patient or because patients performed concentric contractions when pushing against the assessor.

When outliers were excluded, $MDD_{95\%}$ was ≤31% and acceptable for most of the muscles for both HHD and stationary dynamometry (except ankle plantar flexors and hip extensors with stationary dynamometry and hip extensors with HHD). Possibly, the unfamiliar positioning of the ankle and hip muscles led to more variable performances and larger measurement error. The $MDD_{95\%}$ in the present study was overall comparable to HHD in populations without DM1[23–29] and to stationary dynamometry in healthy individuals.[22] For both devices, the MDD values exceeded the MCID values for almost all muscle groups, but it was remarkable only in the hip muscles. This suggests that the measurement error for both devices in the hip muscles may conceal clinical important changes in single individuals. Few genuine outliers widened the 95% CI for measurement error for single individuals on the Bland-Altman plots for some of the distal muscle groups,

Table 4 Regression and adjusted coefficient of determination of balance and functional mobility measures

Independent variable, predictor (x)	Dependent variable, outcome (y)	Slope of regression line (β)	95% CI for β	*p* Value for β	Adjusted R^2	*p* Value for adjusted R^2
TUG	Step test	−1.664	−2.352 to −0.997	<0.0005	0.226	<0.0005
Step test	Muscle strength (ankle plantar and dorsal flexion)	1.055[a]	1.032[a]–1.077[a]	<0.0005	0.286	<0.0005
TUG[a]	Muscle strength (hip extension, knee extension, and ankle dorsal flexion)	0.621[a]	0.412[a]–0.936[a]	0.023	0.070	0.023
STS	10mWT	0.168	0.127–0.209	<0.0005	0.465	<0.0005
STS	Muscle strength (extension in hip and knee)	0.975[a]	0.9599[a]–0.992[a]	0.003	0.103	0.003
10mWT	Muscle strength (hip extension, knee extension, ankle plantar and dorsal flexion)	0.861[a]	0.802[a]–0.925[a]	<0.0005	0.224	<0.0005

Abbreviations: CI = confidence interval; STS = Sit-to-Stand test; 10mWT = 10-m walk test; TUG = Timed Up and Go test.
The muscle strength measures are values from stationary dynamometry (torque, mean of the best value from each muscle group). The data are original data unless otherwise stated.
[a] $Antilog_2$: ratio.

but 5% of the observations can be expected to exceed the normal range due to the predefined 95% statistically probability level.

Despite the clinically relevant disagreements between HHD and stationary dynamometry for single individuals regarding validity, the highest torque did not favor any of the devices. Thus, for single individuals, both devices can be applied to measure maximal muscle strength, but they cannot be used interchangeably. This is consistent with findings in healthy individuals.[23]

The finding that mCTSIB on a balance platform was less reliable than the dynamic balance tests (TUG and step test) on the group and individual levels was unanticipated. The reason may be that the mCTSIB captures even subtle changes in postural control. Despite a learning effect in the dynamic balance tests only, the dynamic balance tests are favored because a learning effect is more fixable by adding familiarization trials and thus less problematic than random error (MDD and SEM).

Despite the lower reliability of mCTSIB compared to the dynamic balance tests, all tests demonstrated acceptably low measurement error for groups ($SEM_{\%}$ <15%). Possible reasons for the higher measurement error with mCTSIB were that static and not dynamic balance was tested and the challenges that accompanied the test when it was performed with closed eyes on a foam surface. It has previously been shown that measurement error increases with difficulty of tests.[30] Foot position at retest was inconsistent, but the importance of foot position on balance performance is controversial.[31,32] In the present study, the $SEM_{\%}$ of mCTSIB agreed with previous findings in the elderly.[30] The static 1-leg stance eyes closed test showed very high absolute measurement error on a group level. This may be due to the complexity of balance[33] and the difficulty of the test. This explanation is supported by previous findings of higher MDD for more difficult tests in elderly persons.[30]

For construct validity, it was surprising that all static balance measurements demonstrated either ceiling or floor effects, especially because the test battery covered easy, moderate, and hard difficulty levels. The findings of the present study of a ceiling effect in the feet-together stance and 1-leg stance test eyes open were in concordance with findings in DM1[5] and hip osteoarthritis.[34] The regression and correlation coefficients of the balance tests were only small to moderate, but the construct validity was considered sufficient because comparator tests have only some aspects in common with balance and the sample size was large, which potentially reduces the correlation coefficients.[35] Compared to previous findings in DM1,[8] the present study found similar statistically significant correlations for step test vs muscle strength but smaller statistically significant correlations for TUG vs muscle strength. Because correlation coefficients are mathematically increased with smaller sample size and larger heterogeneity,[35] correlation coefficients are difficult to compare across studies. No studies of predictions by balance measurements have been published.

Similar to the group level, the mCTSIB was less reliable than the dynamic balance test, and the 1-leg stance eyes closed showed markedly high measurement error on an individual level. The section on group level provides possible explanations for the lower reliability. For the mCTSIB, however, the measurement error was acceptably low without outliers ($MDD_{95\%}$ ≤34%), and the MDD results were consistent with findings in elderly persons.[30] The $MDD_{95\%}$ of TUG in the present study was approximately similar to findings in DM1[5] and non-DM1 populations.[18,36,37] For the step test, the

Table 5 Correlation coefficients for balance and functional mobility measures

Outcome measures	Pearson correlation coefficient	*p* Value[a]	95% CI
TUG			
Step test	−0.486	<0.0005[b]	−0.687 to −0.286
STS	0.439	<0.0005[b]	0.292–0.695
Muscle strength (hip extension, knee extension, and ankle dorsal flexion)	−0.313[c]	0.015[b]	NA
Step test			
Muscle strength (ankle plantar and dorsal flexion)	0.546	<0.0005[b]	0.327–0.764
mCTSIB (composite score)			
Step test	−0.251[c]	0.008[b]	NA
Muscle strength (ankle plantar and dorsal flexion)	−0.220[c]	0.022[b]	NA
STS			
10mWT	0.687	<0.0005[b]	0.519 to 0.855
Muscle strength (hip and knee extension)	−0.340	0.003[b]	−0.563 to −0.117
10mWT			
Muscle strength (hip extension, knee extension, and ankle plantar and dorsal flexion)	−0.488	<0.0005[b]	−0.719 to −0.256

Abbreviations: CI = confidence interval; mCTSIB = modified Clinical Test of Sensory Integration and Balance; NA = not applicable; STS = Sit-to-Stand test; 10mWT = 10-m walk test; TUG = Timed Up and Go test.
NA because 95% CI is not provided for the nonparametric Kendall τ_b test. The muscle strength measures are values based on stationary dynamometry (torque, mean of the best value for each muscle group).
[a] Adjustment for multiple testing: because TUG was correlated with 3 variables independently, the statistically significant level has been Bonferroni corrected at α = 0.05/3 = 0.017. Because the step test was correlated with 3 variables independently, the statistically significant level has been Bonferroni corrected at α = 0.05/3 = 0.017. Because STS was correlated with 3 variables independently, the statistically significant level has been Bonferroni corrected at α = 0.05/3 = 0.017. Because mCTSIB (composite score) was correlated with 2 variables independently, the statistically significant level has been Bonferroni corrected at α = 0.05/2 = 0.025. Because 10mWT was correlated with 2 variables independently, the statistically significant level has been Bonferroni corrected at α = 0.05/2 = 0.025.
[b] Statistically significant.
[c] Kendall τ_b.

present study showed an $MDD_{95\%}$ similar to findings in DM1[5] and hip osteoarthritis.[34] The MDD values were only marginally larger than the MCID values for the dynamic balance tests (TUG and step test) but larger for the static balance tests. Thus, there is a risk that the smallest important difference for a single individual might not be captured with the mCTSIB. There was a discrepancy in the number of participants performing the 1-leg stance with eyes open vs eyes closed. The reason for the discrepancy was that all participants attempted the 1-leg stance eyes open, while only the participants who were able to hold this position for ≥30 seconds qualified for the 1-leg stance eyes closed test.

The section on group level provides a discussion of construct validity.

The 10mWT and STS showed acceptably low measurement error for groups ($SEM_{\%}$ <15%), but the measurement error was lowest for the 10mWT. Our 10mWT results were overall consistent with findings in DM1[5] and non-DM1.[37,38] No studies of absolute reliability of STS have been published in any population.

The construct validity was reasonable for both the 10mWT and STS. However, the correlation between STS and muscle strength in hip and knee extensors was only $r = -0.34$. This could possibly be explained by the different types of muscle contractions (concentric and eccentric contractions during STS and isometric contraction during muscle strength testing). The correlation was only slightly smaller than what was found in rheumatoid arthritis.[39] The correlation findings for 10mWT vs muscle strength in DM1 were overall consistent with previous findings in DM1.[8,40] The statistically significant correlations for 10mWT vs STS in the present study agreed with findings in multiple sclerosis.[41] No studies of predictions by functional mobility measurements have been published.

The findings on a group level for 10mWT and STS were also applicable on an individual level with sufficiently low absolute measurement error ($MDD_{95\%}$ <30% without outliers). The

10mWT results were comparable with findings in DM1[5] and non-DM1.[37,38] The MDD was equal to MCID for the 10mWT, but slightly larger for the STS. Thus, minimal meaningful changes in a single individual might be captured, especially with the 10mWT.

The section on group level provides a discussion of construct validity.

For the majority of tests in our study, sufficient intrarater reliability and concurrent or construct validity were found. Cognitive impairment, but not apathy, was present in 29% of the participants and may influence reliability and validity. However, the reliability and validity results in the present study were overall comparable with results in DM1 and non-DM1 populations. Because of tiredness, 4 participants dropped out between the 2 visits. However, this is not considered to be an issue because it was only a low proportion of the participants. Because of the broad spectrum of demographic and disease characteristics and recruitment from 2 sites in Denmark with 2 assessors, the present study is considered generalizable to the DM1 population.

The nonperfect degree of validity in the present study might be influenced by the nonperfect degree of reliability of the measurements.[42] In addition, validity can be deduced for only 1 of 2 comparator measures because validity can only be as good as the poorest measurement.[9]

Data were missing primarily for the muscle strength and balance measurements. Missing values for the muscle strength measurements were due primarily to inability of the participants to perform the test because of muscle weakness; for the balance measurements, they were due mostly to the test protocol (e.g., for mCTSIB, all subtests should be performed to obtain the composite score, and to qualify for the 1-leg stance eyes closed test, the 1-leg stance eyes open position should be held for ≥30 seconds). The estimation of peak torque for each trial by curve reading for stationary dynamometry was associated with possible errors. Because reliability and validity are not inherent to the measurements, the degree of reliability and validity might differ across patients, assessors, and environments,[3] but the present extensive study provides useful guidelines for clinicians and researchers. Because only a few patients were tested outside the time window, this likely did not affect results considering that DM1 is a slowly progressive disease. Another limitation was that only 2 assessors were used, which reduces the generalizability to a wider pool of assessors with different skills.

For groups and single individuals, both HHD and stationary dynamometry showed acceptable intrarater reliability for most muscle groups, but HHD was superior and more feasible. Stationary dynamometry was more valid than HHD for the extensor muscles on a group level, but for single individuals, both devices can be applied all though not interchangeably.

For groups, the dynamic balance tests (TUG and step test) and the static mCTSIB were sufficiently reliable, but for single individuals, the feasible dynamic balance tests are recommended with familiarization trials. In addition, the dynamic balance tests were more valid than the static balance tests.

Both the feasible 10mwt and STS had acceptable reliability and were valid for groups and individuals, but the 10mWT was most reliable. Responsiveness and intertester reliability remain to be shown.

Study funding

The study is not industry sponsored.

Disclosures

K.L. Knak received grants from Axel Muusfeldt's Foundation, Familien Hede Nielsen's Foundation, and Rigshospitalet's Research Foundation. A.M Sheikh, H. Andersen, N. Witting, and J. Vissing report no relevant disclosures. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* May 15, 2019. Accepted in final form December 5, 2019.

