Exploring outcome measures for adults with Myotonic Dystrophy type 1

By

Aura Cecilia Jimenez Moreno

BSc (Physiotherapy) MRes (Neuromuscular Disorders)

Student ID 13035984

A thesis submitted as requirement for the degree of Doctor of Philosophy with the Institute of Genetic Medicine, Newcastle University;
August 2017.

Academic Supervisors:
Prof. Hanns Lochmuller
Dr. Grainne Gorman
Dr. Sarah Charman

ABSTRACT

Myotonic Dystrophy type 1 (DM1) is a multisystem progressive disorder with high heterogeneity. Novel emerging therapies require assessment tools that can effectively assess the effects of an intervention. The Outcome Measures in Myotonic Dystrophy (OMMYD) Consortium has proposed a battery of functional outcome measures (FOM) identified as relevant for clinical trials in DM1. However, due to the variable nature of the disease and a scarcity of resources, there is a lack of systematic research that properly explores the use of these FOM. The current study examined three of these FOM and one extra related to patients' daily life performance. These are: (1) the ten-meters walk test; (2) the ten-meters walk/run test; (3) the 30-seconds sit and stand test; and, (4) a tri-axial accelerometer. By exploring the reliability, validity and responsiveness of these outcomes, we aimed to establish reference values and standard methodologies that could serve as guidance for clinical trials in DM1. A cohort of DM1 adults screened for the two largest-to-date trials in DM1 (OPTIMSITIC and PHENO-DM1) were examined in relation to a set of pre-specified assessments and disease-burden scores. The results of this thesis supply disease-specific evidence of their validity, reliability and feasibility. The FOM, have shown to be psychometrically robust measures of functionality in DM1 and to be feasible for clinical trials; they can provide a picture of patients' muscle strength and perceived mobility and participation in life. The accelerometer can objectively quantify joints accelerations when walking at different speeds and summarise a DM1 patient's habitual physical activity. The final choice of an outcome measure for a clinical trial in DM1 should be guided by disease domain that an intervention is likely to impact on; but, a disease-specific study like this one will reduce the burden of protocol design whilst providing evidence supporting the decision-making process.

This thesis is dedicated to the team of postgraduates that have shaped my life: my mom - the dreamer; my dad - the healing forehead kiss; and, Simao - the one and only “funny man”.

“Never, never, never give up … and remember: dance a little.”

-Emma Watson and Gloria Steinem

ACKNOWLEDGMENTS

First of all, I want to thank my PhD funding organisations: the Medical Research Council Centre for Neuromuscular Diseases, Consejo Nacional de Ciencia y Tecnologia of Mexico and the Barbour Foundation. .

I want to express my sincere thanks to my academic supervisors who have provided me with all the necessary tools and support throughout my PhD.

-To Hanns Lochmüller *der stärkste baum, an den ich mich immer anlehenen kann*. You saw potential and believed in me to lead the OPTIMISTIC study, working closely with patients and one day becoming a doctor. Your experience in this field has been priceless, and I would never have accomplished as much as I have without this.

-To Grainne Gorman for being the heart and soul in a world where few hard-headed can sometimes make the road a bit bumpy. Your female strength is a true inspiration and a torch for women like me that also want to conquer all.

-To Sarah ‘D’ Charman for your inerasable smile, your immeasurable intelligence and your never ending support as a leader and as a friend. You also brought little Charms to my life and for that I’ll be always grateful.

I want to thank all those non-official supervisors that provided knowledge and advice to help me with my thesis.

-To Jane Newman for your systematic and serene approach when helping me whenever I struggled. Your knowledge, experience and insightful discussions have enhanced my knowledge and you are truly someone I admire.

-To Nikoletta Nikolenko not only for being someone that always saw me and treated me as colleague and not like just a student but also for turning into the nicest cypriot-ukranian friend anyone could ask for.

-To Mike Catt for your kind advice and expertise in a new and advancing which is complex and confusing at times.

-To Libby Wood for being a constant friendly ear, for introducing me to all the famous people and for that smile that always characterises you.

-To Robert McFarland for your unconditional guidance and support, and for allowing me the opportunity and exposure to your magical patients who continuously recharge my batteries every Tuesday.

-To Anna Mayhew for trusting in me even in stormy seasons, and for letting me observe you in action.
- I wish to thank Aidan Wightman for his help in checking the English of this thesis

I also want on record my gratitude for all those that have helped me throughout this journey one way or another, but most of all for supporting me when times have been difficult to achieve my goals and aspirations.

-Germán, por creer en mi capacidad hasta el final. Por cada vez que me dijiste 'te admiro' y por nueve años de crecer juntos.
-Simao y Brenda, por su envidiable manera de vivir la vida y por compartirla conmigo recordándome que todo es posible. Por todos sus dibujos.
-Mis papás, por inculcarme la ambición de nunca dejar de estudiar y de comerme al mundo.
-José, por llegar y alegrar mi vida justo cuando mas lo necesitaba.
-Zava, por tu paciencia con mis tablas mal hechas, por introducirme al arte de EndNote y mas importante por recordarme lo importante que es bailar.
-Dave and Helen for keeping my soul and body fit and happy even in the most difficult times.
-To all the staff in the clinical research facilities that have supported me throughout this PhD with a friendly and patient smile, especially to my pink girl, Ashley Bell.
-To the all the MoveLab team for their always smiley support and early welcoming into the team.

Finally, to all the patients and participants that provided me with their invaluable time and effort that made all this possible. I want to express my admiration and empathy to all the DM1 community; you are warriors behind peaceful faces.

Table of contents

LIST OF TABLES

LIST OF FIGURES

5

CHAPTER 1. BACKGROUND

1.1. INTRODUCTION

Myotonic Dystrophy Type 1 (DM1; Steinert's Disease) is a rare, autosomal-dominant neuromuscular disorder, characterized by progressive muscle weakness, myotonia and multisystem involvement. It is the second most common form of inherited muscular dystrophy and the most common amongst adults with a prevalence of 1:8,000 to 1:10,000 in the general population (1-3). In the northern region of England, patients with DM1 comprise 28.6% of the clinic population registered at the John Walton Muscular Dystrophy Research Centre(4).

Due to the nature of the disease and its heterogeneous phenotype, understanding its molecular and clinical complexity has been a challenging task. Still, potential treatments have emerged in the last decade, thus requiring the establishment of the best methods to measure disease progression and therapeutic impact. The rationale for this thesis has been to explore potential outcome measures and biomarkers suitable for implementation in DM1 clinical trials.

1.2. MYOTONIC DYSTROPHY TYPE 1

DM1 is caused by a repeat expansion (≥50 CTGn) in the 3' untranslated region of the DMPK (myotonic dystrophy protein kinase) gene located at chromosome 19q13.3 (1, 5). Although the detailed molecular understanding of this disorder is not yet fully understood, it is known that the nuclear accumulation of the resulting mutant CUG - RNA segments that remain un-translated inside the nucleus are the major factors resulting in this disorder (5). This mutation itself largely determines disease severity, progression and age of onset; and, the instability of this expansion results in a phenomenon of anticipation, increasing severity and a decreasing age of onset by generations (6, 7).

The main factor for this pathology is a miss-regulation of two RNA-binding protein families: (1) the loss-of-function of the MBNL proteins, which normally are highly expressed in the cardiac muscle, skeletal muscle and brain (8); and, (2) the overexpression of the muscle specific CUG binding protein (CUGBP1) (9). These two proteins regulate the developmental splicing process, and even though >80% of the alterations may be explained particularly by the loss-of-function of MBNL1, it seems that it is the ratio between these two that determines the pathogenic features (10). The following are relevant examples of proteins affected by this disrupted translational process in DM1: [1] the CLCN1 (muscle-specific chloride channel), a protein involved in the active contraction, the aberrant version of CLCN1 in DM1 has been identified as the source of the myotonia (11); [2] the BIN1 (bridging integrator) protein, an organizer of the T-tubule muscular network, when mutated the skeletal muscle presents with reduced strength(12); and, [3] the PKM (pyruvate kinase), an essential enzyme of the glycolysis process, when the embryonic isoform of PKM prevails in DM1 it results in high levels of muscle energy expenditure and muscle fatigue (13).

DM1 is a multi-systemic disorder with a high heterogeneity in cognitive, physical and functional levels among patients. The most affected organs involve post-mitotic tissues, such as skeletal muscle, cardiac conduction system and the central nervous system (3, 14). DM1 is typically characterized by progressive muscle wasting and weakness combined with the ‘myotonia’ phenomenon. This associated muscle weakness is a slow but persistent process that finally limits functional mobility (15).

With a highly variable clinical manifestations and age of onset, DM1 can be classified into three somewhat overlapping phenotypes but with different speeds of disease progression and severity (2, 3, 16):

a) The **congenital phenotype**, that is almost exclusively maternally inherited, is considered the most severe form and is characterized by symptoms at birth or in the postnatal period such as generalized muscle weakness and hypotonia, talipes, mental retardation, feeding problems and severe cardiorespiratory complications; the latter being the main cause of death.

b) The *classic phenotype* or **adult onset** type, presents with a relatively slow progressive muscle weakness, this progresses from distal to proximal, starting in the distal limbs (finger flexors, and wrist and ankle extensors), neck extensors and facial muscles. Patients have been historically portrayed with the typical myotonic-face characterized by bilateral ptosis and wasting of the jaw and temporal musculature. Myotonia is the recognized hallmark of the disorder when compared to other muscular dystrophies and is commonly observed in the hand after a voluntary handshake or elicited grip.
These patients also present with premature cataracts, nasal speech, cardiac conduction abnormalities, gastrointestinal tract involvement, fatigue, excessive daytime sleepiness (EDS) and cognitive impairment. Cardiac manifestations commonly include myocardial fibrosis and conduction system abnormalities in 65%-90% of patients having conduction abnormalities with a high risk for sudden cardiac death (17).

c) The **late onset** form can sometimes be considered as *mild* or asymptomatic as it is not usually diagnosed until pedigree screening or specific clinical assessments most commonly detect it. Typical, non-specific manifestations include cataracts and mild myotonia. It is associated with a normal life span.

Collectively these clinical manifestations lead to physical impairment and restricted social participation impacting considerably on the health-related quality of life indices of patients with DM1 and their families (18). Unfortunately, there is not yet a proven treatment that will relieve disease impairments, reduce limitations and optimise participation altogether. Measuring the impact of an intervention at all these levels and identifying possible cut-off points that impact on the patient's quality-of-life allows potential therapeutics to act based on relevant results.

1.3. FUNCTIONALITY AND DISABILITY IN DM1

In 2001 the World Health Organisation (WHO) established a new way for measuring health and disability providing a classification system that will standardize the language to describe an individual's health status not only

based on their own disease-related characteristics but also related to each individual's environment (19). This classification system is the International Classification of Functioning, Disability and Health (ICF) and is structured around three broad components: Body functions and structure (impairments at the body structure or functioning level); activities (experienced limitations at the individual's activity level); and participation (level of involvement as a member of society); these interact between each other based on the personal and environmental factors that affect each individual (Figure 1). One of the aims of the ICF is to reach an understanding of health and health-related outcomes in a common language that allows comparison (19).

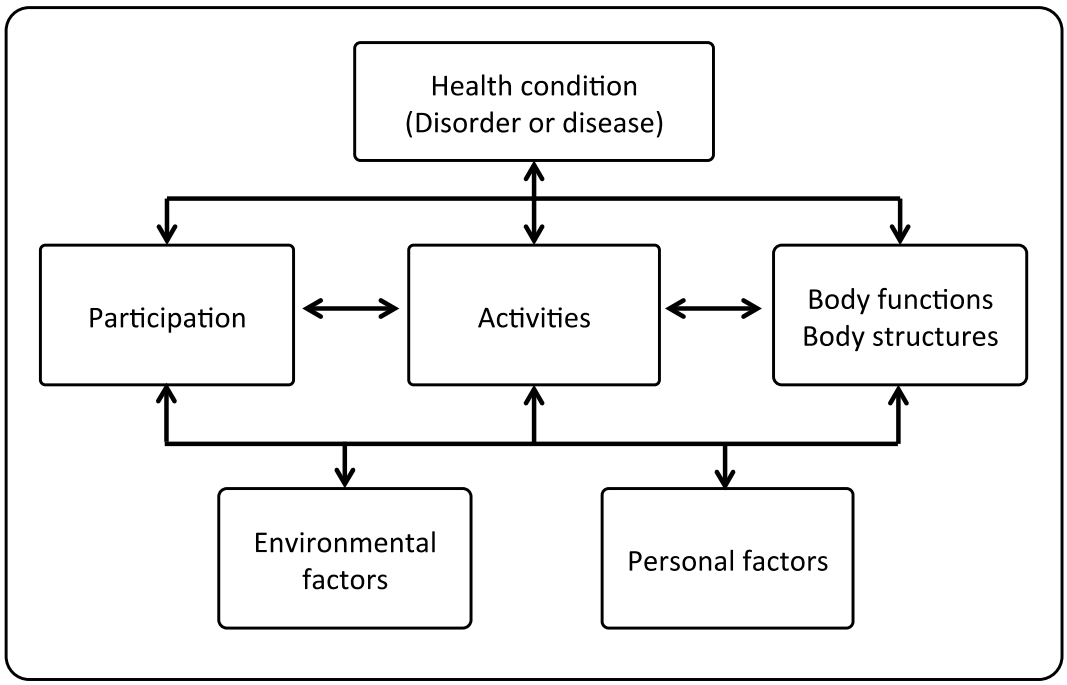


Figure 1 Model of the International Classification of Functioning, Disability and Health (ICF) - this is the World Health Organisation (WHO) framework for measuring health and disability.

Furthermore, when describing someone's activity and participation levels, these can be described as either *capacity* or *performance* (19). The latter refers to what someone accomplishes in his or her real-life and current environment and differs from the first one that assesses what someone's health status allows to be completed when requested and under ideal circumstances. This checklist-based system itself can be used as an outcome measure in rehabilitation (20); however, its practicality and specificity has been criticized before (21, 22). Still,

in research, this taxonomic system allows the identification of relevant outcome measures and the reporting of results in a common language (23, 24).

Kierkegaard et al. (25) performed a cross-sectional study aiming to describe and analyse self-related perceived functioning, disability and environmental factors related to disease severity in adults with DM1. A selection of 23 OMs with different formats was used to assess a wide spectrum of the ICF components. The number of impairments or restrictions identified ranged from one to 55 with a prevalence correlation to disease severity. Twenty per cent of participants perceived problems with 19 out of the 29 body-function categories assessed; excessive daytime sleepiness and muscle power were rated as the most common areas of burden (76-80%). More than 20% of participants perceived difficulties in 23 out of 52 activities and participation categories queried. Fifty-nine per cent to 74% reported difficulties in mobility-demanding activities. Finally, nine of the 23 environmental factors were identified as facilitators with family members and transport services as the most common (29-37%) with none were identified as barriers. This was the first attempt of classifying OMs according to the ICF checklist in DM1. These findings emphasize the multi-systemic nature of DM1 and the need for a multidisciplinary approach when caring and assessing patients with this disease. Indeed, Kierkegaard concluded her study by emphasizing the lack of standardized and validated OMs in the DM1 population and the possibility of developing a new disease-specific OM (26).

1.4. FUNCTIONAL OUTCOME MEASURES IN DM1

From 2011 a series of international workshops has been conducted with the purpose of selecting outcome measures (OM) suitable for RCTs in DM1. The Outcome Measures in Myotonic Dystrophy type- 1 (OMMYD) meetings pursue the filtered selection of condition-specific outcome measures recommended for longitudinal studies and potentially randomized controlled trials (RCTs) in DM1. These meetings started by defining a core set of disease-related domains recommended to be measured (i.e. quality of life, muscle strength, cognition, fatigue and daytime sleepiness and functional autonomy). This was followed by the identification of available evidence-based OM that could fit on any of the

core set domains; the pre-defined criteria for an ideal OM included: [a] the test must be valid, reliable and sensitive to change; [b] normative data should be available; [c] good test-retest reliability; and, [d] simple to administer (27). After that, continuous investment towards the experimentation, validation and methodology standardisation of these outcomes has been made (27, 28). This expert-lead consensus process has followed the methodology of the Outcome Measures in Rheumatology initiative (OMERACT)(29-31).

In 2011, the group in charge of the Functional Capacity Outcome Measures (FCOM) (i.e. upper and lowed extremity functions) agreed that the selected set of outcomes should measure the reflection of daily life movements and capacities and the domains proposed included: balance, walking capacity, global lower extremity function, dexterity and upper extremity speed and function (27). Due to the early stages of the DM1 research field in this domain, the FOM selection focused only on two criteria: [a] the OM must have sound metrological properties, according to the OMERACT filter (29); and [b] it must be easy to administer in research and in clinical practice among different countries (27); after this, it was agreed that further experimentation should follow. From an initial set of suitable FCOM (Table **1**), five were considered as most appropriate for DM1 based on published evidence and team members' experience with the disease and the different outcomes discussed. The final set of proposed FCOM included: (1) Six-Minute Walk Test (6MWT), (2) timed 10-Meter Walk Test (10m-WT), (3) timed 10-Meter Walk/Run Test (10m-W/RT), (4) 30-Second Sit Stand Test (30SSST), and (5) Nine-hole Peg Test (9HPT)(28). On the last OMMYD meeting in 2015, the FCOM team suggested standardizing operational procedures (SOPs) for a set of identified outcomes relevant when assessing physical capacity in DM1 clinical trials. Details of the FOM selection process are presented in chapter 3.

1.5. CURRENT CLINICAL TRIALS IN DM1

By 2017, ClinicalTrials.gov shows 77 registered studies, 59 assessing a type of intervention with 15 studies actively recruiting. Functional outcome measures (not related to the PHENO-DM1 trial) identified were: 10-m walk and walk/run test (**ClinicalTrials.gov Identifier:** NCT02858908 and **ClinicalTrials.gov**

identifier: NCT00577577), 6MWT (**ClinicalTrials.gov Identifier:** NCT02118779) and TUG (**ClinicalTrials.gov Identifier:** NCT02312011 and **ClinicalTrials.gov Identifier:** NCT02251457).

ClinicalTrials.gov Identifier: NCT02858908 is a single-bling, phase II study to evaluate the safety and efficacy of Tideglusib in adolescent and adult subjects with congenital or juvenile-onset myotonic dystrophy type 1. Tideglusib is a selective and irreversible **GSK-3β** inhibitor with therapeutic potential already identified in other pathologies (32, 33). GSK-3β or glycogen synthase kinase 3 is a protein responsible for the phosphorylation and regulation of different factors of the transcription process and is involved in the regulation of the cell cycle, apoptosis, and survival (34). In DM1, there is an increase in stability and activity of the GSK-3β (35). Preliminary findings indicate that Tideglusib may act on the central nervous system, however due to the molecular level of action and preclinical findings, it might also be beneficial for muscle function in patients with DM1 (35). The 10-mWT and the 10-mW/RT, if adequate, should identify these changes, if any.

ClinicalTrials.gov identifier: NCT00577577 is a phase II RCT study to evaluate the safety and efficacy of a complex of Recombinant Human Insulin-Like Growth Factor In Myotonic Dystrophy Type 1. The insulin-like growth factor has shown protein synthesis and differentiation of DM1 muscle cells in culture (36). Ambulation and muscle function and strength are part of this study primary outcomes, the first one measured with the 10-mWT.

ClinicalTrials.gov Identifier: NCT02118779 is the OPTIMISTIC trial which will be introduced later on in this chapter.

ClinicalTrials.gov Identifier: NCT02312011 is a phase 1/2a blinded study testing safety, tolerability, and pharmacokinetics of multiple escalating doses of IONIS-DMPK-2.5Rx administered subcutaneously to adult patients with DM1. This is antisense-based compound targeting the toxic RNA translated from the mutated DMPK gene in the muscle. From preliminary findings there is small but encouraging trend in biomarker and splicing changes, therefore Ionis has reported a setback on the program until improved compound is ready (11, 37).

ClinicalTrials.gov Identifier: NCT02251457 is a phase II study to determine if ranolazine is a safe and effective treatment for the symptoms of myotonia congenital, myotonia congenita, and myotonic dystrophy type 1. As primary outcomes this study included: [1] patient reported outcomes related to quality of life; [2] electromyography to measure myotonia; and, [3] the time up and go (TUG) test as muscle task. Preliminary findings identified significant changes after four weeks of treatment with a TUG time reduction (p = 0.03) (38).

Table 1. Functional capacity outcome measures (FCOM) identified as suitable to introduce in DM1 clinical trials in 2011 by the OMMYD consortium.