Appendix Authors

Name	Location	Contribution
Kirsten L. Knak, PT, MSc	Rigshospitalet, Copenhagen, Denmark	Design and conceptualized study; major role in the acquisition of data; analyzed and interpreted the data; performed statistical analysis; drafted the manuscript for intellectual content
Aisha M. Sheikh, PT, MSc	Rigshospitalet, Copenhagen, Denmark	Major role in the acquisition of data; revised the manuscript for intellectual content
Henning Andersen, MD, PhD	Aarhus University Hospital, Aarhus, Denmark	Major role in the acquisition of data; revised the manuscript for intellectual content
Nanna Witting, MD, PhD	Rigshospitalet, Copenhagen, Denmark	Design and conceptualized study; major role in the acquisition of data; interpreted the data; revised the manuscript for intellectual content
John Vissing, MD, PhD	Rigshospitalet, Copenhagen, Denmark	Design and conceptualized study; major role in the acquisition of data; interpreted the data; revised the manuscript for intellectual content

References

1. Kierkegaard M, Petitclerc E, Hébert LJ, Gagnon C. Is one trial enough for repeated testing? Same-day assessments of walking, mobility and fine hand use in people with myotonic dystrophy type 1. Neuromuscul Disord 2017;27:153–158.
2. Smith CA, Gutmann L. Myotonic dystrophy type 1 management and therapeutics. Curr Treat Options Neurol 2016;18:52.
3. Portney L, Watkins M. Reliability of measurements. In: Foundations of Clinical Research: Applications to Practice, 3rd ed. Cranbury: Pearson Education; 2009.
4. Portney L, Watkins M. Validity of measurements. In: Foundations of Clinical Research: Applications to Practice, 3rd ed. Cranbury: Pearson Education; 2009.

5. Hammarén E, Ohlsson JA, Lindberg C, Kjellby-Wendt G. Reliability of static and dynamic balance tests in subjects with myotonic dystrophy type 1. Adv Physiother 2012;14:48–54.
6. Hébert LJ, Remec JF, Saulnier J, Vial C, Puymirat J. The use of muscle strength assessed with handheld dynamometers as a non-invasive biological marker in myotonic dystrophy type 1 patients: a multicenter study. BMC Musculoskelet Disord 2010;11:1.
7. Roussel MP, Hébert LJ, Duchesne E. Intra-rater reliability and concurrent validity of quantified muscle testing for maximal knee extensors strength in men with myotonic dystrophy type 1. J Neuromuscul Dis 2019;6:233–240.
8. Hammarén E, Kjellby-Wendt G, Kowalski J, Lindberg C. Factors of importance for dynamic balance impairment and frequency of falls in individuals with myotonic dystrophy type 1—a cross-sectional study—including reference values of Timed up & Go, 10m walk and step test. Neuromuscul Disord 2014;24:207–215.
9. Altman D. Some common problems in medical research. In: Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
10. Valentin G, Maribo T. Hand-held dynamometry fixated with a tripod is reliable for assessment of back extensor strength in women with osteoporosis. Osteoporos Int 2014;25:2143–2149.
11. Portney L, Watkins M. Statistical measures of reliability. In: Foundations of Clinical Research: Applications to Practice, 3rd ed. Cranbury: Pearson Education; 2009.
12. Wikipedia. Bland–Altman plot. Available at: en.wikipedia.org/wiki/Bland%E2%80%93Altman_plot. Accessed January 29, 2019.
13. Kierkegaard M, Tollbäck A. Reliability and feasibility of the six minute walk test in subjects with myotonic dystrophy. Neuromuscul Disord 2007;17:943–949.
14. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231–240.
15. Schwenk M, Gogulla S, Englert S, Czempik A, Hauer K. Test–retest reliability and minimal detectable change of repeated Sit-to-Stand analysis using one body fixed sensor in geriatric patients. Physiol Meas 2012;33:1931.
16. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med Auckl NZ 1998;26:217–238.
17. Faria CDCM, Teixeira-Salmela LF, Neto MG, Rodrigues-de-Paula F. Performance-based tests in subjects with stroke: outcome scores, reliability and measurement errors. Clin Rehabil 2012;26:460–469.
18. Huang SL, Hsieh CL, Wu RM, Tai CH, Lin CH, Lu WS. Minimal detectable change of the Timed "Up & Go" test and the dynamic gait index in people with Parkinson disease. Phys Ther 2011;91:114–121.
19. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
20. Portney L, Watkins M. Concepts of measurement. In: Foundations of Clinical Research: Applications to Practice, 3rd ed. Cranbury:Pearson Education; 2009.
21. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–592.
22. Otten R, Whiteley R, Mitchell T. Effect of subject restraint and resistance pad placement on isokinetic knee flexor and extensor strength: implications for testing and rehabilitation. Sports Health 2013;5:137–142.
23. Martins J, da Silva JR, da Silva MRB, Bevilaqua-Grossi D. Reliability and validity of the belt-stabilized handheld dynamometer in hip- and knee-strength tests. J Athl Train 2017;52:809–819.
24. Thorborg K, Bandholm T, Hölmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. Knee Surg Sports Traumatol Arthrosc 2013;21:550–555.
25. Davis PR, McKay MJ, Baldwin JN, Burns J, Pareyson D, Rose KJ. Repeatability, consistency, and accuracy of hand-held dynamometry with and without fixation for measuring ankle plantarflexion strength in healthy adolescents and adults. Muscle Nerve 2017;56:896–900.
26. Wollin M, Purdam C, Drew MK. Reliability of externally fixed dynamometry hamstring strength testing in elite youth football players. J Sci Med Sport 2016;19:93–96.
27. Buckinx F, Croisier JL, Reginster JY, et al. Reliability of muscle strength measures obtained with a hand-held dynamometer in an elderly population. Clin Physiol Funct Imaging 2017;37:332–340.
28. Kimura M, Ogata Y, Akebi T, Ochi M, Hachisuka K, Saeki S. Development and reliability of a hand-held dynamometer device to measure the ankle dorsiflexor muscle strength. J Rehabil Assist Technol Eng 2018;5:2055668318802558.
29. Mañago MM, Hebert JR, Schenkman M. Psychometric properties of a clinical strength assessment protocol in people with multiple sclerosis. Int J MS Care 2017;19:253–262.
30. Scaglioni-Solano P, Aragón-Vargas LF. Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and sensory integration in highly functional older adults. Int J Rehabil Res 2014;37:138–143.
31. Riemann BL, Lininger M, Kirkland MK, Petrizzo J. Age related changes in balance performance during self-selected and narrow stance testing. Arch Gerontol Geriatr 2018;75:65–69.
32. Wrisley DM, Whitney SL. The effect of foot position on the modified Clinical Test of Sensory Interaction and Balance. Arch Phys Med Rehabil 2004;85:335–338.
33. Shumway-Cook A, Woollacott M. Normal postural control. In: Motor Control: Translating Research Into Clinical Practice, 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2007.
34. Choi YM, Dobson F, Martin J, Bennell KL, Hinman RS. Interrater and intrarater reliability of common clinical standing balance tests for people with hip osteoarthritis. Phys Ther 2014;94:696–704.
35. Portney L, Watkins M. Correlation. In: Foundations of Clinical Research: Applications to Practice, 3rd ed. Cranbury: Pearson Education; 2009.
36. Kristensen MT, Bloch ML, Jønsson LR, Jakobsen TL. Interrater reliability of the standardized Timed Up and Go Test when used in hospitalized and community-dwelling older individuals. Physiother Res Int J Res Clin Phys Ther 2019;24:e1769.
37. Marques A, Cruz J, Quina S, Regêncio M, Jácome C. Reliability, agreement and minimal detectable change of the Timed Up & Go and the 10-meter walk tests in older patients with COPD. COPD 2016;13:279–287.
38. Andersen CW, Kristensen MT. Performance stability and interrater reliability of culturally adapted 10-meter walking test for Danes with neurological disorders. J Stroke Cerebrovasc Dis 2019;28:2459–2467.
39. Newcomer K, Krug H, Mahowald M. Validity and reliability of the timed-stands test for patients with rheumatoid arthritis and other chronic diseases. J Rheumatol 1993;20:21–27.
40. Jimenez-Moreno AC, Nikolenko N, Kierkegaard M, et al. Analysis of the functional capacity outcome measures for myotonic dystrophy. Ann Clin Transl Neurol 2019;6:1487–1497.
41. Nilsagård Y, Andreasson M, Carling A, Vesterlin H. Examining the validity and sensitivity to change of the 5 and 10 Sit-to-Stand tests in people with multiple sclerosis. Physiother Res Int 2017;22:e1681.
42. Bland JM, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–160.
43. Mathieu J. Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. Neurology 2001;56:336–340.

PAPER II

Responsiveness of Outcome Measures in Myotonic Dystrophy Type I

Knak KL, Sheikh AM, Witting N, and Vissing J

Published in *Annals of Clinical and Translational Neurology*, 2020

ANNALS of Clinical and Translational Neurology
Open Access
AMERICAN NEUROLOGICAL ASSOCIATION

RESEARCH ARTICLE

Responsiveness of outcome measures in myotonic dystrophy type 1

Kirsten L. Knak, Aisha M. Sheikh, Nanna Witting & John Vissing

Rigshospitalet, Department of Neurology, Juliane Maries Vej 28, Copenhagen,DK-2100, Denmark

Correspondence
Kirsten L. Knak, Rigshospitalet, Department of Neurology, Copenhagen, Denmark. Tel: +45 35456217; Fax: 35 45 61 38; E-mail: kirsten.lykke.knak@regionh.dk

Received: 21 April 2020; Revised: 7 June 2020; Accepted: 18 June 2020

***Annals of Clinical and Translational Neurology* 2020; 7(8): 1382–1391**

doi: 10.1002/acn3.51129

Abstract

Objective: As myotonic dystrophy type 1(DM1) evolves slowly and interventional trials often have a short duration, responsive outcomes in DM1 are needed. The objective of this study was to determine the responsiveness of muscle strength, balance, and functional mobility measurements after a 1-year follow-up period in individuals with DM1. **Methods**: Sixty-three adults with noncongenital DM1 completed the following assessments at baseline and at 1-year follow-up: Handheld dynamometry (lower limbs), stationary dynamometry (lower limbs), step test, timed-up-and-go test (TUG), modified clinical test of sensory integration and balance (mCTSIB), feet-together stance, tandem stance, one-leg stance, 10-meter walk test, and sit-to-stand test. **Results**: Change was captured by stationary dynamometry (proximal flexor and extensor muscles), handheld dynamometry (proximal flexor and distal extensor muscles), TUG, and mCTSIB ($P \leq 0.04$). Ceiling or floor effects were shown for most static balance tests. **Interpretation**: Overall, adequate responsiveness was shown for both muscle strength dynamometers, TUG and mCTSIB. These outcomes are therefore likely candidate endpoints for clinical trials lasting 1 year. Most static balance tests are not responsive and not recommended in a heterogeneous DM1 population.

Introduction

Myotonic dystrophy type 1 (DM1) is the most common muscular dystrophy in adults,[1] characterized by distal limb, facial and bulbar muscle weakness, myotonia and multisystemic affection involving cognitive impairment, cardiac disease, metabolic abnormalities, and cataracts.[2] There is an unmet need for evidence-based outcomes in individuals with DM1[3,4] as disease-modifying clinical trials are emerging.[5] Validity and reliability of muscle strength, balance, and functional mobility outcomes have recently been established in DM1,[6] but knowledge about responsiveness is lacking, which hampers the possibility to design and pick appropriate endpoints for interventional trials. Responsiveness is a tool's ability to detect change in a condition over time.[7] Responsiveness of timed-up-and-go test (TUG) and handheld dynamometry (HHD) in the lower limbs has previously been investigated in DM1,[8] but the follow-up period was 9 years. Because clinical trials often have a maximum duration of 1 year, the challenge is to identify responsive endpoints within 1 year despite the slowly progressive nature of DM1.[4] Only the 30-second sit-to-stand test (STS) and 10-meter walk test (10mWT, walk/run max pace) have been investigated after 1 year.[9] Responsiveness of commonly used endpoints such as stationary dynamometry, step test, feet-together stance, tandem stance, one-leg stance, modified clinical test of sensory integration and balance (mCTSIB), 10mWT (walk, fast pace), and 10-times STS in DM1 is still unknown.