FCOM	Characteristics	Published Evidence with DM1
6-minutes walking test (6MWT)	Total distance walked over 6 minutes as sub-maximal walking speed. Evaluates walking capacity and aerobic capacity.	The difference for a single subject should be greater than 33 m or 6% of total distance to be considered a real clinical change (39). Thirty-seven children with congenital DM performed the first 2 minutes of the 6MWT and showed correlation with leg lean mass muscle (*r* 0.62) (40).
10-meters walk test (10-mWT)	Assesses walking speed at a selected comfortable pace.	In older adults most small meaningful changes are from 0.04 to 0.06 m/s on gait speed. Even though selected walking speed in the clinical environment might still be slower than selected speed when tested in the daily life environment this test differentiates disease severity (41).
10-meters walk/run test (10-mW/RT)	Assesses walking speed at fastest possible pace.	This test has been studied as a possible fall predictor in DM1 (42-44).
Timed Up & Go (TUG)	Time it takes a patient to rise from a seated position in an armchair, walk 3 metres at	After five years of follow-up DM1 patients showed a statistically significant ($p<0.001$) deterioration (from 9.6 secs at baseline to 12 secs at year 5) (44, 45). This test correlates to falls risk in DM1.

	selected speed, turn back and sit down.	In the elderly, it differentiates between fallers and non-fallers in older people (cut off >14 seconds) (46).
Step Test	As many “full steps” as possible over an 8 cm high block.	After 5 years of follow up, a mean reduction of 19% was observed in DM1 patients (45). It has been correlated with the number of falls a patient may experience in one year
30 seconds sit and stand (30SSS)	Measures functional lower limb strength and dynamic balance.	Reliability and validity studies not yet published for DM1. A similar test, timed-stand test (TST) has been suggested to separate normal from abnormal performance on adult muscular conditions, but has the risk of excluding participants on the more severe spectrum of the phenotype that could not perform the test (i.e. Muscular Impairment Rating Scale of 4 or 5) (25, 47).
Nine-hole Peg test (9HPT)	Assesses upper extremity function, specifically fine dexterity.	Good to very good intra- and inter-reliability were reported in adults with DM1 (48). Discriminates between participants with distal weakness or non weakness at all and participants with weakness present in proximal muscles (i.e. Muscular Impairment Rating Scale stages 1-3 and 4-5) (47).

1.6. ACCELEROMETERS: POTENTIAL OUTCOME MEASURES OF PERFORMANCE IN DM1

All the previous mentioned FCOMs are examples of *capacity* assessments. There are, however, other options to assess directly and objectively: matters of *performance*. By accurately measuring physical activity in a free-living environment, someone's functional performance in daily life can be assessed (19, 49). The ICF considers the individual's PA levels and *participation* alongside the individual's environment as factors to consider when classifying someone's health status.

Current technology allows for measuring activity behaviours with good accuracy and detail and an example of this is the use of accelerometers (50). Accelerometer outputs come from detecting and recording body acceleration and deceleration and are usually recorded as units of acceleration over time (counts) that can then be further transformed into more meaningful outputs such as energy expenditure or step counts. Despite advances in the use and development of accelerometers, careful consideration is still needed when employing them to capture clinically meaningful outcome measures (51). When thinking about activity monitors (in this case accelerometers) for cohorts accompanied by functional limitations such as DM1, certain factors should be considered to support and validate the application of a device in the population of interest such as altered biomechanics and a slower gait speed when walking (52, 53). Increasingly, this technology has been implemented into research and many of these tools have been correlated to long-term health outcomes and motor capacity in diseases with impaired mobility such as Parkinson's disease (54), stroke (55) and cerebral palsy (56). Their use and interpretation in DM1 and other neuromuscular disorders are still in their infancy (51, 57).

1.7. CHARACTERISTICS OF A GOOD OUTCOME MEASURE

Any outcome measure gives the opportunity to assign a number to an observation and to quantify a phenomenon (58). When choosing measures or

tools as relevant OMs for any population in particular, in this case DM1, it is important to consider the assessment properties: validity, reliability and responsiveness, plus the feasibility of its implementation (59).

Reliability: defines the accuracy level of a measurement. A significant component of the process of development or validation of an outcome measure is to reduce the error of measurement as much as possible. A good outcome measure will be reliable not only within the test in particular but at different times or when performed by different assessors also (58, 59). Internal reliability can be measured by the Cronbach's Alpha obtained from the Intraclass Correlation Coefficient (ICC) measurement; and, by comparing two different scores either from test to test or from one tool to another external reliability can be estimated (60-62).

Validity: reflects the extent to which an instrument measures what it is supposed to measure. By correlating an outcome measure to other validated tools or scores in the same area we can validate the level of comparability within measures (63, 64).

Sensitivity: is the ability of a test to correctly classify an individual as 'affected' (or characteristic present) (65).

Specificity: is the ability of a test to correctly classify an individual as characteristic-free (65).
The specificity and sensitivity probabilities of a test or its accuracy to discriminate different disease cases from others (or from clinically unaffected) may be evaluated using Receiver Operating Characteristic (ROC) curve analysis (65-67).

Responsiveness: is the ability of a measure to detect any real change over a prespecified time frame. With progressive diseases, a longitudinal analysis can test an outcome measure's sensitivity to change and can help to establish the minimum change to be considered as a clinically meaningful change over that specific period of time (68).

There might already be other outcome measures that quantify the same or very similar characteristics to the one expected, but, if there is not enough evidence of the reliability and validity of these in the specific targeted population then this measure should be initially tested and validated in the population of interest before implementing it in a clinical trial (58, 59).

CHAPTER 2. STUDY OUTLINE

2.1. GENERAL AIM

The aim of this research is to explore outcome measures and assess functional capacity and performance in adults with myotonic dystrophy type 1 (DM1), searching for tools and data that maybe suitable for use in clinical trials.

2.2. SPECIFIC AIMS

1. To explore the feasibility, validity and reliability of the OMMYD selected functional outcome measures: 30 seconds sit and stand, 10 meters walk test and 10 meters walk/run test; in adults with DM1.
2. To explore the validity and feasibility of activity monitors (accelerometers) as outcome measures of daily life performance (habitual physical activity) in DM1 adults.
3. Describe assessment protocols for the selected outcomes that can be replicated in clinical practice and/or clinical trials.
4. Describe a source of reference values that can help the design of future clinical trials selecting any of the above as an outcome measure of their study.

2.3. THESIS HYPOTHESIS

- The functional capacity of DM1 adults can be assessed effectively with the following functional outcome measures: 30 seconds sit and stand, 10 meters walk test and 10 meters walk/run test.
- With the use of ankle-worn accelerometers, we can objectively measure habitual physical activity of patients with DM1.

2.4. STUDY DESIGN

The results presented in this thesis are derived in their majority from two on-going trials in DM1: [1] OPTIMISTIC (**ClinicalTrials.gov Identifier:** NCT02118779), and [2] PHENO-DM1 (**ClinicalTrials.gov Identifier:** NCT02831504). Both studies target genetically confirmed DM1 adults with the former study an intervention based randomised control study; and the second one, a natural history observational study. Additionally, two other cohorts were included in study 1 as comparisons groups, one formed by healthy volunteers and another formed by chronic fatigued patients (Figure 2. Thesis outline).

The first study (chapter 3) focuses on the OMMYD selected functional outcome measures (FOM) and the second study (chapter 4) investigates the use of an accelerometry-based device (GeneActiv accelerometers) in DM1 adults. Both studies start with an initial description of the sample as a cross-sectional single-visit study followed by a smaller -sample longitudinal analysis.

The statistical methodology followed by both studies is very similar including both a set of the following: [1] descriptive statics and normality testing (Shapiro-Wilk test); [2] comparison between groups and sub-groups (independent sample t-test and Mann-Whitney test); [3] reliability and validity testing with the use of intraclass correlation tests, binary correlations and Bland-Altman plotting; and, [4] progression over time analysis (paired T-test) and standard error of measurement (SEM) estimations to identify a minimum expected change when declaring a real change. Additionally, a Receiver Operating Characteristic (ROC) analysis was performed for the FOM chapter.

Specific details of each study are presented in their respective chapters.

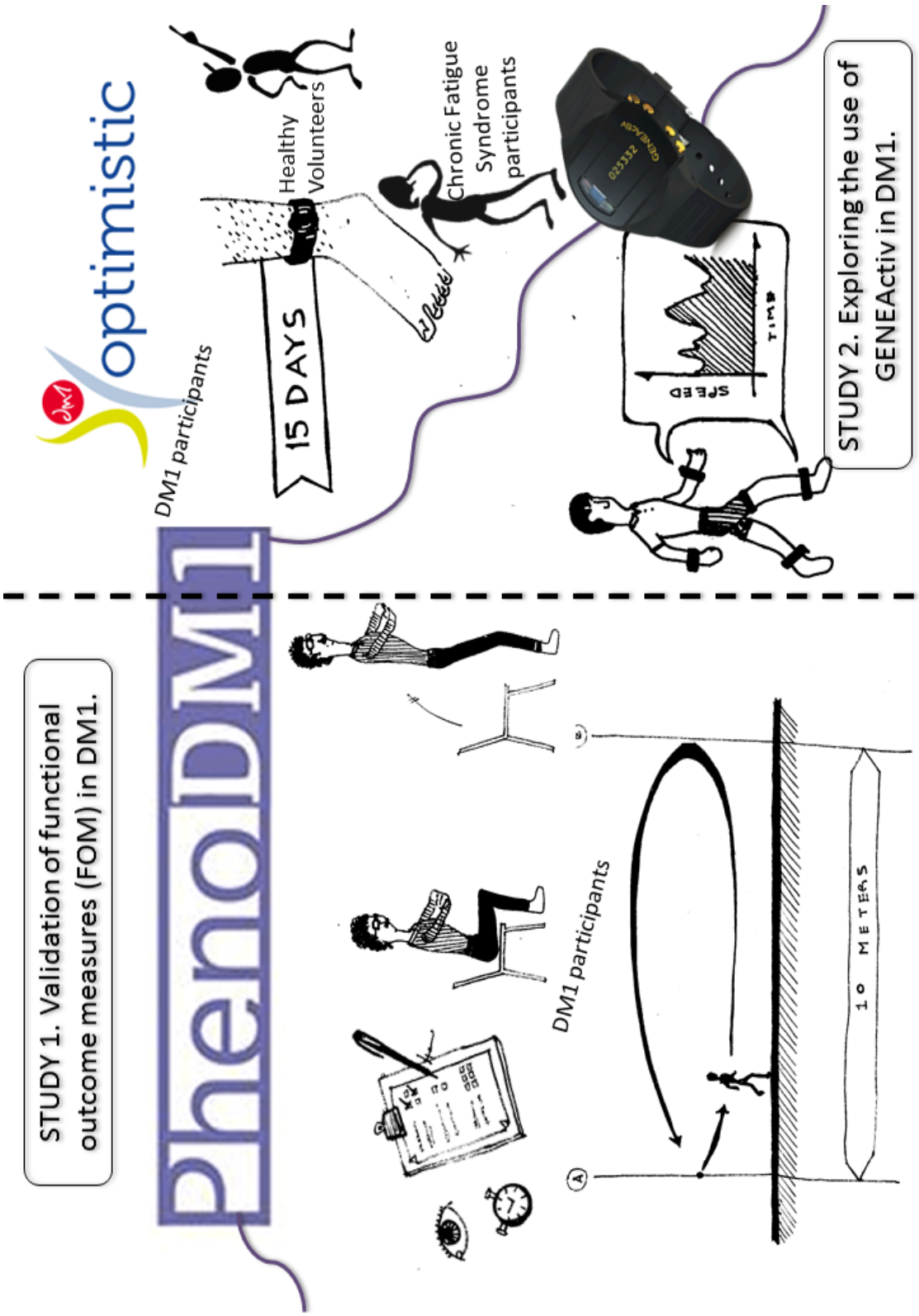


Figure 2. Thesis outline.

5

2.4.1. *OPTIMISTIC trial (**ClinicalTrials.gov Identifier:** NCT02118779)*

Observational Prolonged Trial in Myotonic Dystrophy type 1 to Improve Quality of Life Standards, Target Identification Collaboration (OPTIMISITC). OPTIMISTIC is a two-arm, multi-centre, randomized controlled trial with the main aim of improving clinical practice in the management of patients with DM1. It has been designed to compare standard management regimes against an active group. The active component is based on a cognitive and behavioral change therapy (CBT) developed particularly to increase physical activity as it includes a special component of graded physical activity. OPTIMISTIC is the first international clinical trial in myotonic dystrophy type 1 as collaboration between: the Netherlands, Germany, Paris and Newcastle, with one recruitment site in each (69).

The rationale behind this study intervention comes from the importance and prevalence of severe fatigue (>70%) in DM1 (70). Severe fatigue is a perpetuating factor that impacts on people's social participation and quality of life. After a DM1 longitudinal study, a fatigue model was created; this showed associations between reported fatigue and lack of physical activity, sleep disturbances, pain and the disease-associated lack of motivation of these patients. It was concluded that by alleviating at least one of these influencing factors, experienced fatigue or the way the patient copes with it could be improved and by consequence their general health status and quality of life (71, 72).

This study aim will be assessed by the impact on the DM1-ActivC patient reported outcome (73) (as primary outcome) and with a wide range of secondary outcomes that include: fatigue reported outcomes (Checklist Individual Strength (CIS) fatigue score); 6-minute walk test (6MWT); and, Habitual Physical Activity (HPA) levels measured for 15 consecutive days after each visit by using an ankle worn tri-axial accelerometer (GENEActiv). All outcome measures (or as many as possible), together with any adverse events

will be measured at screening/baseline visit (i.e. combined visit 1 and 2), visit 3 (5 months), visit 4 (10 ± 1 months) and visit 5 (16 ± months)(69).

The inclusion criteria include adult patients (≥18 years), who are severely fatigued (as measured with a CIS score ≥35), with the ability to walk independently (orthotics and walking assistive devices allowed) and capable to provide informed consent.

2.4.2. *PHENO-DM1 trial (**ClinicalTrials.gov Identifier:** NCT02831504)*

PHENODM1- Myotonic Dystrophy type 1 (DM1) includes deep phenotyping to improve delivery of personalized medicine and assist in the planning, design and recruitment of clinical trials. PHENO-DM1 is a multicentre (Newcastle and London), natural history study in the UK that has completed recruitment and baseline assessments. PhenoDM1 will use patient reported outcomes to assess levels of pain, fatigue, quality of life and disease burden in this cohort. Clinical and functional outcomes will look at muscle wasting and levels of myotonia and blood samples (RNA, DNA, HbA1c, thyroid hormones and androgens in males) will be collected from all patients so that additional genetic and molecular biomarker analysis can be performed.

Inclusion criteria limited to those over 18 years of age, with a genetic confirmation of DM1 who are able to provide informed consent and walk for at least 10 meters independently (orthotics and walking assistive devices allowed). One of the aims of this study is to identify population subgroups and understand independently the nature of disease-progression in different phenotypes. This unrestrictive approach will enable the assessment of a wide and comprehensive spectrum of the population, including those with early, adult and late onset phenotypes.

This study aimed to recruit 200 to 400 patients with a 1:1 men and women ratio. It involves two to three study visits approximately 12 months apart.
Strength and function assessments include: [1] Manual muscle testing and Quantitative Muscle Testing (Hand Held Myometry and Hand-Grip Dynamometry; [2] Muscular impairment rating scale (MIRS)(74); and [3]

Functional outcome measures (FOM) in parallel with OMMYD agreement (Nine Hole Peg Test, Six Minute Walk Test, 30 Seconds Sit and Stand Test, Timed 10-Meter Walk Test and timed 10-Meter Walk/Run Test).

2.5. SAMPLE RECRUITMENT AND CHARACTERISTICS

The two main sources of patient recruitment for these trials were the UK Myotonic Dystrophy Patient Registry (75) and directly at clinics when attending their standard annual specialist appointment. The Myotonic Dystrophy Patient Registry in the UK is a nationwide, self-completed system where DM patients can register and provide basic information of their condition and consent to be approached in case suitable clinical trials. It also includes a section filled in by the health-professional involved with the participant with a brief understanding of their disease-status allowing identification of potential participants based on characteristics matching trial selection criteria. However, strategies aimed at raising awareness through patient organizations or at platforms at neuromuscular diseases conferences were also implemented. Implementation of these strategies has provided me with an unprecedented number of DM1 participants for my studies.

2.6. ETHICAL APPROVAL

The collection of data from each cohort is under the ethical approval of the corresponding study protocol.

OPTIMISTIC (IRAS project 137613) has been funded by the EU Seventh Framework Programme (#305697) and has been ethically approved by the NRES committee North East – Tyne & Wear South (REC: 13/NE/0342) and developed in collaboration with The Newcastle upon Tyne Hospitals NHS Foundation Trust in the UK. Dr. Grainne Gorman is the principal investigator at Newcastle site.

PHENODM1 (IRAS project 180510) has been funded by the National Institute of Health Research (NIHR) Rare Disease Translational Research Collaboration

and The Wyck Foundation and has been ethically approved by the NRES committee North East – Tyne & Wear South (REC: 15/NE/0178) and developed in collaboration with The Newcastle upon Tyne Hospitals NHS Foundation Trust under the lead of Professor Hanns Lochmuller and University College London Hospitals NHS Foundation Trust under the lead of Dr. Chris Turner.

CHAPTER 3. EXPLORING FUNCTIONAL OUTCOMES IN DM1

3.1. INTRODUCTION

Myotonic Dystrophy type 1 (DM1) is a multisystem disease with high heterogeneity, which represents an obstacle when defining outcome measures that can be valid for different phenotypes yet still be sensitive to address change over time. Novel emerging therapies for DM1 require a deep understanding of the natural progression of the disease through multifactorial assessment tools that can be applied to disease cohorts. The OMMYD Consortium has proposed a set of Functional Outcome Measures (FOM) that were highlighted for consideration in clinical trials for DM1 (27, 28). A cohort of 213 patients was enrolled in the natural history study of PHENO-DM1 (Myotonic Dystrophy Type 1 Deep Phenotyping to Improve Delivery of Personalized Medicine and Assist in the Planning, Design and Recruitment of Clinical Trials) with the aim to assess the validity, reliability and possible sensitivity to change of three of the five OMMYD functional outcome measures (FOM). The protocol includes: [1] Standard medical history; [2] Strength assessments (myometry and manual muscle testing); [3] Functional outcome measures (six-minute walk test, 30 seconds sit and stand test, timed 10 m walk test, timed 10 m walk/run test, 9-hole peg test); and [4] patient-reported outcomes including the gold standard tools for DM1 (DM1-Activ-c and MDHI). By comparing selected functional outcome measures to clinical manifestations of the disease and to the reported burden of illness, we expect to establish the feasibility and validity of these tests in large-scale studies. By analysing the variability from each test retest we expect to assess their reliability and by stratifying our population according to sex or clinical phenotype, we expect to establish reference values for their use in clinical trials in DM1.

SELECTING FUNCTIONAL OUTCOME MEASURES AS PART OF THE OMMYD CONSORTIUM

As introduced earlier, the international OMMYD consortium was initiated with the aim of selecting the best available outcome measures to be used in

research and clinical trials in DM1. The team members of the Functional Capacity Outcome Measures (FCOM) have reached consensus of four tests considered relevant and robust enough to assess physical function and capacity in DM1. These tests are: [1] six-minute walk test (6MWT), [2] 30 seconds sit and stand test (30SSS), [3] timed 10 m walk test (10-mWT), [4] timed 10 m walk/run test (10-mW/RT), and, [4] 9-hole peg test (9HPT). The selection process resulting in these FCOM started in 2011 and for the last 18 months prior to this thesis writing, their standard operating procedures were developed.

The FCOM selection process followed closely the OMERACT initiative, a data driven interactive process in which a group of relevant stakeholders from different fields of interest in DM1 participated to endorse valid, responsive and feasible health outcome measures/scales relevant to disease specific health-domains(31). On the first meeting, an agreement on a minimum set of outcomes worthy to investigate further was reached. Presentations and discussions focused on potential outcome measures which were identified through an initial pre-meeting questionnaire. For this first meeting, five attendees conformed the FCOM group (initially named upper and lower extremity functions). The first step consisted of reviewing existing tools that could assess the established disease-domains related to upper and lower limb functionality. This was accomplished through a systematic literature review selecting tools previously used in DM1 or other diseases with similar characteristics. It was agreed that at that point experts were unable to create a DM1-specific FCOM battery of tests due to the lack of validity assessments specifically reported for DM1. However, they produced a list of 24 potential FCOM (including balance tests) emphasising the strengths and weaknesses raised during the discussion (27).

Two years a second meeting was convened with a team of six experts conforming the FCOM group. The aim of this second meeting was to reach consensus on the creation of a universally feasible battery of tests. This will be accomplished by refining the previously selected outcomes based on three component criteria: truth (validity), discrimination (sensitivity and specificity) and feasibility. After that, a minimum of three tests would be expected as consistently used in future clinical trials in DM1. A series of discussions around

the outcomes identified at OMMYD-1 resulted in a classification system for the tests defining them as: “must”, “highly recommended” or “to be evaluated”, based on their relevance when assessing an intervention in DM1 that may impact on patients’ functionality. Four outcomes were identified as “must” and are presented in this chapter. Two main criteria led to their classification including previous results reported in DM1 and published or presented opportunity to discriminate disease severity (i.e. discriminate MIRS stages 1-3 and 4-5). Evidence of test-retest reliability (regardless of population tested) was essential. Finally, we focussed on identifying outcomes that represented a functional activity as close as possible to a daily life activity (28). Outcome measures such as climbing stairs, timed up and go (TUG) and the Berg balance scale were labelled as “highly recommended” but not considered a “must” due to their risk for either flooring or ceiling effect (76-78). The 30 second sit to stand (30SSS) however, which is a close assessment to the TUG but avoids any floor effects was also deemed appropriate as could potentially assess fatigue (79).

On the third and final meeting, there was an initial confirmation of the consensus around the four FCOM previously selected and a critical review of any new evidence supporting or rejecting the decision. An initial draft of the procedures to follow when implementing these tools was developed taking into consideration specific characteristics of this disease and sites’ feasibility. The refinement of these SOPs have been an on-going process and the final version of the outcome of these endeavours will be submitted to a peer-reviewed journal at the end of this year.

3.2. METHODS

3.2.1. Sample

This study sample represents 213 patients screened for the ongoing observational natural history PHENO-DM1 study. The inclusion criteria for this study included a cohort of genetically confirmed DM1 participants 18 years old or older, with the ability to provide individual informed consent and walk independently (assistive devices and orthotics allowed) for at least 10 metres. A

cohort of 34 of these patients had completed a follow-up visit (12 months after baseline) at the time of this analysis and were included for progression-over-time analysis.