The objective of this study was to investigate responsiveness of muscle strength, balance, and functional mobility measurements after 1 year in individuals with noncongenital DM1.

Methods

Patients

From November 2017 to September 2019, 63 individuals with DM1 were recruited from a DM1 cohort[6] at the Rigshospitalet (n = 60) and Aarhus University Hospital (n = 3) in Denmark (see Fig. 1). The inclusion criteria were genetically confirmed DM1 (CTG repeats> 80), 18-60 years, able to stand up from a chair with no arm

support, able to walk at least 10 meters (with or without gait aids) and reside close to Copenhagen or Aarhus. The exclusion criteria were congenital DM1 (defined as disease onset before 1 year), cognitive impairment preventing test adherence, non-DM1-related disorders or medicine consumption which confound muscle strength, balance, or functional test results, abuse of drugs or alcohol within 3 months, serious medical illness (e.g., symptomatic coronary artery disease and cancer), pregnancy, and clinically significant medical illness within 30 days.

Clinical assessments

Assessment of muscle strength, balance, and functional mobility were done twice separated by 1 year. All assessments have previously been described in detail.[6] The same order of tests and procedures were repeated at follow-up by the same assessor for each patient at the same time of the day. Neither the patient nor the assessor were blinded to the test results, but recall bias is limited after 1 year, and therefore likely did not influence the results. The patients were asked to wear closed, flat comfortable shoes and asked to refrain from exhausting or unusual physical activity the day before each visit to eliminate bias from muscle soreness or fatigue.

Muscle strength measurements

Maximal isometric muscle torque was tested with HHD (microFET2, Hoggan Scientific, LLC, Salt Lake City, UT) and stationary dynamometry (Biodex System 3 or 4 PRO, Biodex Medical Systems, NY). Newton from HHD was converted to Newton-meter: *Newton-meter (Nm) = Newton (N) * meter (m)*. Muscle strength was tested over the ankle, knee, and hip joints in the dominant leg. Two practice trials followed by three recorded trials were performed with standardized encouragement.

Balance measurements

The static balance measurements comprised of 60-second feet-together stance, 40-second tandem stance, 40-second one-leg-stance with eyes open and closed, and mCTSIB

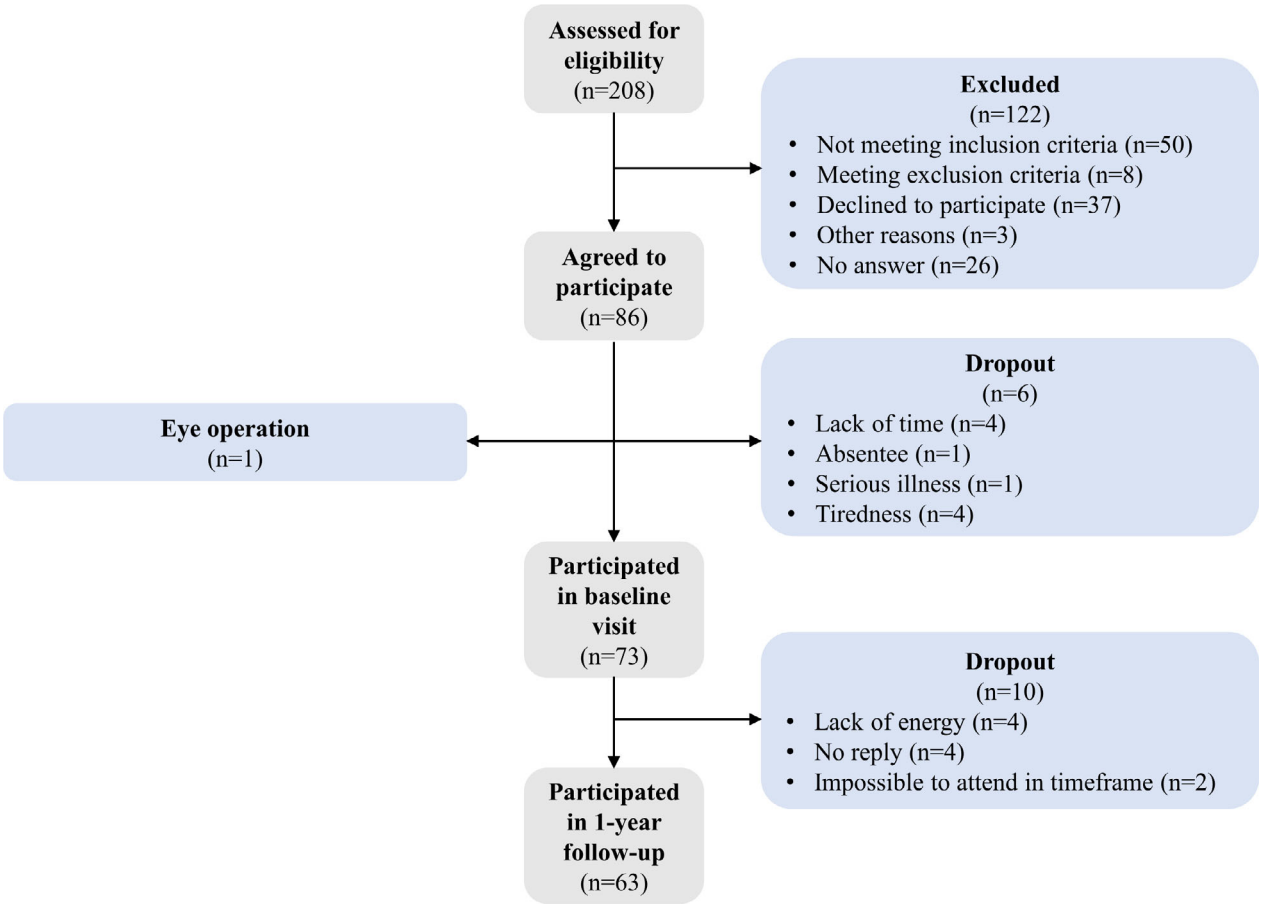


Figure 1. Flowchart of recruitment of patients in this study.

(eyes open and closed on a firm and foam surface) on a balance platform (BioSway Portable Balance System 950-460, Biodex Medical Systems, NY). The dynamic balance measurements consisted of 15-second step test and TUG. Two recorded trials were conducted with no encouragement. All balance measurements were performed with comfortable shoes and insoles and ankle–foot orthosis (AFO) were allowed. Other habitual walking aids such as a cane were only allowed for the TUG.

Functional mobility measurements

The functional mobility measurements included 10-times STS and 10mWT (both at the fastest possible pace). Two recorded trials were conducted with no encouragement.

Questionnaires

The International Physical Activity Questionnaire (short version) was applied to investigate the physical activity level.[10] At follow-up, the patients completed a patient-reported global rating scale (GRS) to investigate whether a change had occurred or not from a subjective perspective, which the objective outcomes were compared to. The GRS questions for the objective tests were as follows: (I) Ankle muscle strength tests: Has your muscle strength in the lower limb/crus changed since the last visit (more difficulties standing on toes, tendency to stumble, slapping foot, walking longer distances)? (II) Knee and hip muscle strength tests: Has your muscle strength in the thigh and buttock changed since the last visit (more difficulties with rising from a chair (use of arms), climbing stairs (use of arms), walking longer distances? (III) Dynamic balance tests: Has your balance during movement changed since the last visit (tendency to fall, need to lean on objects)? (IV) Static balance tests: Has your balance when you are standing still changed since the last visit (tendency to fall, need to lean on objects)? (V) STS test: Has your ability to rise from a chair changed since the last visit (more difficulties with rising from a chair (use of arms))? (VI) 10mWT: Has your ability to walk shorter distances changed since the last visit (e.g., slower walking at home)? The possible answers were as follows: (I) Much deterioration, (II) Some deterioration, (III) Stability, (IV) Some improvement, or (V) Much improvement. Each objective test was compared to the GRS-question that addressed the investigated construct of the objective test.

Statistical analysis

Linear mixed model was conducted to investigate statistically significant change between baseline and follow-up (mean $\pm$ SE) with family as a random effect, visit as a covariate and with unstructured covariance to account for repeated measurements over time in the same patients. If the model assumptions were not fulfilled, data were $\log_2$-arithmetic transformed for analyses and antilog_2-arithmetic back-transformed for interpretation. In case of genuine outliers, sensitivity analysis of data with and without outliers was conducted. Based on a previous study,[11] floor and ceiling effects were defined as >15% of the patients scoring the lowest or highest score, respectively.

Secondary analyses

Receiver operating characteristic curve and area under the curve were conducted using GRS as anchor for whether a change had occurred or not. The model assumptions were checked, and the five GRS categories were dichotomized into Worse versus Stable/Better. Area under the curve estimates how good the measurements are to correctly classify change or no change compared to the anchor,[12] and the following guideline was applied[13]: 0.50 = no discrimination; 0.50-0.70 = poor discrimination; 0.70-0.80 = acceptable discrimination; 0.80-0.90 = excellent discrimination; 0.90-1.00 = outstanding discrimination.

Subanalyses of patients able to perform ankle dorsal flexion in the stationary dynamometry were done, because a floor effect was found in 21% of patients. This is relevant for clinical trials with an inclusion criterion of preserved dorsal ankle flexion strength to perform dynamometry assessment. Subanalyses based on age, age at onset of disease and CTG repeat size might also be relevant, but was deselected to avoid the risk of mass-significance with multiple testing.

A difference between patients who completed the study versus patients who declined to participate or dropped out was tested by unpaired t-test (continuous data), Mann–Whitney test (ordinal data), and Fisher's exact test (dichotomous data) if the model assumptions were fulfilled.

Mann–Whitney test was used to analyze if there was a difference in cognition and apathy among patients with agreement and disagreement between the subjective GRS ratings and the objective measurements, respectively. Because these secondary analyses were only exploratory, Bonferroni correction was not applied.

Ethics

The Regional Committee of Health Research Ethics in Denmark approved the study (H-17017556) and informed written consent was obtained.

Table 1. Demographic data.

Sex, no.	
Female	30
Male	33
Age (years), mean (SD)	41 (10)
BMI[1], median (IQR[2])	24 (21-27)
MIRS[3], no.	
Grade 1	0
Grade 2	13
Grade 3	2
Grade 4	42
Grade 5	6
Walking aid, no.	
Insoles	1
AFO[4]	8
Three-wheeled scooter	1
Cane	1
Walker	1
AES-S[5], median (IQR[2])	12 (8-16)
Apathy, no.	0
STROOP[6], median (IQR[2])	
Word score	32 (27-37)
Cognitive impairment, no.	24
Color score	34 (31-40)
Cognitive impairment, no.	11
Color-Word score	37 (34-45)
Cognitive impairment, no.	3
Interference score	50 (50-51)
Cognitive impairment, no.	0

Missing values for STROOP word (n = 11) due to test was not implemented (n = 7), invalid score (n = 1), patient was color blind (n = 2), and patient was unable to read (n = 1).
Missing values for STROOP color, color/word and inferences (n = 10) due to test was not implemented (n = 7), patient was color blind (n = 2), and patient was unable to read (n = 1).
[1]BMI = Body mass index (kg/m^2), $\frac{weight(kg)}{height(m^2)}$.
[2]IQR = Interquartile range.
[3]MIRS = Muscular impairment rating scale. Grade 1 = no muscular impairment, grade 2 = minimal weakness, grade 3 = distal weakness, grade 4 = mild to moderate proximal weakness, grade 5 = severe proximal weakness.[22]
[4]AFO = Ankle–foot orthosis.
[5]AES-S = Apathy evaluation scale (Self-rated). A score> 34=apathy. Missing values (n = 8) due to test was not implemented (n = 7) and incomplete test (n = 1).
[6]Verbal STROOP color and word test (Adult version). A higher score means better cognitive performance. The 95% CI for normal cognition score measured by STROOP is 30.4 to 69.91.