3.2.2. Procedures

This research is covered under the ethical approval of the PHENO-DM1 study by The Newcastle and North Tyneside Ethics committee (Reference: NE/15/0178).

This study focuses on the validity and reliability of the following functional outcome measures (FOM): [1] thirty seconds sit and stand (30SSS); [2] ten meters walk test (10-mWT); and [3] ten meters walk/run test (10-mW/RT). The following outcomes were considered for comparisons: [1] muscle strength and capacity (including: quantitative muscle testing (QMT) of hand-grip strength, knee extensors; hip flexors and ankle dorsiflexors, plus the Muscular Impairment Rating Scale (MIRS) which is a method of assessing disease progression as measured by muscle weakness manifestations (74); [2] additional performance tests, including the six-minute walk test (6MWT) and severity of ataxia rating scale (SARA) which assess balance and movement co-ordination and have been reported as possible assessments of disease severity in DM1 (39, 42); [3] disease-specific patient-reported outcomes (PROM) which include the DM1-Activ-c Rasch built scale and the Myotonic Dystrophy Health Index (MDHI) questionnaire (73, 80, 81).

The standard operating procedures (SOPs) for each test were discussed prior to submitting the protocol and those tests selected from the OMMYD pack attempted to follow the SOPs discussed at the OMMYD's last meeting (Paris 2015). The final consensus of these FOM were not finalized nor published at the time that this protocol was submitted (27, 28), hence, minor variances with these final SOPs can be identified.

The following section describes the assessment methods:

30 second chair sit and stand test (TSST): this test measures the functionality of lower limb, core muscles' strength and dynamic balance (82).

Equipment used: [1] a chair 45 cm high with no armrests collocated straight back towards a firm wall and [2] a stopwatch.

Procedure: the test starts with the patient sitting down; back straight and arms crossed against his/her chest, and is instructed to keep the arms in that position throughout the test. The patient is instructed "This test is called the 30 seconds sit and stand test, therefore you'll have to do as many sit and stands as possible within 30 seconds. You'll start when I say 'go' and you'll do as many repetitions as possible until I say 'stop'. It will have to be a full-stand (i.e. extended knees) for me to count it. Any questions? Are you ready to start?" The number of times the participant reaches the full standing position are counted. If the patient could not perform the test and needed arms support to complete the stand, "0" stands were recorded.
Three full trials were attempted. Before starting each new trial the participant was asked "Are you feeling OK?" or, "How do you feel for another try?" and if the answer was "no", the test would stop.

10 meters walk test (10-mWT): this test measures the short duration of a comfortable walking speed (i.e. self-selected pace) (83). This test has been validated among a wider range of conditions and promises generalizable conclusions (84).

Equipment used: [1] stopwatch and [2] a walkway of 12 metres of walking course, leaving one metre at the beginning for acceleration before the time-start point and at least 1 metre at the end after the time-end point for deceleration.

Procedure: the test starts with the patient standing still at the first mark and receives the following instructions: "This is the 10 meters walking test and you are going to walk at a comfortable speed from this mark (or cone) to the one at the end of the line (indicating the last mark). Remember it is a comfortable speed, it means the normal speed you choose in your daily life activities". The time recorded is from the moment the first foot of the patient crosses the second

mark to the moment the first foot crosses the third mark (10 metres). This means that is a timed 10-meter walk test at comfortable speed with flying start and flying finish.

There is a standing rest (10-20 sec approx.) before asking the patient to repeat the same procedure aiming for 3 trials if possible.

If the first two trials showed consistent results it was up to the assessor's judgment to decide to go for a third trial or not.

10 meters walk/run test (10-mW/RT): this test measures the participant's ability to run and the maximum speed pace (83).

Equipment, corridor length and measurement methodology are as the 10-mWT but with the following instructions: "This is the 10 meters walk and run test and the aim is for you to go from this mark (first) to that last one as fast as you can safely go. If you feel you can run you can. You'll have more than one trial allowing you to test the surface on the first attempt. Remember it is safe".

Muscular Impairment Rating Scale (MIRS): is an ordinal five-point rating scale established according to the clinically recognized muscle strength loss progression from distal to proximal (74). This test has been proposed to monitor disease progression and has been used as disease-severity classification system in other trials exploring outcome measures in DM1 (39, 44).

The scale grades are as followed:

1. No muscular impairment
2. Minimal signs (myotonia, jaw and temporal wasting, facial weakness, neck flexor weakness, ptosis, nasal speech, no distal weakness except isolated digit flexor weakness)
3. Distal weakness (no proximal weakness except isolated elbow extensor weakness)
4. Mild to moderate proximal weakness
5. Severe proximal weakness (i.e. a proximal muscle with an MRC score ≤3/5; MRC: Modified Medical Research Council Scale)

6-minute walk test (6MWT): based on the proposed American Thoracic Society guidelines (85, 86), the 6MWT was performed with the following variations: [1]

corridor length of 25 metres (87); and [2] one examiner at each end of the corridor, and when considered needed (i.e. risk of falls or an unstable patient) one examiner walking behind the patient. Encouragement inputs were provided every minute.

Severity of Ataxia Rating Scale (SARA): this test regardless been originally created as an ataxia rating scale (88), it has been identified as reliable and valid when measuring disease severity in DM1 with strong correlations to disease biomarkers (CTG length), strength and functional outcomes and to disease-specific reported outcomes (42). The SARA test includes eight performance-based items, each with different scoring ranges, but to all of them it applies that zero implies no dysfunction and higher scores imply a degree of impairment, with a total score between 0 and 40 (88).

Exploring the utility of SARA as scale to measure disease severity has been a parallel project of this study and the first exploratory results have already been published (appendix A) (42).

Quantitative Muscle Strength (QMT): the MicroFet-2 (MicroFet, Draper, UT) was used to assess the isometric strength of the following muscle groups: ankle dorsiflexors, knee extensors and hip flexors. All these measure in pounds (lbs). A Saehan DHD-2 Digital dynamometer was utilized to measure grip strength, measured in kg. Initial training for each examiner and standard written instructions were utilized throughout the assessments to increase intra-rater reliability. Each test was performed at least three times, searching for a variance of no more than 10% between tests but if the participant's score continued increasing or one of the results was a big outlier, a fourth trial would be performed to confirm uniformity. The highest score within the 10% variability options was considered for this analysis.

Rasch-built DM1-ActivC scale: a 25-item rasch-built scale that measures DM1 patients participation level in daily life activities based on the ICF concepts of functioning and disability. Has proven high internal consistency and good tes-retest reliability. Correlates with manually-tested muscle strength and to MIRS

score (73, 89). This questionnaire has been selected as the primary outcome on the OPTIMISTIC trial (69).

The Myotonic Dystrophy Health Index (MDHI) questionnaire: is a disease-specific- patient reported outcome instrument created under guidance from the FDA. Composed by 114-items each one representing a possible symptom of the disease. Scores from zero to 100 with the higher the score the higher the reported disease severity. There is evidence of excellent test-retest reliability and of sensitivity and specificity to differentiate DM1 severity. On the validation study, fatigue was the symptom with the highest mean scoring and the biggest difference shown between groups was between the employed and the unemployed (80, 81, 90). The MDHI is currently being used as a patient relevant outcome measure in clinical trials including the Phase II IONIS-DMPKRx Therapeutic Treatment Trial for DM1 adults and as secondary outcome on natural history studies in the United States and on the OPTIMISTIC trial.

Inter-rater reliability: aiming to reduce inter-rater variability as much as possible the following strategies were implemented: [1] SOP for each of the FOM and strength assessments were redacted including pictures of the participant and examiner positioning; [2] an initial training session for each of the assessors involved was led by the site's lead physiotherapists who have wide experience in muscle strength assessment in neuromuscular conditions and clinical research training standards looking for an inter-rater agreement level considered "satisfactory" in the eyes of the trainer; and, [3] manual muscle testing and QMT were performed three times consecutively looking for a difference between scores no bigger than 10% and when this happened a fourth assessment was performed. This methodology replicates strategic procedures commonly used by physiotherapists when performing clinical trials in other neuromuscular disorders such as Duchenne Muscular Dystrophy and Spinal Muscular Atrophy (91-94).

3.2.3. Statistics

Normality was tested by utilizing the Shapiro-Wilk test. Independent samples (t-test and Mann-Whitney U-test) were used to analyse differences between subgroups and for the Bland-Altman analysis. Correlation tests are presented as Pearson's rho scores and scores ≥0.50 have been highlighted as strong values (95).

Intraclass correlation coefficients ($ICC_{2,1}$) were tested for test-to-test relative reliability (60). The paired t-test and Mann-Whitney test were used for differences between tests and over time and to assess responsiveness over time (96).

Bland-Altman plots were used to check the distribution of the difference between scores and the agreement level between tests, identifying any possible systematic bias (63).

Receiver operating characteristic (ROC) was used to test accuracy and to determine any possible cut-off points with optimal sensitivity (proportion of true positive results) and specificity (proportion of false positives). In a ROC curve the sensitivity is plotted as a function of the false positive rate (100-specificity) for different cut-off points. Each point representing a sensitivity/specificity pair-ratio that matches to a particular outcome-result threshold-score. A test with perfect discrimination (no overlap between sensitivity and specificity percentages to distinguish the presence or absence of a defined characteristic) has a ROC curve that passes through the upper left corner. Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test (65, 66). The area under the curve (AUC) from this test then identifies participants with or without certain characteristics (e.g. disease or not, or disease sign or not); the closer to 1 the better discriminative ability the test has (97). In this study, the FOM have been tested for specificity and sensitivity to detect the following two outputs (or disease characteristics): ROC [1] participants with a SARA score of eight or above; and ROC [2] participants with proximal weakness (i.e. a MIRS score of IV or V).

Standard error of measurement (SEM) and SEM% were used to determine the limit needed for the smallest change in a group mean to be considered a real

clinical change. SEM is the within-subject standard deviation (square root of the within-subject variance) (98).

SEM=SD√(variance within scores)

SEM=SD(√(SD^2_1+SD^2_2+SD^3_n/n)

SEM%=SEM/mean of means *100

For sample size calculations, the standard deviation (SD) at 12 months was multiplied by $\sqrt{(1 - r^2)}$ to estimate for a sample size robust enough to detect a significant change in a randomised controlled trial if any of these FOM are selected as primary outcomes (99). Correlation, ICC, ROC and SEM scores presented are those obtained from estimations using the average of the three trials of each FOM. For all results, only p values ≤0.05 have been considered statistically significant and rest are shown as ns (non statistically significant).

3.3. RESULTS

Two hundred and thirteen (n=213) participants were screened between the two sites, 110 in Newcastle and 103 in London. Data from three participants were excluded from the analysis due to significant missing data. Age and 6MWT were the only variables identified with a normal (i.e. Gaussian) distribution. There was a similar distribution of male and female allowing within-sex comparisons. Twelve per cent of the participants reported using a wheelchair either part-time (such as for long distances) or full-time in their daily life but none were wheelchair-dependent as all were capable of completing at least the 10-mWT independently (orthotics or assistive devices allowed) as per inclusion criteria. The big majority of our sample (81%) presented with a MIRS score between II and IV, and the most commonly reported limitation to perform the functional tests was due to poor neuromuscular control, which included balance problems. Demographics are summarized in Table 3. 1.

Table 3. 2 presents the strength and FOM results for the whole sample and between the sexes at visit 1 (baseline). Statistically significant differences between males and females were identified for all assessments except hand-strength measurements. The 30SSS only showed significant differences between sexes when the best, the second and third trials were compared.

5 **Table 3. 1 Participant's Demographics.**

	All		Male		Female	
N=	210		103	49%	107	
		SD		SD		SD
Age	45.1	14.7	47	15.3	43.3	13.9
Wheelchair users in DLA	26	12%	11	11%	15	14%
MIRS						
I: no muscular impairment	22	10%	15	14%	6	6%
II: minimal muscular impairment	58	27%	23	22%	36	33%
III: distal weakness	46	22%	26	25%	20	19%
IV: mild proximal weakness	67	32%	29	28%	38	35%
V: severe proximal weakness	17	8%	10	10%	7	7%
Walking accessory						
none	171	81%	83	80%	88	82%
cane	26	12%	14	14%	12	11%
crutches	2	1%	2	2%	0	
walker	2	1%	1	1%	1	1%
Limiting Factors reported by examiner						
Pain	17	8%	6	6%	11	10%
Poor Neuromuscular Control	34	16%	23	22%	11	10%
Paresis	1	0%			1	9%
Fatigue	10	5%	4	4%	6	6%
Other	1	0%			1	1%

N: number of participants per sample, SD: standard deviation, MIRS: muscular impairment rating scale

Table 3. 2 Strength values and Functional Outcome Measures (FOM) at baseline.

Outcome Measure	All		Male		Female		sig. between sexes
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	
SARA	5.4	4.6	5.7	4.8	5.1	4.3	ns
Grip strength (QMT – kg)	16.8	12.8	19.8	15.3	13.9	7.7	ns
Wrist Extensors (QMT – lbs)	15.6	8.4	17.3	10.3	14.0	5.8	ns
Knee Extensors (QMT – lbs)	45.8	20.1	54.1	19.5	38.0	17.4	<0.001
Ankle Dorsiflexors (QMT – lbs)	25.7	13.4	28.9	15.3	23.1	11.0	0.02
Hip Flexors (QMT – lbs)	33.4	13.0	39.1	13.2	28.0	10.4	<0.001
6MWT (metres)	415.4	148.7	440.4	155.7	389.9	138.1	0.02
10-mWT (s)	9.9	4.4	9.3	4.1	10.5	4.6	0.008
10-mWT 2nd Trial	9.6	3.8	8.8	2.6	10.4	4.5	0.002
10-mWT 3rd Trial	9.5	4.3	8.4	2.6	10.4	5.0	<0.001
Average 10m-WT	9.3	3.5	9.2	4.1	10.5	4.6	0.03
Best 10m-WT	9	3.3	8.8	4	10	4.3	0.04
Worst 10-mWT	10.2	4.6	9.5	4.1	10.9	4.9	0.03
10-mW/RT (s)	6.1	3.4	5.4	3.4	6.8	3.3	<0.001
10-mW/RT 2nd Trial	5.6	2.8	4.9	2.9	6.2	2.6	<0.001
10-mW/RT 3rd Trial	5.4	2.7	4.5	2.6	6.2	2.6	<0.001
Average 10-mW/RT	5.7	3.1	5.3	3.3	6.6	3.3	0.003
Best 10-mW/RT	5.4	3.1	5	3.3	6.4	3.2	0.003
Worst 10mW/RT	6.3	3.4	5.6	3.5	6.9	3.3	0.006
30SSS (times)	10.7	6.1	11.2	6.6	10.2	5.5	ns
30SSS 2nd Trial	13.4	5.6	14.2	6.0	12.6	5.1	0.05
30SSS 3rd Trial	16.6	11.0	16.9	5.8	14.6	5.4	0.05
Average 30SSS	11.5	6.5	12.1	6.9	10.9	5.9	ns
Best 30SSS	13.4	6.2	14.4	6.6	12.4	5.7	0.03
Worst 30SSS	10.5	6	11.1	6.6	10.1	5.5	ns

Total stands							
30SSS	30.7	20.2	32.4	21.4	29.1	18.9	ns

SARA: severity of ataxia rating scale, QMT: quantitative muscle test, 6MWT: six-minutes walking test, 10-mWT: ten meter walk test (comfortable speed), 10-mW/RT: ten meter walk/run test (as fast as possible), 30SSS: 30 seconds sit and stand test, Total stands 30SSS: total stands accomplished in all completed test trials, SD: standard deviation, ns: not significant.

There was a significant correlation between FOM, strength values and disease severity as scored by the SARA (r= 0.6 to 0.7, p<0.01) and the patient reported outcomes (PRO) (r=0.5 to 0.7, p<0.01) (tables 3.3a and 3.3b). There is a strong correlation between the four-selected FOM (r=0.6 to r=0.8, p<0.01). Age only correlated slightly but was significant with the 6MWT (r=0.1, p<0.05), 30SSS (r=0.2, p<0.01), and 10-mW/RT (r=0.2, p<0.01); with directions for each FOM suggesting a mild disease impact (i.e. less meters in 6MWT = higher age, more seconds in 10-mW/RT = higher age and less stands = higher age). 30SSS (following the 6MWT) (r=0.4 to 0.6, p<0.01) showed the strongest correlation values to muscle strength, in particular to lower limb strength (r=0.5, p<0.01) (Table 3. 3). Table 3. 4 shows the correlation values of the FOM with the PRO with significantly strong correlation values with DM1-ActivC overall score (r=0.6 to 0.7, p<0.01) and with MDHI mobility (r=0.7, p<0.01) and ability to perform activities (r=0.6, p<0.01) sub-scales. From all FOM, the walking-capacity assessments (i.e.6MWT and 10-mWT) showed the strongest correlation values with the patient reported outcomes (r=0.6 to 0.7, p<0.01). Reported fatigue, pain, social performance and upper extremity functionality showed moderate association with all the FOM and the SARA score but with less strength to the 10-mW/RT (r=0.4 to 0.5, p<0.01).

Table 3. 3 Correlations between FOM assessments and strength - Spearman's rho correlation coefficients.

	Age	SARA	grip strength	wrist ext.	knee ext.	ankle dorsiflex	hip flex.	6MWT	10-mWT	10-mW/RT
6MWT	-.14*	-.64**	.59**	.45**	.47**	.46**	.51**	1.0		
10-mWT	ns	.63**	-.51**	-.37**	-.50**	-.42**	-.45**	-.81**	1.0	
10-mW/RT	.21**	.55**	-.49**	-.33**	-.49**	-.46**	-.52**	-.80**	.81**	1.0
30SSS	-.2**	-.67**	.44**	.47**	.57**	.52**	.54**	.69**	-.64**	-.65**

*Correlation is significant at the ≤0.05 level (2-tailed)

**Correlation is significant at the≤ 0.01 level (2-tailed)

SARA: severity of ataxia rating scale, wrist ext.: wrist extensors quantitative muscle strength (QMT), knee ext.: knee extensors QMT, ankle dorsiflex.: ankle dorsiflexors QMT, hip flex.: hip flexors QMT, 6MWT: six minute walking test, 10-mWT: ten-meter walk test (average of three trials), 10-mW/RT: ten-meter walk/run test (average of three trials), 30SSS: 30-seconds sit and stand test (average of three trials).

Strong correlation rho values (≥0.50)

Table 3. 4 Correlations between assessments and disease-specific patient reported outcomes (DM1-ActivC and MDHI questionnaires) - Spearman's rho correlation coefficients.

Patient Reported Outcome	6MWT	10-mWT	10-mW/RT	30SSS	SARA
DM1-ActivC score	.69**	.67**	.59**	.65**	.72**
MDHI- mobility subscale	.73**	.73**	.66**	.65**	.71**

MDHI-Upper Extremity Functionality subscale	.49**	.46**	.38**	.46**	.54**
MDHI-Ability to perform activities subscale	.64**	.62**	.56**	.57**	.60**
MDHI-Communication subscale	.25**	.26**	.18*	.28**	.39**
MDHI-Social satisfaction subscale	.43**	.42**	.34**	.43**	.45**
MDHI-Social performance subscale	.45**	.45**	.36**	.43**	.45**
MDHI-Fatigue subscale	.50**	.51**	.43**	.48**	.45**
MDHI-Pain subscale	.54**	.49**	.44**	.50**	.47**
MDHI-Myotonia subscale	.52**	.48**	.40**	.48**	.52**
MDHI-Gastrointestinal issues subscale	.23**	.26**	.17*	.30**	.19**
MDHI-Swallowing subscale	.22**	.25**	.16*	.28**	.24**
MDHI-Vision subscale	.27**	.28**	.24**	.31**	.36**
MDHI-Emotional issues subscale	.33**	.34**	.26**	.31**	.37**
MDHI-Sleep subscale	.29**	.35**	.26**	.33**	.31**
MDHI-Cognition subscale	.22**	.27**	.21**	.25**	.25**
MDHI-Hearing subscale	.15*	.15*	ns	ns	ns
MDHI-Breathing subscale	.33**	.33**	.27**	.34**	.31**
Total MDHI	.56**	.56**	.46**	.54**	.55**

*Correlation is significant at the ≤0.05 level (2-tailed)

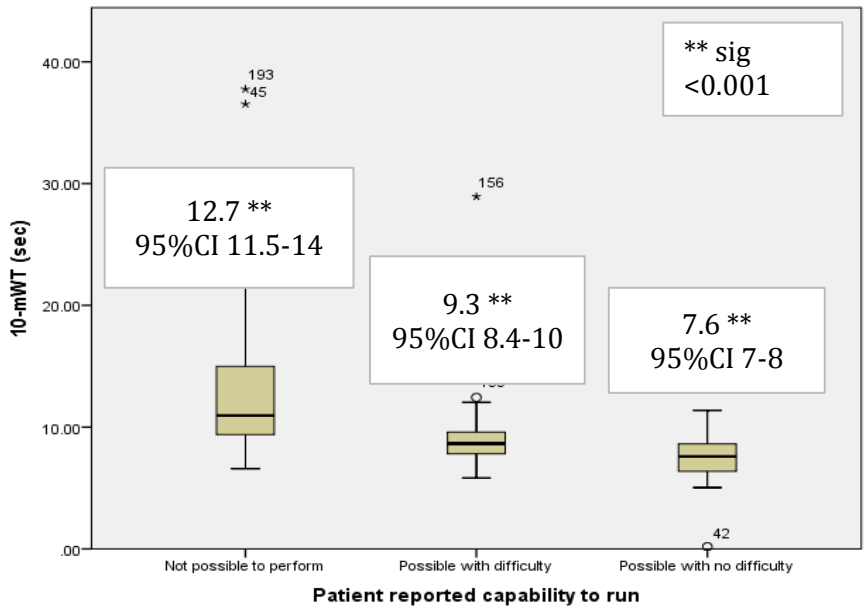
**Correlation is significant at the≤ 0.01 level (2-tailed)

6MWT: six minute walking test, 10-mWT: ten-meter walk test (average of three trials), 10-mW/RT: ten-meter walk/run test (average of three trials), 30SSS: 30-seconds sit and stand test (average of three trials), DM1-ActivC: Rasch-built DM1-ActivC daily life activities performance questionnaire, MDHI: myotonic dystrophy health index questionnaire, ns: not significant.