Results

For demographic data, see Table 1. The follow-up visit was performed after a median of 12 months (IQR 11.75-12.5 months), and the time of day between baseline and follow-up varied with a median of 0.5 hours (IQR 0.25-2.00 hours). For feet-together stance, tandem stance, and one-leg-stance eyes open, inference statistics were not calculated because of null-inflation (i.e., no true variation among patients) due to ceiling effects.

Muscle strength measurements

A change in muscle strength from baseline to follow-up was captured by stationary dynamometry in the flexor and extensor muscles over the knee and hip joints, and by HHD in the ankle plantar flexors, knee flexors (without 2 outliers), and hip flexors ($P \leq 0.03$) (Table 2).

Twenty-one percent of the patients were unable to overcome the threshold in the stationary dynamometry for ankle dorsal flexors. Subanalysis without these patients did not change the results regarding change.

Balance measurements

All static balance tests, except mCTSIB, showed either a ceiling effect (maximum score) or a floor effect (minimum score) at both baseline and follow-up in many patients (Fig. 2). The balance tests that were able to detect changes at follow-up were the dynamic balance test TUG and the static balance test mCTSIB ($P \leq 0.035$) (Table 3).

Functional mobility measurements

None of the functional mobility tests (10mWT and STS) captured a change at follow-up ($P \geq 0.88$) (Table 3).

Secondary analyses

Outcome measures against GRS

The objective outcome measure results were generally not reflected in the subjective perceptions of change or no change as measured by GRS, because only the 10mWT and ankle plantar flexors with stationary dynamometry reached acceptable agreement with the GRS (area under the curve> 0.70, Fig. 3).

For the 10mWT, the cognition was better in the patients with disagreement between 10mWT change and GRS change (median, IQR; 43.00, 39.25-48.00) versus the patients with agreement between the 10mWT and the GRS (38.25, 35.50-41.13) ($P = 0.02$). For the ankle dorsal flexors with HHD, apathy was less pronounced in patients with disagreement between HHD change and GRS change (median, IQR; 9, 6-12) compared to the patients with agreement between the HHD and the GRS

Table 2. Baseline and follow-up muscle strength measures and their change.

	Baseline Mean (SD)	FU Mean (SD)	Change Mean (95% CI) for absolute and percentage differences	*P*-value
Stationary dynamometry[1] (Nm)				
Ankle plantar flexors	27.25 (14.80; 39.20)[2]	26.95 (14.60; 47.10)[2]	1.03 (0.94; 1.12); 3% (−6%; 12%)[3]	0.52
Ankle dorsal flexors	16.80 (7.70; 27.30)[2]	17.15 (8.50; 27.80)[2]	(0.97; 1.06); 1% (−3%; 6%)[3]	0.50
Knee extensors	139.78 (64.49)	133.50 (66.73)	−6.02 (−11.06; −0.98); −4.31% (−7.91%; −0.70%)	0.02*↓
Knee flexors	61.14 (27.22)	58.20 (26.37)	−2.87 (−4.95; −0.79); −4.69% (−8.09%; −1.30%)	0.009*↓
Hip extensors	128.40 (98.20; 176.70)[2]	106.75 (72.00; 143.60)[2]	−28.65 (−37.88; −19.42); −20.16% (−26.66%; −13.67%)	<0.0001*↓
Hip flexors	69.42 (28.42)	85.59 (30.45)	16.35 (12.41; 20.29); 23.55% (17.88%; 29.23%)	<0.0001*↑
HHD[4] (Nm)				
Ankle plantar flexors	18.08 (10.37; 25.72)[2]	15.96 (10.36; 24.65)[2]	−1.33 (−2.35; −0.31); −7.17% (−12.66%; −1.68%)	0.01*↓
Ankle dorsal flexors	18.21 (5.70; 25.74)[2]	18.14 (6.66; 27.54)[2]	(0.93; 1.10); 1% (−7%; 10%)[3]	0.79
Knee extensors	105.30 (38.24)	108.58 (41.77)	3.61 (−1.31; 8.53); 3.43% (−1.24%; 8.10%)	0.16
Knee flexors	68.14 (26.41)	65.73 (25.60)	−2.27 (−5.03; 0.49); −3.33% (−7.39%; 0.72%)	0.11
Hip extensors	64.26 (22.63)	59.93 (19.98)	−3.58 (−7.32; 0.16); −5.57% (−11.40%; 0.25%)	0.065
Hip flexors	65.32 (20.43)	61.30 (22.36)	−3.83 (−7.16; −0.50); −5.86% (−10.96%; −0.76%)	0.03*↓

Note. **P*-value ≤ 0.05. ↑improvement, ↓deterioration. The 95% CI for difference is based on SE. FU = 1-year follow-up.
[1]Stationary dynamometry: Ankle plantar flexors: missing values (n = 1) because of technical issues. Ankle dorsal flexors: missing values (n = 1) because of technical issues. Knee extensors: missing values (n = 2) because of not conducted due to knee pain (n = 1) and technical issues (n = 1). Knee flexors: missing values (n = 1) because of technical issues. Hip extensors: missing values (n = 1) because of technical issues. Hip flexors: missing values (n = 1) because of technical issues.
[2]Median (IQR) because mean – (1.96*SD) resulted in a negative value which is meaningless.
[3]Antilog$_2$: ratio geometric mean (95% CI); percentage mean (95% CI).
[4]HHD: Ankle plantar flexors: missing values (n = 0). Ankle dorsal flexors: missing values (n = 0). Knee extensors: missing values (n = 1) because of knee pain. Knee flexors: missing values (n = 0). Hip extensors: missing values (n = 2) because of patient compensation (n = 1) and tester unable to hold position (n = 1). Hip flexors: missing values (n = 1) because tester was unable to hold position (n = 1).

(13, 10-17) ($P = 0.048$). For the other outcomes, there was no difference regarding cognition or apathy between the two groups ($P \geq 0.078$).

Comparison with dropouts

There was no difference between patients who completed the study versus dropouts regarding sex, age, BMI, and level of muscle affection (Muscular Impairment Rating Scale (MIRS)) ($P \geq 0.15$). The patients who completed the study were younger (41 ± 10 years) than the patients who declined to participate (47 ± 9 years) ($P = 0.008$), but there was no difference in sex ($P = 0.49$).

Physical activity level

There was no significant difference in physical activity level from baseline (median, IQR, 375.0, 200.0 to 620.0 min) to follow-up (median, IQR, 372.5, 180.0 to 810.0 min) ($P = 0.32$).

Discussion

The main findings of this study are that muscle strength and the balance measurements TUG and mCTSIB are responsive to change over a 1-year period in a cohort of noncongenital DM1. In contrast, all other static balance

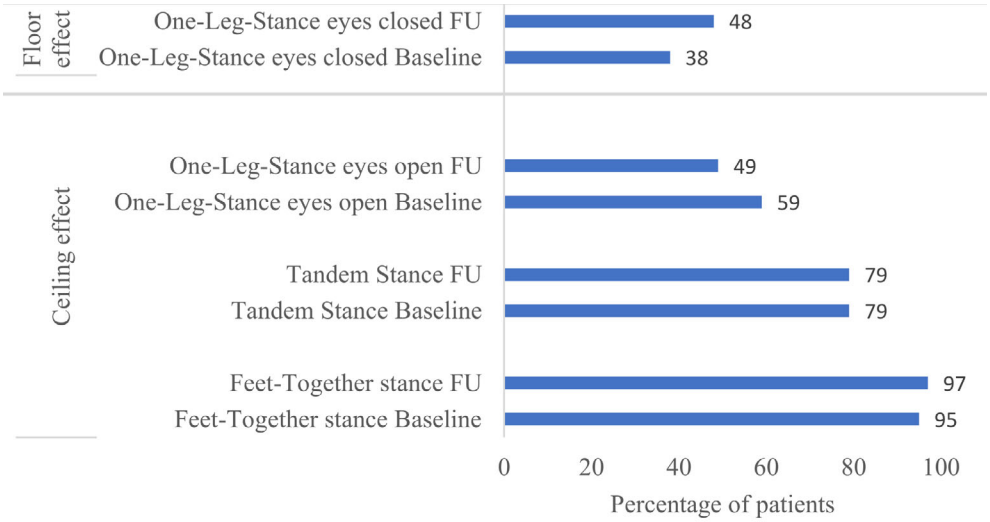


Figure 2. Ceiling and floor effects. The percentage of patients with either ceiling or floor effects is shown for both baseline and 1-year follow-up (FU).

measurements than the mCTSIB are unresponsive because of either ceiling or floor effects.

Muscle strength measurements

Both stationary and handheld dynamometers captured a change in the proximal leg muscles. The study shows that although distal muscles are well-known to be more affected in DM1, proximal muscles capture change better than distal, likely because distal muscles are very weak in DM1 and have reached an end-stage. Thus, 3-21% of the patients were too weak to exceed the threshold of the dynamometers, and therefore unresponsive to change. Moreover, 76% of the patients demonstrated proximal weakness (MIRS $\geq$ 4). Although HDD and stationary dynamometry agreed for most measures, there was some discrepancy for a few tests. Thus, change in strength was only recorded by HHD ankle plantar flexors, which could relate to easier patient-tester cooperation and smaller strength variation within the patients assessed by HHD, whereas the change in knee extensors, which was only captured by stationary dynamometry, could be due to tester-independency of stationary dynamometry. A significant increase in hip flexor strength assessed by stationary dynamometry was found in this study as well as in a previous DM1 study.[8] However, this is considered a spurious type II error finding because of no change in physical activity level in this study and the loss of hip flexor strength recorded by HHD in the present- and a previous DM1 study.[14] This study demonstrated loss of strength in knee and hip flexors and ankle plantar flexors assessed by HHD. Our study did not show change in knee extensors and ankle dorsal flexors with HHD, but studies of longer duration in DM1 have shown this.[8,14] It has previously been shown that 1 year is too short to register progression in these muscles in DM1.[15] Compared to previous findings of significant decline in the knee extensors, but not in the knee flexors using stationary dynamometry in DM,[16] this study found a reduction of strength in both knee extensors and flexors. The discrepancy may be because Lindemann et al.[16] did not specify DM-type, investigated isokinetic torque, and had a smaller sample size, which reduces power.

The HHD has limitations when a subject is stronger than the investigator, but in this study, this problem was minimal as only one patient could overcome the assessor. Reversely, HHD is superior to stationary dynamometry as illustrated by 21% of the patients who could not exceed the threshold in the stationary dynamometry, whereas this number was only 13% for HHD.

Balance measurements

The TUG and mCTSIB captured change, but the step test and one-leg-stance eyes closed test, measuring different aspects of the same construct, failed to do so. The failure of capturing a change in the one-leg-stance eyes closed test was probably caused by a floor effect, missing values, and heterogeneity of the patients' performances. After 5–9 years of observation, deterioration in DM1 has been shown not only in the TUG[8,14] but also in the step test.[14]

The ceiling effects in the feet-together stance, tandem stance, and one-leg-stance eyes open and the floor effect

Table 3. Baseline and follow-up balance and functional mobility measures and their change.

	Baseline Mean (SD)	FU Mean (SD)	Change Mean (95% CI) for absolute and percentage differences	*P*-value
Dynamic balance				
TUG[1] (s)	8.11 (1.68)	8.45 (1.41)	0.35 (0.17; 0.53); 4.32% (2.14%; 6.49%)	0.0003*↓
Step Test[2] (no.)	18.53 (5.77)	18.77 (6.12)	0.24 (–0.27; 0.75); 1.30% (–1.45%; 4.05%)	0.37
Static balance				
mCTSIB[3] (deg.)	0.99 (0.25)	1.05 (0.25)	0.06 (0.001; 0.119); 6.06% (0.12%; 12.00%)	0.035*↓
Feet-together stance[4] (s)	60 (60; 60)[5]	60 (60; 60)[5]	NA[6]	NA[6]
Tandem stance[7] (s)	40 (40; 40)[5]	40 (40; 40)[5]	NA[6]	NA[6]
One-leg-stance eyes open[8] (s)	40.00 (11.87; 40.00)[5]	38.59 (12.19; 40.00)[5]	NA[6]	NA[6]
One-leg-stance eyes closed[9] (s)	5.09 (3.70; 12.47)[10]	7.83 (3.60; 17.56)[10]	1.14 (0.84; 1.54); 14% (–16%; 54%)[11]	0.42
Functional mobility				
10mWT[12] (s)	5.588 (1.33)	5.582 (1.54)	–0.009 (–0.15; 0.13); –0.16% (–2.62%; 2.29%)	0.90
STS[13] (s)	16.15 (4.32)	16.09 (5.02)	–0.06 (–0.75; 0.63); –0.37% (–4.62%; 3.88%)	0.88

Note. *P-value ≤ 0.05. ↓deterioration. The 95% CI for difference is based on SE. FU = 1-year follow-up.
[1]TUG: missing values (n = 0).
[2]Step test: missing values (n = 1) because patient did not finish the test due to knee pain.
[3]mCTSIB: missing values (n = 14) because of technical issues (n = 2), inability to complete the test (n = 10), and inability to initiate the test (n = 2).
[4]Feet-together stance: missing values (n = 0).
[5]Median (IQR) because of null inflation (i.e., no true variation across patients).
[6]Impossible to estimate due to null inflation (i.e., no true variation across patients).
[7]Tandem stance: missing values (n = 0).
[8]One-leg-stance eyes open: missing values (n = 0).
[9]One-leg-stance eyes closed: missing values (n = 31) because of inability to initiate the one-leg-stance eyes open test (n = 1) and inability to stand ≥ 30 s in one-leg-stance eyes open test (n = 30) which qualified for the one-leg-stance eyes closed test.
[10]Median (IQR) because mean – (1.96*SD) resulted in a negative value which is meaningless.
[11]Antilog$_2$: ratio geometric mean (95% CI); percentage mean (95% CI).
[12]10mWT: missing values (n = 0).
[13]STS: missing values (n = 1) because of incomplete test.

in the one-leg-stance eyes closed test suggests that these tests are unresponsive outcomes in a heterogeneous DM1 cohort. A ceiling effect has also previously been shown in the feet-together stance in DM1.[17] The mCTSIB was close to reaching the threshold for floor effects with 13-14% of the patients being unable to complete the test. This indicates that the subparts of mCTSIB with eyes closed and standing on a foam surface may be too challenging for patients with more severe balance impairments.

Functional mobility measurements

None of the functional mobility tests in this study showed change in 1 year, which is at variance with other studies in DM1 using these tests.[9,18] The discrepancies may be due to differences in test methods,[9,18] unspecified DM-type,[18] and a larger sample size.[9] Changes in the 10mWT after a longer period of time has previously been shown in DM1.[14]

General discussion

Outcome measures against GRS

The objective measurements were generally poorly reflected by the subjective scoring of changes in muscle strength, balance, and functional mobility using the GRS after 1 year. This relationship was better matched when the observation period was 9 years,[8] where larger changes occur that can be more easily perceived. However, overall the true agreement between the objective measurements and subjective assessments in this study is somewhere

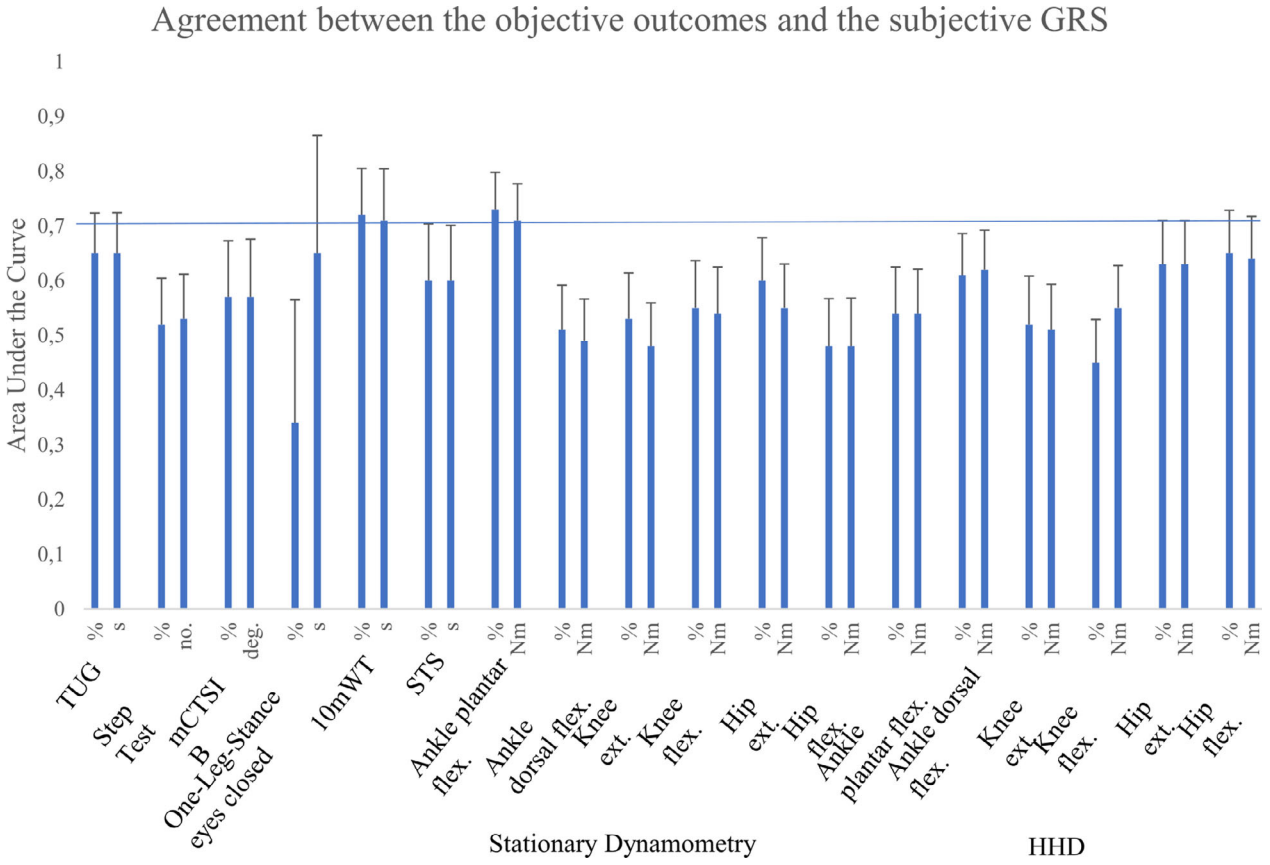


Figure 3. Agreement between the objective muscle strength, balance and functional mobility measurements and the subjective GRS. Area under the curve (Y-axis) is reported for both absolute and relative change values for each outcome measure with 95% CI (X-axis). Flex.=flexors, ext.=extensors.

between no agreement to excellent agreement with 95% confidence. This uncertainty suggests that GRS as an anchor of change or no change is unsuitable in DM1. Thus, defining responsive outcome measures using an anchor for the slowly progressive DM1 disease within a clinical trial duration of 1 year has proven problematic. This may be caused by the several limitations of GRS such as recall bias (inaccuracy of retrieving previous experiences[19]),[8] response-shift bias (a shift in internal perceptions[8]), difficulty to perceive slow, gradual decline,[8] well-being at the day of rating, cognition,[8] and apathy. Moreover, the classification accuracy of change or no change by the objective measurements can only be as good as the anchor. Cognition was impaired in 5-38% of the patients in this study, but none of the patients reached the threshold for apathy (apathy score> 34[20]). However, the patients with disagreement between GRS and the objective measurements did not show lower cognition, which suggests that the impact of cognition may be less than anticipated.

Implications for clinical trials

For clinical trials it is important to select the most responsive outcomes, so that small therapeutic effects are not concealed, and larger study cohorts can be avoided. Thus, based on this study, the best outcomes for clinical trials within 1 year are the stationary dynamometry or HHD with measurements of the proximal muscle groups and the balance assessments TUG or mCTSIB, because these tests capture subtle, but highly significant changes after 1 year of no intervention. For functional mobility, modified or other mobility tests, or novel outcomes such as gait analyses could be investigated for responsiveness to define additional responsive mobility outcomes for clinical trials of 1 year. However, the outcome measures that did not capture change after 1 year with no intervention may still capture a change after a 1-year interventional trial if the treatment is very effective, and therefore can still be considered as outcome. Responsiveness may be improved in a more homogeneous sample,[21] but the

present sample was not large enough for a subgroup analysis.

The strength of the heterogeneous DM1 cohort investigated in this study is that it is generalizable to the majority of the noncongenital DM1 population.

Study limitations

The study was limited by dropouts, but since these constituted only 14% of the sample and did not differ clinically from the completers, this is not considered to influence conclusions significantly. The patient-rated GRS, in contrast to clinician-rated GRS, may be a limitation in DM1 patients due to possible symptoms such as lack of insight, apathy, and impaired cognition, but on the other hand patient-rated GRS represents the patients' own perceptions, which should be acknowledged. Efforts should therefore be directed at developing more suitable patient-reported-outcomes for DM1.

Conclusion

In conclusion, both muscle strength dynamometers and the balance measurements TUG and mCTSIB showed reasonable responsiveness by detecting subtle changes after 1 year in the slowly, progressive disease, DM1. All static balance measurements are not recommended as responsive outcomes due to either ceiling or floor effects, except the mCTSIB.

Acknowledgment

We thank grants from Axel Muusfeldt's Foundation, Familien Hede Nielsen's Foundation, and Rigshospitalet's Research Foundation. We also thank the patients for participating in the study.

Authors' contributions

Aisha Munawar Sheikh contributed with acquisition of data and drafting the manuscript. Nanna Witting contributed with design of the study, analysis of data, and drafting the manuscript. John Vissing contributed with design of the study, analysis of data, and drafting the manuscript.

Conflicts of interest

KLK reports grants from Axel Muusfeldt's Foundation, grants from Familien Hede Nielsen's Foundation, grants from Rigshospitalet's Research Foundation during the conduct of the study. AMS, NW, and JV have nothing to disclose.

References

1. Hammarén E, Kjellby-Wendt G, Kowalski J, Lindberg C. Factors of importance for dynamic balance impairment and frequency of falls in individuals with myotonic dystrophy type 1 – A cross-sectional study – Including reference values of Timed Up & Go, 10m walk and step test. Neuromuscul Disord 2014;24(3):207–215.
2. Andersen G, Ørngreen MC, Preisler N, et al. Muscle phenotype in patients with myotonic dystrophy type 1: Muscle Phenotype in DM1. Muscle Nerve 2013;47 (3):409–415.
3. Foff EP, Mahadevan MS. Therapeutics development in myotonic dystrophy type 1. Muscle Nerve 2011;44 (2):160–169.
4. Gagnon C, Meola G, Hébert LJ, et al. Report of the second Outcome Measures in Myotonic Dystrophy type 1 (OMMYD-2) international workshop San Sebastian, Spain, October 16, 2013. Neuromuscul Disord 2015;25 (7):603–616.
5. Nguyen C-TE, Campbell C. Myotonic dystrophy type 1. Can Med Assoc J 2016;188(14):1033.
6. Knak K, Sheikh A, Andersen H, et al. Intra-rater reliability and validity of outcome measures in myotonic dystrophy type 1. Neurology In press.
7. Portney L, Watkins M. Validity of Measurements. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.
8. Kierkegaard M, Petitclerc É, Hébert LJ, et al. Responsiveness of performance-based outcome measures for mobility, balance, muscle strength and manual dexterity in adults with myotonic dystrophy type 1. J Rehabil Med 2018;50(3):269–277.
9. Jimenez-Moreno AC, Nikolenko N, Kierkegaard M, et al. Analysis of the functional capacity outcome measures for myotonic dystrophy. Ann Clin Transl Neurol 2019;6 (8):1487–1497.
10. International Physical Activity Questionnaire. International Physical Activity Questionnaire [Internet]. 2020. Available from: https://sites.google.com/site/theipaq/questionnaire_links.
11. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42.
12. Portney L, Watkins M. Statistical Measures of Validity. In: Foundations of Clinical Research - Applications to Practice. 3rd Edition. New Jersey: Pearson Education; 2009.
13. Laerd Statistics. ROC Curve [Internet]. Available from: https://statistics.laerd.com/premium/spss/blr/binomial-logistic-regression-in-spss-12.php#roc.
14. Hammarén E, Kjellby-Wendt G, Lindberg C. Muscle force, balance and falls in muscular impaired individuals with myotonic dystrophy type 1: A five-year prospective cohort study. Neuromuscul Disord 2015;25(2):141–148.