Strong correlation rho values (≥0.50)

Participants reported their capability of running as part of the DM1-ActivC questionnaire that exhibited the following distribution: [1] not possible to perform: 37%; [2] possible with difficulty: 27%; and [3] possible with no difficulty: 36%. These results were used to identify cut-off points to distinguish runners from non-runners with the 10-mWT and the 10-mW/RT (Figure 3. 1).

3.1a



3.1b

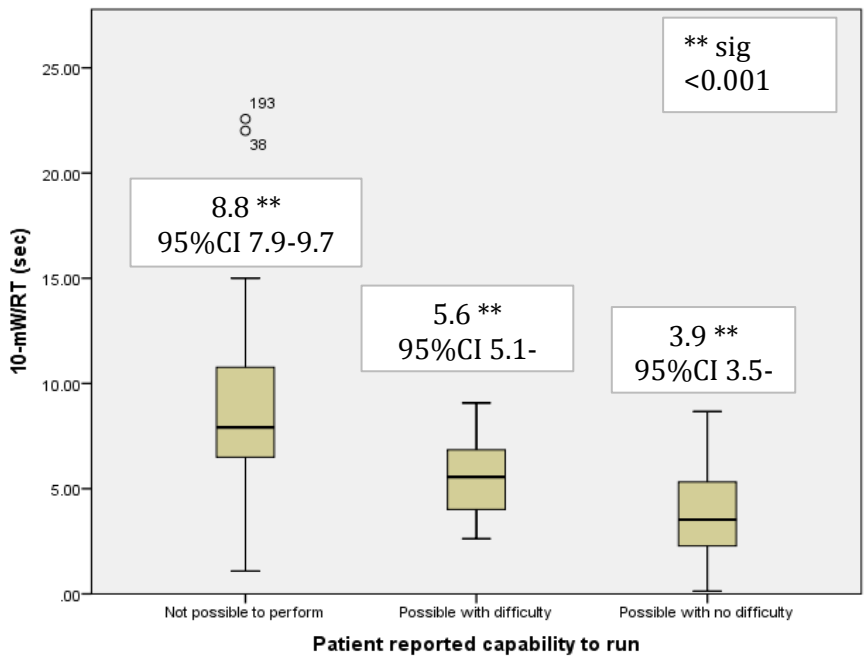


Figure 3. 1 Patient reported capability to run.

Figure 3.1a (10-mWT: ten meter walk test) and **Figure 3.1b** (10-mW/RT: ten meter walk/run test): data presented indicate the mean values per group with their

95% confidence intervals (95%CI). Groups were defined by participants' reported capability of running.

10-mWT, 10-mW/RT and 30SSS were attempted in three consecutive trials whenever possible. More than 90% of the participants capable of performing the test performed a second trial. If participants not capable of performing the 30SSS are scored as 'zero' on their first attempt, then 84% continue for a second trial; but, when only those capable of performing the test are considered, then 92% performed a second trial of the test. More than half of the participants completed the three trials of each test (Table 3. 5). The most common reason not to carry out a second or third trial was fatigue followed by fear of falling from either the examiners' or participants' point of view. Table 3. 5 presents the descriptive statistics per assessment trial for each FOM. More than half of the participants performed their best (or only) attempt at the first trial of the 10-mWT and the 10-mRT. The 30SSS had a similar distribution between the trials but with the majority of the milder participants (i.e. MIRS I and II) scoring better at the second or third trial. The more severe participants (i.e. MIRS V) scored their best at the first attempt. There was a statistically significant change ($p<0.001$) from the first trial to the second on all three FOM and between the second and third trials of the 30SSS (Table 3. 2).

The ICC test revealed a very strong Cronbach's alpha of: [1] 0.992 for the 10-mWT; [2] 0.987 for the 10-mW/RT; and [3] 0.979 for the 30SSS. For the Bland-Altman analysis, when the mean difference between the means was significantly different from zero (i.e. first vs. second trial of the 10-mWT, 10-mW/RT and 30SSS and second vs. third trial of the 30SSS) a plot is not suitable as these measurements do not agree with each other. Between the second and the third trials of the 10-mWT and 10-mW/RT there was a mean difference of 0.04 (95% interval of agreement = -2 to 2.1) and 0.6 (95% interval of agreement = -1.1 to 1.2) respectively (Figure 3. 2).

ROC curve results showed that 30SSS and 10mW/T are good tests to discriminate between participants with a SARA score of eight or above and are

fair tests to discriminate between participants with proximal muscle weakness (Figure 3. 3). The 10mW/RT showed fairness for both of the outputs (Figures 3.3e and 3.3f). Table 3. 6 presents a range of suitable cut-off points when considering any of these FOM to predict the tested outcomes and those in **bold** have highlighted those considered more suitable for consideration based on a balance between sensitivity and specificity aiming for a sensitivity >60% and a specificity <40% and that could apply equally for both tested outcomes. For the 30SSS test, 11 full stands are recommended as the most appropriate cut-off value.

The results of the SEM (SEM%) estimations were: [1] 0.53 sec (5%) for the 10-mWT; [2] 0.44 sec (8%) for the 10-mW/RT; and [3] 1.4 times (10%) for the 30SSS. 34 participants assessed for a second time 12.1 (SD0.8) months after the baseline showed a significant disease progression as detected by all FOM in at least one score, the SARA and knee extensors strength (Table 3.6). All FOM showed at least one output with a significantly higher change score than the minimal expected SEM and SEM%. In order from more to less, the sequence of change from baseline (%) was: 1st SARA score (32%), 2nd total stands of 30SSS (26%), 3rd the best (22%), the average and the worst (21% both) 10-mW/RT, 4th the best 30SSS (19%), 5th average (16%), the best (15%) and the worst (14%) 10-mWT and the average 30SSS (16%) and, last the 6MWT and the worst 30SSS (13% both)(Table 3. 7).

Table 3. 5 Descriptive statistics and performance prevalence per trial for all FOM sub-divided by participant's MIRS score. Presented are: mean and standard deviation (SD) for each trial, prevalence (percentage) of participants performing the test at each trial and frequency (percentage) of participants performing their best at each trial.

		1st trial				2nd trial				3rd trial				Best of	
	MIRS	Part.	mean	SD	% best	Part.	mean	SD	% best	Part.	mean	SD	% best	mean	SD
30 SSS	I	100%	16.2	6.0	10%	100%	18.1	6.9	38%	76%	19.1	5.8	52%	18.7	7.1
	II	100%	13.1	4.9	28%	96%	14.7	5.4	39%	68%	17.2	5.0	33%	15.1	5.7
	III	100%	11.1	6.3	39%	80%	14.5	5.1	28%	57%	15.9	5.6	33%	13.2	6.9
	IV	100%	8.5	4.3	35%	86%	10.4	3.4	34%	48%	11.6	4.4	31%	9.9	5.0
	V	100%	3.5	4.2	71%	47%	8.1	4.3	18%	18%	14.0	0.0	12%	4.4	5.4
		N= 210	**10.7**	**6.0**	**34%**	**177**	**13.4**	**5.6**	**33%**	**115**	**15.6**	**5.6**	**33%**	**12.5**	**6.9**
10-m WT	I	100%	7.0	1.1	45%	100%	7.0	1.2	32%	64%	7.1	1.4	23%	7.3	1.2
	II	100%	8.4	2.4	52%	98%	8.1	2.1	31%	76%	7.7	1.4	17%	8.6	2.4
	III	100%	10.2	4.2	52%	100%	10.0	4.1	35%	80%	10.1	4.7	13%	10.4	4.3
	IV	100%	11.2	4.6	52%	97%	10.9	4.5	34%	70%	11.1	5.1	13%	11.5	4.8
	V	100%	13.4	7.9	59%	88%	11.8	3.0	29%	47%	11.7	4.9	12%	13.6	7.8
		N= 210	**9.9**	**4.4**	**52%**	**205**	**9.6**	**3.8**	**33%**	**150**	**9.5**	**4.3**	**15%**	**10.2**	**4.6**
10-m W/RT	I	100%	3.3	1.4	77%	100%	3.0	1.4	23%	86%	3.0	1.3	0%	3.3	1.4
	II	100%	4.9	2.6	64%	97%	4.5	1.9	28%	81%	4.4	2.0	9%	5.1	2.7
	III	100%	6.3	3.4	57%	98%	6.1	3.2	30%	70%	5.9	2.4	14%	6.5	3.4
	IV	100%	7.3	3.5	58%	95%	6.7	2.5	23%	68%	6.9	2.7	19%	7.5	3.4
	V	100%	8.6	4.0	75%	69%	8.1	3.3	19%	50%	8.1	2.6	6%	8.8	3.9
		N= 202	**6.1**	**3.4**	**63%**	**191**	**5.6**	**2.8**	**25%**	**147**	**5.4**	**2.7**	**12%**	**6.2**	**3.4**

Part. column presents the percentages (%) grand total (N) of participants performing the test at each trial, mean: average score at each trial, SD: standard deviation, % best: percentage of participants performing their best test at each trial, MIRS: muscular impairment rating scale (I: no muscular impairment, II: minimal muscular impairment, III: distal weakness, IV: mild proximal weakness, V: severe proximal weakness), 30SSS: 30 seconds sit and stand, 10-mWT: ten meter walk test, 10-mW/RT: ten meter walk/run test.