15. Sedehizadeh S, Brook JD, Maddison P. Body composition and clinical outcome measures in patients with myotonic dystrophy type 1. Neuromuscul Disord NMD 2017;27 (3):286–289.
16. Lindeman E, Leffers P, Spaans F, et al. Deterioration of motor function in myotonic dystrophy and hereditary motor and sensory neuropathy. Scand J Rehab Med 1995;27:59–64.
17. Hammarén E, Ohlsson JA, Lindberg C, Kjellby-Wendt G. Reliability of static and dynamic balance tests in subjects with myotonic dystrophy type 1. Adv Physiother 2012;14 (2):48–54.
18. Nitz JC, Burns YR, Jackson RV. A longitudinal physical profile assessment of skeletal muscle manifestations in myotonic dystrophy. Clin Rehabil 1999;13(1):64–73.
19. Wikipedia. Recall bias. 2020. Available from: https://en.wikipedia.org/wiki/Recall_bias
20. Shirley Ryan Abilitylab. Apathy Evaluation Scale (Self, Informant, and Clinician Versions). Available from: https://www.sralab.org/rehabilitation-measures/apathy-evaluation-scale-includes-three-forms-self-informant-and-clinician
21. Lencioni T, Piscosquito G, Rabuffetti M, et al. Responsiveness of gait analysis parameters in a cohort of 71 CMT subjects. Neuromuscul Disord NMD 2017;27 (11):1029–1037.
22. Mathieu J. Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. Neurology 2001;56:336–340.

PAPER III

Physical Activity in Myotonic Dystrophy Type I

Knak KL, Sheikh AM, Witting N, and Vissing J

Published in *Journal of Neurology*, 2020

Journal of Neurology (2020) 267:1679–1686
https://doi.org/10.1007/s00415-020-09758-8

ORIGINAL COMMUNICATION

Physical activity in myotonic dystrophy type 1

Kirsten Lykke Knak[1] · Aisha Munawar Sheikh[1] · Nanna Witting[1] · John Vissing[1]

Received: 3 January 2020 / Revised: 10 February 2020 / Accepted: 10 February 2020 / Published online: 17 February 2020

Abstract

Background Physical inactivity is associated with lifestyle diseases and exercise of moderate intensity seems beneficial in DM1, but knowledge about physical activity and predictors of physical activity in individuals with myotonic dystrophy type 1 (DM1) is limited. The objective of this study is to assess physical activity and predictors of physical activity in individuals with DM1.

Methods Sixty-seven adults with DM1 and 39 healthy adults were recruited. Physical activity was monitored by accelerometry and assessed using the International Physical Activity Questionnaire. Age, marital status, education, apathy, cognition, fatigue, and muscle strength were assessed as predictors of physical activity in DM1.

Results The individuals with DM1 were on average − 187 min ($p < 0.00001$) objectively and − 48% ($p = 0.001$) subjectively less physically active per week compared to healthy controls. Education was the only predictor of physical activity in DM1 ($p = 0.02$).

Conclusions Individuals with DM1 are less physically active compared to healthy controls and only half of the patients fulfilled minimum requirements for recommended physical activity. Education is the only predictor of physical activity in DM1. Thus, enhancement of physical activity in individuals with DM1 might be suggested, and especially in the less educated individuals, but RCT studies are needed to guide exact recommendations.

Keywords Myotonic dystrophy · Physical activity · Accelerometry · IPAQ

Background

Myotonic dystrophy type 1 (DM1) is the most prevalent adult-onset muscular dystrophy [1]. It is characterized by distal muscle weakness in the extremities and multiorgan manifestations [2]. Strength and aerobic exercise of moderate intensity appear to be safe and beneficial in DM1, but the body of evidence for this is limited [3]. It is well known that physical activity has a preventive effect on lifestyle-related diseases for the general population [4, 5]. Although there is no direct evidence showing that physical inactivity causes additional physical impairment in DM1, muscle strength is decreased with bed rest in healthy individuals and physical activity seems to prevent disuse atrophy in muscle diseases [6]. Thus, it seems likely that physical inactivity worsens existing physical impairments in DM1. Subjective reporting [7–9] has demonstrated physical inactivity in DM1, but these studies may be confounded by recall bias and social desirability bias. Only two studies [10, 11] have objectively investigated physical activity in few DM1 patients, but no studies have documented physical activity objectively in a larger cohort of individuals with DM1. To promote physical activity in DM1, it is important to understand predictors of physical activity, so the correct sub-group of patients can be targeted, but such predictors are unknown for DM1.

The objective was to investigate physical activity and to investigate the predictors of physical activity in individuals with DM1 compared to healthy controls.

Methods

Participants

The DM1 participants ($n = 67$) were recruited from the MyDOM cohort from Copenhagen and Aarhus in Denmark between 2017 and 2018. Thirty-nine healthy controls

✉ Kirsten Lykke Knak
kirsten.lykke.knak@regionh.dk

[1] Department of Neurology, Rigshospitalet, Juliane Maries Vej 28, 2100 Copenhagen, Denmark

were recruited from Copenhagen via advertisement in newspapers and Facebook and via ads at the hospital for staff. The inclusion criteria for individuals with DM1 were 18–60 years of age, living in Copenhagen or Aarhus, CTG repeats > 80 (the cutoff criteria for diagnosing DM1 at the Department of Genetics, Rigshospitalet), ability to rise from a chair without using arms, and ≥ 10 m walking distance. The inclusion criteria for healthy controls were 18–60 years of age, living in Copenhagen, and in general being healthy.

The exclusion criteria for individuals with DM1 were pregnancy, alcohol or drug abuse, recent illness, onset of DM1 disease < 1 year of age, and diseases unrelated to DM1 or medicine consumption that confounds interpretation of muscle strength, balance or function. The exclusion criteria for healthy controls were pregnancy, alcohol or drug abuse, recent illness, and physical disability, illness or medication interfering with balance.

Protocol

This was an observational study with one visit. The following data were obtained: age, body mass index (BMI), marital status (cohabitant or not), education (classified as none, elementary school, high school, college with three different education lengths, and postgraduate education), apathy (Apathy Evaluation Scale self-rated [12]), cognition (STROOP [13]), fatigue (Fatigue Severity Scale-7 [14]), and isometric ankle dorsal flexor muscle strength (Newton meter) measured by Stationary Dynamometry (Biodex System 3 and 4 PRO, Biodex Medical Systems, NY).

The participants were instructed to wear a hip-worn accelerometer (wGT3X-BT, Timik Medical, Herlev, Denmark) over 7 consecutive days for 24 h, removing it only when showering or performing water activities. Acceleration intensity and duration of hip motion (e.g., steps and other movements such as rising from a chair) were monitored in three axes (up–down, left–right, forward–backward). Light, moderate, vigorous, and very vigorous intensities were extracted, which were defined as: rest ≤ 99 counts per minute (CPM); light = 100–1951 CPM; moderate = 1952–5724 CPM; vigorous = 5725–9498 CPM; and very vigorous ≥ 9499 CPM [15]. The number of counts increases with the frequency and intensity of movement. Wear time ≥ 4 days during daytime was accepted as the cutoff compliance for inclusion. Non-wear time was defined as ≥ 60 min of continuous zero acceleration [16]. On the last day of accelerometry monitoring, the subjects completed the International Physical Activity Questionnaire (IPAQ) (short version [17]) regarding walking, moderate, and very vigorous physical activity intensity within the last 7 days.

Statistical analyses

SAS Enterprise Guide 7.1 was used for data analyses. To investigate differences in physical activity between DM1 and healthy controls, a linear mixed model was applied with family as a random effect (correction for family members in the cohort) and with group, sex, age, BMI, marital status and education as covariates (correction for possible differences on these parameters between the groups). The primary confirmatory analysis was conducted on the mean physical activity (min/week) regarding accelerometry and IPAQ, respectively. The $p = 0.05$ for the two primary outcomes were Bonferroni corrected (p value$_{\text{Bonferroni}} = 0.025$). Because the different activity intensities were only exploratory outcomes, Bonferroni correction was not applied to these outcomes since this would be too conservative. To investigate predictors of physical activity, a linear mixed model was conducted with family as a random effect and with age, marital status, education, apathy, fatigue, and ankle dorsal strength as covariates (predictors) in DM1, and a multiple linear regression with age, marital status, and education as covariates in healthy controls. For linear mixed model, the following model assumptions were checked and in case of violation, the analyses were conducted on log10-transformed data: multivariate normal distribution and homogeneity of residuals, and convergence criteria. For multiple linear regression, the following model assumptions were checked: independency of observations, linearity of covariates, homogeneous and normally distributed residuals. For both groups, the number of predictors was limited to sample size divided by ten to avoid mass significance. Non-collinearity between predictors was checked. Difference in characteristics among subjects who adhered to wear time and subjects who did not was investigated by unpaired t test for continuous data, by Mann–Whitney test for ordinal data, and by Fisher's exact test for dichotomous data. The associated model assumptions were checked.

Results

Recruitment of participants is shown in Fig. 1, and demographic data are shown in Table 1. There were missing values for: (I) AES-S ($n = 11$) because AES-S was not implemented in the beginning of the study ($n = 8$), drop out before AES-S assessment ($n = 1$), not possible to attend the AES-S assessment ($n = 1$), and incomplete AES-S ($n = 1$); (II) STROOP Word ($n = 14$) because the test was not implemented ($n = 8$), withdrawal before the assessment ($n = 1$), not possible to attend the assessment

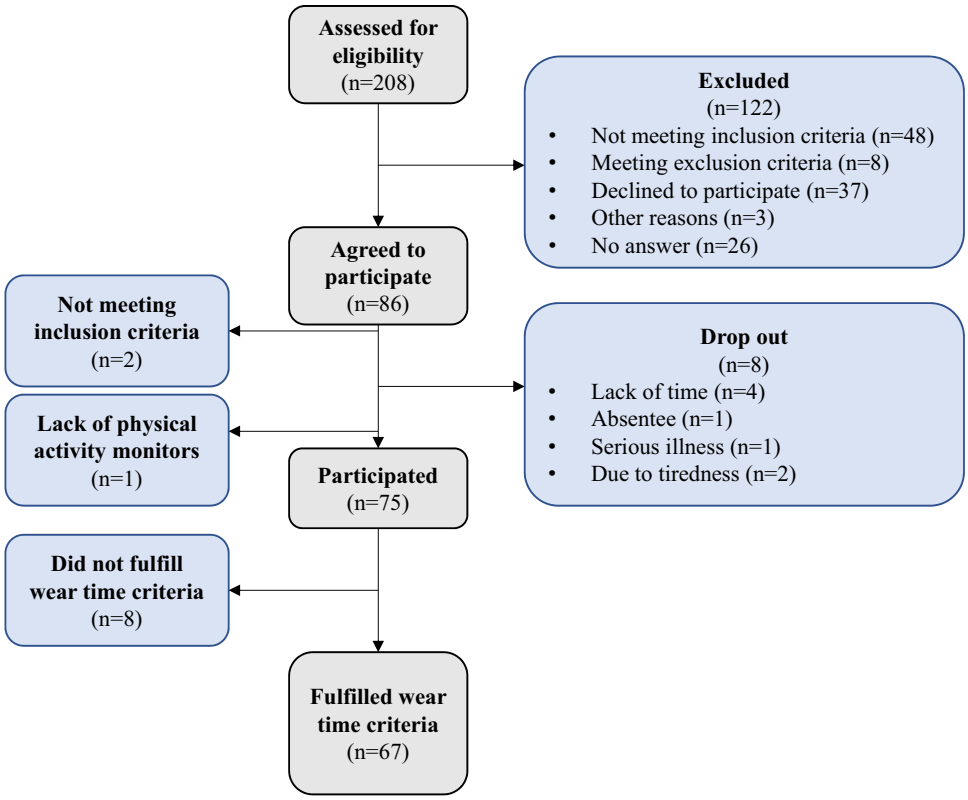


Fig. 1 Flowchart of recruitment of individuals with DM1

($n=1$), subject was unable to read ($n=1$), subject was color blind ($n=2$), and invalid score ($n=1$); (III) STROOP Color ($n=13$) because the test was not implemented ($n=8$), withdrawal before the assessment ($n=1$), not possible to attend the assessment ($n=1$), subject was color blind ($n=2$), and subject was unable to read ($n=1$); (IV) STROOP Color/Word ($n=14$) because the test was not implemented ($n=8$), withdrawal before the assessment ($n=1$), not possible to attend the assessment ($n=1$), subject was unable to read ($n=1$), subject was color blind ($n=2$), and subject did not want to complete the test ($n=1$); (V) STROOP inferences ($n=13$) because the test was not implemented ($n=8$), withdrawal before the assessment ($n=1$), not possible to attend the assessment ($n=1$), subject was unable to read ($n=1$), and subject was color blind ($n=2$) and; (VI) FSS-7 ($n=1$) because of incomplete FSS-7 ($n=1$).

The individuals with DM1 were on average less physically active objectively (mean ± SD, 485 ± 144 min) as assessed by accelerometry compared to the healthy controls (695 ± 138 min), a difference of -187 min (− 248 to − 127 min, $p<0.00001$). The individuals with DM1 were also on average less physically active subjectively (median and IQR; 380 min, 215–720 min) as assessed by the questionnaire compared to the healthy controls (550 min, 368–983 min), a mean difference of -48% (95% CI − 65 to − 23%, $p=0.001$). The different physical intensities are visualized in Fig. 2. Moreover, the individuals with DM1 fulfilled the WHO recommendations of physical activity poorly and remarkably less than the healthy controls (Fig. 3).

Education was the only significant predictor of physical activity in DM1 showing that when educational level is increased by one step (a higher degree of education), the weekly mean activity level is increased by 29 min (5–53 min, $p=0.02$) (Table 2). Age, marital status, and education were not significant predictors of physical activity in healthy individuals ($p\geq0.20$).

Twelve percent of the subjects with DM1 and 8% of the healthy controls did not adhere to wearing the activity monitor for at least 4 days. The only significant difference between the subjects who adhered and the subjects who did not was marital status in DM1 ($p=0.009$) with cohabitants being more compliant and age in healthy controls ($p=0.014$) with the oldest being more compliant.

Discussion

This study demonstrates that individuals with DM1 are less physically active than healthy individuals and that education is the only predictor of physical activity in subjects with DM1.

Our study is the first prospective investigation of physical activity assessed by objective and subjective measures

Table 1 Demographic data of the subjects with DM1

	DM1	Healthy controls
Sex, no	67	39
Women	32	21
Men	35	18
Kindred providing > 1 family member, no	11	0
Age (years), mean (SD)	41 (10)	39 (11)
BMI, mean (SD)[a]	24.3 (5.1)	23.5 (2.6)
Marital status, no		
Cohabitant	50	31
No cohabitant	17	8
Accomplished education, no		
None	1	0
Elementary school	10	1
High school	9	3
College (2–2½ years)	17	6
College (3–4½ years)	16	13
College (5–6 years)	11	15
Postgraduate education	3	1
AES-S, median (5–95 percentile)[b]	12 (3–24)	NA
Apathy, no	0	
STROOP, median (5–95 percentile)[c]		NA
Word score	32 (18–44)	
Cognitive impairment, no	20	
Color score	34 (22–55)	
Cognitive impairment, no	11	
Color–Word score	38 (32–55)	
Cognitive impairment, no	1	
Interference score	50 (49–51)	
Cognitive impairment, no	0	
FSS-7, median (5–95 percentile)[d]	4.3 (1.9–6.6)	NA
Abnormal fatigue, no	39	
Ankle dorsal flexion muscle strength (Nm)[e], median (5–95 percentile)	17 (0.00–45.70)	NA
Too weak to activate the test, no	13	

NA not applicable, not investigated

[a]BMI (body mass index) (kg/m^2), $\frac{\text{weight (kg)}}{\text{height (m)}^2}$

[b]AES-S Apathy Evaluation Scale (self-rated). A score of > 34 = apathy

[c]Verbal STROOP Color and Word test (adult version). Higher score indicates better cognitive performance. 95% CI for STROOP (30.4; 69.91)

[d]FSS-7 Fatigue Severity Scale (without item 1 and 2 from FSS-9), mean score. Higher score = greater fatigue severity. Minimum mean score = 1. Maximum mean score = 7

[e]Peak torque (Nm = Newton meter) measured by Stationary Dynamometry (Biodex System 4 Pro, Biodex Medical Systems, NY)

in a large cohort of patients with DM1. The finding of less physical activity in DM1 compared to healthy subjects is consistent with previous case series and subjective assessments of physical activity in DM1 [7, 10, 11]. Possible explanations for the decreased physical activity level in DM1 compared to healthy controls in the present study are that 42% of the subjects with DM1 showed cognitive impairment, 59% showed abnormal fatigue (FSS score ≥ 4 [18]), and 78% were affected by distal/distal and proximal muscle weakness (MIRS ≥3). In line with this, education, which partly expresses cognition, was a significant predictor of physical activity in DM1. Although fatigue was not a significant predictor, it was almost significant at 5.7% probability. The subjects with DM1 showed BMI ranging from underweight to obese but the mean BMI was normal weight. Despite a correlation between BMI and physical activity in

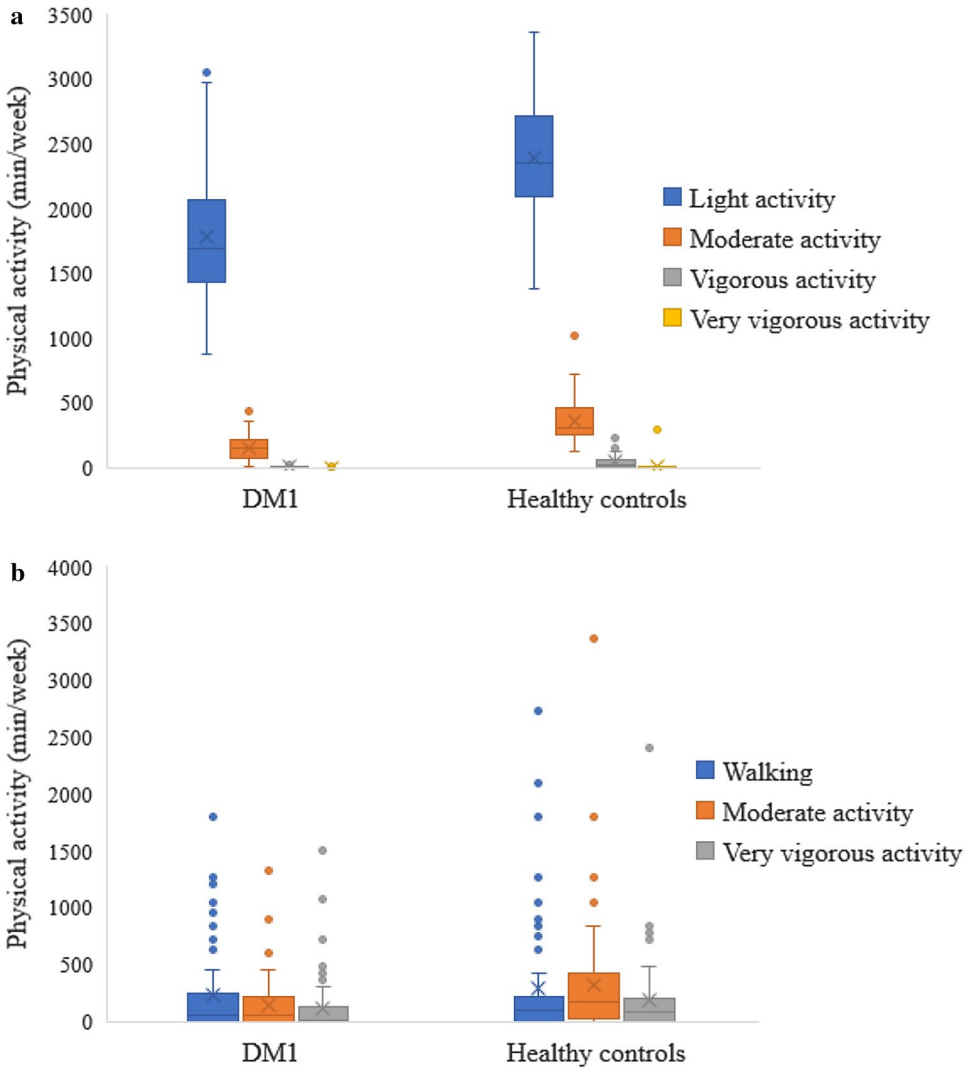


Fig. 2 Physical activity levels in individuals with DM1 and healthy controls. Part A shows objective accelerometry data and part B shows subjective IPAQ data. *$p \leq 0.05$. *Y*-axis: minutes of physical activity per week. The boxes represent different activity intensities. Box: 25%—75% percentile. Horizontal line: median/50% percentile. For the boxes with outliers, the whiskers represent 1.5 * box length, but for the boxes without outliers, the whiskers represent minimum and maximum values

healthy [19], BMI has not been shown to be a predictor of physical activity in healthy [20]. The difference in physical activity level cannot be explained by possible differences in demographic parameters between the two groups, because we adjusted for that in our analyses. This means when sex, age, BMI, marital status and education are similar in the two groups, DM1 demonstrated a lower physical activity level compared to the healthy.

WHO recommends that adults on a weekly basis perform physical activity ≥ 150 min of moderate intensity or ≥ 75 min of vigorous intensity [4]. For both the objective and subjective measurements, only around 50% of the subjects with DM1 fulfilled the recommendation of moderate intensity and 0–37% fulfilled vigorous or very vigorous intensity. This shows that persons affected by DM1 were not only less physically active than healthy subjects, but half or more of DM1 patients also failed to adhere to the WHO recommendations. The decreased physical activity level in patients with DM1 is concerning because it increases the risk of lifestyle-related diseases [4] and it also contributes to disuse muscle atrophy, which can exacerbate existing disabilities in DM1. Moreover, individuals with DM1 are predisposed to diabetes mellitus [21], which makes physical activity imperative in this condition.

The finding that education was a significant predictor of physical activity in DM1 is inconsistent with findings for healthy subjects in the present study, but consistent with previous findings in healthy adults [20, 22]. However, the previous studies [20, 22] investigated physical activity subjectively. The finding that age and marital status were not significant predictors of physical activity in DM1 is inconsistent with previous findings in healthy adults [22, 23]. However, in the present study, age ($p = 0.07$) was close to the 5% significance level. The differences between the present study and the previous studies [22, 23] may be due to different characteristics of studied populations, analyses or sample size. Unexpectedly, apathy was not a significant predictor of physical activity level in DM1 patients. However,

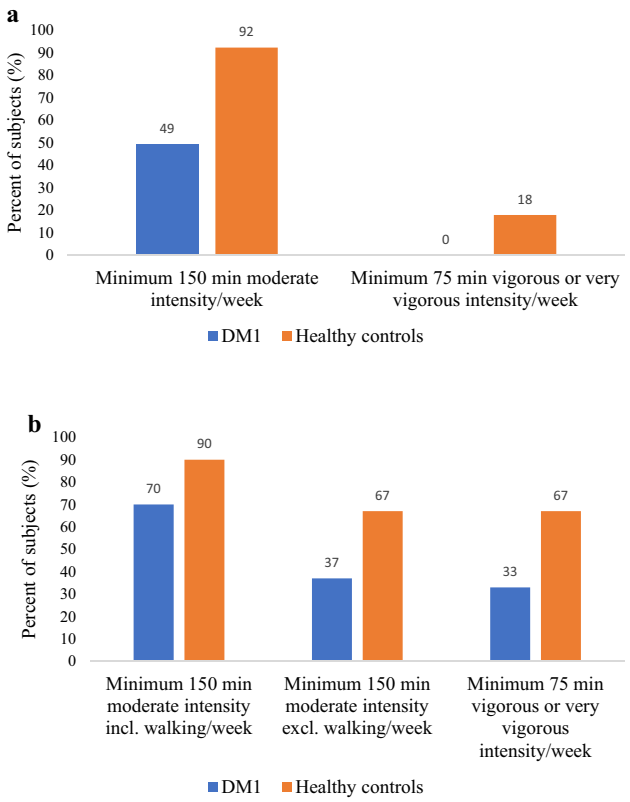


Fig. 3 Percent of subjects with DM1 and healthy controls fulfilling WHO's physical activity recommendations. Part A shows objective accelerometry data and part B shows subjective IPAQ data. There are two columns for moderate intensity, including and excluding walking, because walking can both be light and moderate intensity. *X*-axis: physical intensities according to WHO recommendations for healthy adults. Blue columns: DM1. Orange columns: Healthy controls. *Y*-axis: percent of subjects fulfilling the WHO recommendations

Table 2 Predictors of physical activity in subjects with DM1

Predictor variable	β	SE_β	$95\%CI_\beta$	*P* value
Age	− 3.05	1.68	− 6.34; 0.25	0.07
Marital status				
Cohabitant	59.03	42.39	− 24.05; 142.11	0.16
No cohabitant				
Education	28.93	12.24	4.95; 52.91	0.02*
AES-S	− 4.36	3.11	− 10.46; 1.73	0.16
FSS-7	− 24.55	12.87	− 49.78; 0.68	0.06
Ankle dorsal flexion muscle strength	− 0.47	1.40	− 3.21; 2.27	0.74

β regression coefficient, SE_β standard error of the regression coefficient, *AES-S* Apathy Evaluation Scale (self-rated), *FSS-7* Fatigue Severity Scale (without item 1 and 2 from FSS-9), Peak torque (Nm = Newton meter) measured by Stationary Dynamometry

*$p < 0.05$

apathy cannot be ruled out as a physical activity predictor because even though the apathy score varied in the present study, none of the subjects demonstrated apathy (score >34 [24]. The finding of a higher physical activity level in DM1 individuals with higher level of education, which at least partly expresses cognition, stresses the importance of targeting exercise programs for less educated or cognitively impaired individuals with DM1. In the present study, fatigue tended to predict physical activity ($p = 0.06$), which suggests that fatigue should also be managed to enhance physical activity in DM1. Exercise is one approach to reduce fatigue, because a causal relationship between increased exercise and reduced fatigue has previously been shown in DM1 [25]. Barriers to physical activity in individuals with DM1 such as physical problems and fatigue have previously been reported [7], and an RCT in DM1 [26] showed that physical activity can be enhanced by cognitive behavioral therapy. Thus, an approach to improve physical activity level in DM1 has been suggested but as the barriers indicate, promoting physical activity in individuals with DM1 is complex. Exercise studies in DM1 have shown promising results [25, 27–30], but strong evidence for this is lacking [3, 31, 32], thus an RCT is needed. There is a risk of injury and falls when being physically active, but no studies have documented detrimental effects of physical activity in DM1. In fact, the opposite may be expected as a correlation between increased lower limb muscle strength and decreased number of falls has been demonstrated in DM1 [10], which suggests that strength exercise might decrease the risk of fall.

The study was strengthened by an objective measurement of physical activity level for 7 days, thus, all days of the week with different daily activities were captured. Also, the accelerometer did not only monitor steps, but also duration and intensity of other hip motions. The DM1 sample size is considered generalizable to the background DM1 population in terms of variation in demographic characteristics, except for the most disabled individuals with DM1.

Limitations

Heart rate monitoring could register physical activity without or with minimal hip motion (e.g., bicycling, upper limb strength exercise) but because heart arrhythmia is a symptom in some individuals with DM1 affecting the heart rate monitoring, heart rate was not monitored in the subjects with DM1 and healthy controls. However, future studies could use heart rate monitoring in a subset of patients without arrhythmias. There is a risk that sedentary behavior has been misclassified as non-wear because we used Troiano et al.'s stratification [16], which classifies no acceleration as non-wear. However, sedentary behavior is often different from non-wear time, because the body often creates small

movements despite being sedentary. The subjects with DM1 and the healthy controls who did not adhere to wearing the accelerometer may be less physically active than the compliant participants. Thus, there is a risk that the physical activity level in both groups is overestimated, which stresses the main findings of the study. However, the subjects who were non-compliant constituted ≤ 12% of all subjects, and the variables that differed between compliant and non-compliant participants (marital status and age) were not predictors of physical activity in DM1. Thus, the bias from non-compliant participants is likely small. Also, there was a risk of social desirability bias with subjects increasing their physical activity level for the activity monitoring. The rather long sampling time of 7 days was used to counteract this, and likely the risk of social desirability is similar in healthy controls. We therefore do not think this potential bias was important in our study.

Conclusion

Individuals with DM1 are less physical active than healthy controls, to a degree which is likely clinically meaningful, as only half of patients fulfilled minimum requirements for recommended physical activity. Education was the only predictor of physical activity in DM1.

Acknowledgement We are grateful for the grants from Axel Muusfeldt's Foundation, the Family Hede Nielsen's foundation, and Rigshospitalet's Research Foundation.

Compliance with ethical standards

Conflicts of interest K. Knak received grants from Axel Muusfeldt's Foundation, Familien Hede Nielsen's Foundation, and Rigshospitalet's Research Foundation. J. Vissing, N. Witting, and A. Sheikh declare that they have no conflict of interest.

Ethical standards The study was approved by the Regional Committee on Health Research Ethics in Denmark (H-17017556) and written informed consent was obtained.

References

1. Hammarén E, Kjellby-Wendt G, Kowalski J, Lindberg C (2014) Factors of importance for dynamic balance impairment and frequency of falls in individuals with myotonic dystrophy type 1 a cross-sectional study including reference values of Timed Up & Go, 10m walk and step test. Neuromuscul Disord 24(3):207–215
2. Turner C, Hilton-Jones D (2014) Myotonic dystrophy: diagnosis, management and new therapies. Curr Opin Neurol 27(5):599–606
3. Voet NB, van der Kooi EL, Riphagen LL, Lindeman E, van Engelen BG, Geurts AC (2013) Strength training and aerobic exercise training for muscle disease. The Cochrane Collaboration, editor. Cochrane Database of Systematic Reviews. John Wiley and Sons, Ltd., Chichester
4. World Health Organization. Physical Activity (2018). https://www.who.int/ncds/prevention/physical-activity/gappa Accessed 17 Sep 2018
5. Sundhedsstyrelsen. Anbefalinger om fysisk aktivitet (2017). https://www.sst.dk/da/sundhed-og-livsstil/fysisk-aktivitet/anbefalinger?sc_site=website Accessed 10 Jan 2017
6. Vignos P (1983) Physical models of rehabilitation in neuromuscular disease. Muscle Nerve 6:323–338
7. Gagnon C, Chouinard M-C, Laberge L et al (2013) Prevalence of lifestyle risk factors in myotonic dystrophy type 1. Can J Neurol Sci 40(01):42–47
8. Eichinger K, Dekdebrun J, Dilek N, Pandya S, Chen D (2014) Physical activity profile and barriers to physical activity in individuals with myotonic dystrophy. Neurology 82(10):84
9. Phillips M, Marr L (2009) P3-12 Participation in physical activity by people with myotonic dystrophy. Med Genet
10. Wiles CM (2005) Falls and stumbles in myotonic dystrophy. J Neurol Neurosurg Psychiatry 77(3):393–396
11. Ayvat F, Ayvat E, Kılınç OG, Kılınç M, Yıldırım SA, Tan E (2019) Objective and subjective assessment of physical activity in adults with muscle diseases. Turk J Neurol Turk Noroloji Derg 25(3):117–122
12. Marin R, Biedrzycki R, Firinciogullari S (1991) Reliability and validity of the apathy evaluation scale. Psychiatry Res 38:143–162
13. Van der Elst W, Van Boxtel M, Van Breukelen G, Jolles J (2006) The Stroop color-word test influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment 13(1):62–79
14. Lerdal A, Kottorp A, Gay C, Aouizerat B, Portillo C, Lee K (2011) A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. Qual Life Res 20:1447–1456
15. Freedson PS, Melanson E, Sirard J (1998) Calibration of the computer science and applications, Inc accelerometer. Med Sci Sports Exerc 30(5):777–781
16. Troiano RP (2007) Large-scale applications of accelerometers: new frontiers and new questions. Med Sci Sports Exerc 39(9):1501
17. International Physical Activity Questionnaire. International Physical Activity Questionnaire. https://sites.google.com/site/theipaq/questionnaire_links. Accessed 14 Feb 2020
18. Werlauff U, Højberg A, Firla-Holme R, Steffensen BF, Vissing J (2014) Fatigue in patients with spinal muscular atrophy type II and congenital myopathies: evaluation of the fatigue severity scale. Qual Life Res Int J Qual Life Asp Treat Care Rehabil 23(5):1479–1488
19. Ofori EK, Angmorterh SK (2019) Relationship between physical activity, body mass index (BMI) and lipid profile of students in Ghana. Pan Afr Med J 33:30
20. Ransdell LB, Wells CL (1998) Physical activity in urban white, African-American, and Mexican-American women. Med Sci Sports Exerc 30(11):1608–1615
21. Kaminsky P, Poussel M, Pruna L, Deibener J, Chenuel B, Brembilla-Perrot B (2011) Organ dysfunction and muscular disability in myotonic dystrophy type 1. Medicine (Baltimore) 90(4):262–268
22. Herazo-Beltrán Y, Pinillos Y, Vidarte J, Crissien E, Suarez D, García R (2017) Predictors of perceived barriers to physical activity in the general adult population: a cross-sectional study. Braz J Phys Ther 21(1):44–50
23. Tittlbach SA, Jekauc D, Schmidt SCE, Woll A, Bös K (2017) The relationship between physical activity, fitness, physical complaints and BMI in German adults results of a longitudinal study. Eur J Sport Sci 17(8):1090–1099
24. Shirley Ryan Abilitylab. Apathy Evaluation Scale (Self, Informant, and Clinician Versions). https://www.sralab.org/rehabilita

tion-measures/apathy-evaluation-scale-includes-three-forms-self-informant-and-clinician. Accessed 14 Feb 2020
25. Jimenez-Moreno A, Okkersen K, Maas D et al. (2017) Implementing a graded exercise therapy to severely fatigued DM1 Patients: experience from the OPTIMISTIC Study 2017; International Myotonic Dystrophy Consortium Meeting
26. Okkersen K, Jimenez-Moreno C, Wenninger S et al (2018) Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial. Lancet Neurol 17(8):671–680
27. Hammarén E, Kjellby-Wendt G, Lindberg C (2015) Muscle force, balance and falls in muscular impaired individuals with myotonic dystrophy type 1: A five-year prospective cohort study. Neuromuscul Disord 25(2):141–148
28. Orngreen MC, Olsen DB, Vissing J (2005) Aerobic training in patients with myotonic dystrophy type 1. Ann Neurol 57(5):754–757
29. Missaoui B, Rakotovao E, Bendaya S et al (2010) Posture and gait abilities in patients with myotonic dystrophy (Steinert disease) Evaluation on the short-term of a rehabilitation program. Ann Phys Rehabil Med 53:387–398
30. Aldehag A, Jonsson H, Lindblad J, Kottorp A, Ansved T, Kierkegaard M (2013) Effects of hand-training in persons with myotonic dystrophy type 1 a randomised controlled cross-over pilot study. Disabil Rehabil 35(21):1798–1807
31. Gianola S, Pecoraro V, Lambiase S, Gatti R, Banfi G, Moja L (2013) Efficacy of muscle exercise in patients with muscular dystrophy: a systematic review showing a missed opportunity to improve outcomes. PLoS ONE 8(6):e65414
32. Roussel M-P, Hébert LJ, Duchesne E (2019) Intra-rater reliability and concurrent validity of quantified muscle testing for maximal knee extensors strength in men with myotonic dystrophy type 1. J Neuromuscul Dis 6(2):233–240