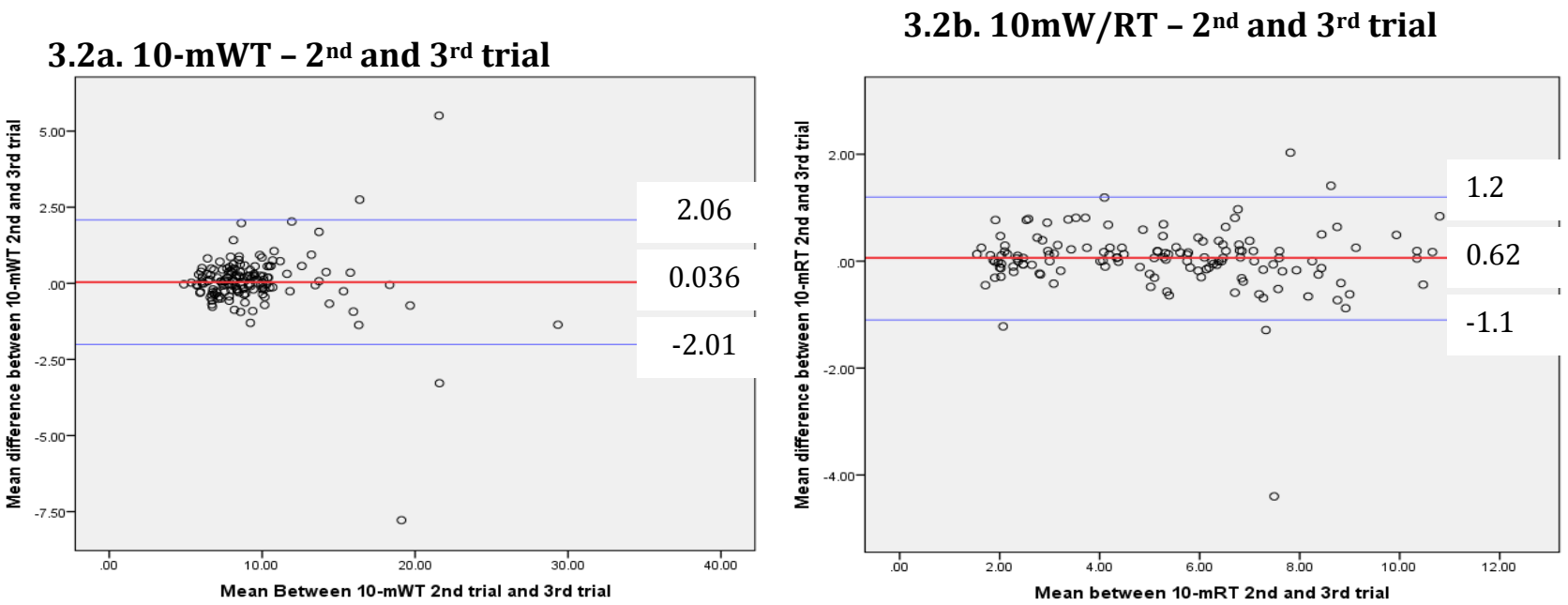


Figure 3. 2 Bland-Altman Plots (identified agreement within test) between the second and third trial.

Figure 3.2a. Bland-Altman Plots between the (10-mWT) ten-meter walk test second and third trials; and, **Figure 3.2b**. Between the (10-mW/RT) ten-meter walk/run test second and third trials.

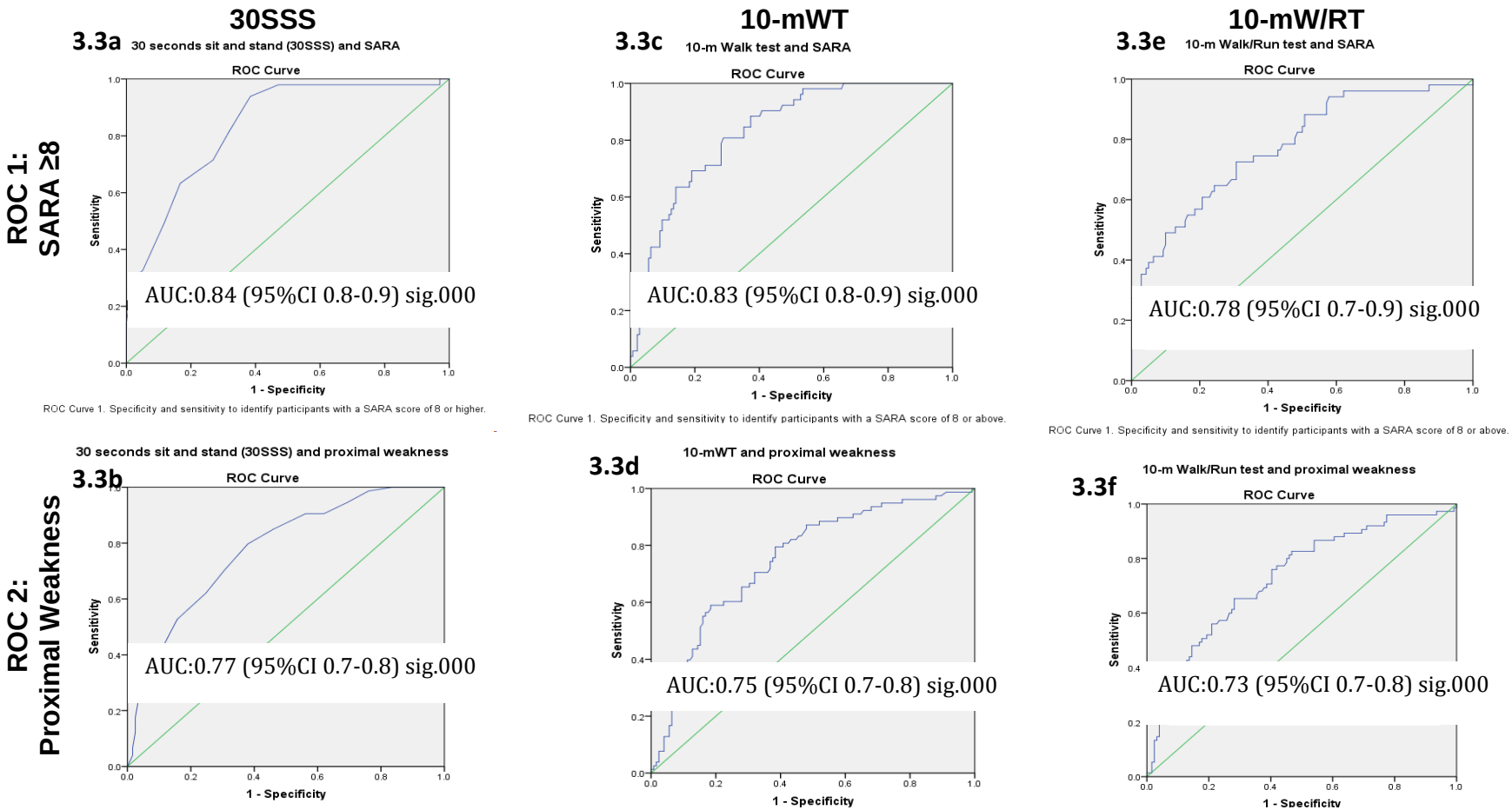


Figure 3. 3 ROC curves.

Figures 3.3a to 3.3f. Receiver operating characteristics (ROC) curve to estimate cut-off points for identifying patients with the presence (sensitivity) or not (specificity) of the following outcomes representing disease severity: ROC 1 (top): a SARA score of 8 or above; and ROC 2 (bottom): proximal weakness (i.e. MIRS score of IV or V). 3.3a and 3.3b: 30SSS (thirty-seconds sit and stand test); 3.3c and 3.3d: 10-mWT (timed ten-meters walk test); and, 3e and 3f: 10-mW/RT (timed ten-meters walk/run test.

Table 3. 6 Assessment cut-off values with respective sensitivity and specificity levels to identify the following outcomes:

[1] a SARA score of 8 or above; and [2] presence of proximal weakness measured by a MIRS score of IV or V.

Test	Outcome	Cut Off Value	Sensitivity	Specificity
30SSS	SARA score ≥8	7.5 stands	49%	12%
		10.5 stands	**82%**	**32%**
		11.5 stands	**94%**	**38%**
		12.5 stands	98%	47%
	Proximal weakness	7.5 stands	42%	11%
		10.5 stands	**70%**	**31%**
		11.5 stands	**80%**	**38%**
		12.5 stands	85%	46%
		15.5 stands	95%	69%
10-mWT	SARA score ≥8	10 secs	63%	15%
		9 secs	**77%**	**28%**
		8 secs	92%	48%
		7.9 secs	96%	53%
	Proximal weakness	10 secs	52%	15%
		9 secs	**65%**	**28%**
		8 secs	87%	48%
		7 secs	95%	71%
10-mW/RT	SARA score ≥8	8 secs	47%	10%
		7 secs	58%	21%
		6 secs	**73%**	**34%**
		5 secs	84%	50%
		4 secs	96%	62%
	Proximal weakness	7 secs	49%	18%
		6 secs	**65%**	**33%**
		5 secs	81%	46%
		2.7 secs	95%	77%

30SSS: 30 seconds sit and stand test, 10-mWT: ten meter walk test, 10-mW/RT: ten meter walk/run test, SARA: severity of ataxia rating scale. In **bold** are presented the cut-off values considered most appropriate based on a balance of good sensitivity level and low specificity.

Table 3. 7 SARA, Strength and Functional Outcome Measures (FOM) progression over time.

Results of a paired t-test analysis.

Outcome Measure	N=	Mean	SD	Mean Change	95 % CI	% from baseline	sig. of change
SARA	33	7.6	5.9	1.8	2.8 to 0.9	32%	0.00
Grip strength (kg)	33	18.3	15.1	-0.4	2.0 to -2.8	2%	ns
Wrist Extensors (lbs)	32	15.8	9.1	0.2	1.9 to -1.4	2%	ns
Knee Extensors (lbs)	33	52.4	22.1	5.9	11.5 to 0.4	13%	0.04
Ankle Dorsiflexors (lbs)	22	26.7	14.6	2.2	7.4 to -3.0	9%	ns
Hip Flexors (lbs)	27	37.6	11.7	4.0	10.3 to -2.2	12%	ns
6MWT (metres)	34	399.6	194	-45.7	-17 to -75	-10%	0.00
10-mWT (s)	34	10	4.8	1.2	2.2 to 0.1	13%	0.03
10-mWT 2nd Trial	29	8.8	1.9	0.7	1.1 to 0.2	8%	0.00
10-mWT 3rd Trial	10	7.8	1.1	0.9	1.6 to 0.1	11%	0.04
Average 10-mWT	34	10.2	4.8	1.4	2.4 to 0.3	16%	0.01
Best 10-mWT	34	9.8	4.8	1.4	2.5 to 0.3	15%	0.01
Worst 10-mWT	34	10.3	4.7	1.3	2.3 to 0.3	14%	0.02
10-mW/RT (s)	31	6.3	4.4	1.0	1.9 to 0.2	20%	0.02

10-mW/RT 2nd Trial	25	4.8	2	0.6	1.0	to	0.1	13%	0.02
10-mW/RT 3rd Trial	9	3.7	1.5	0.3	1.3	to	0.7	8%	ns
Average 10-mW/RT	31	6.1	4.4	1.0	1.9	to	0.1	21%	0.02
Best 10-mW/RT	31	5.9	4.4	1.1	2.0	to	0.1	22%	0.03
Worst 10-mW/RT	31	6.3	3.4	1.1	1.9	to	0.2	21%	0.03
30SSS (times)	34	8.9	7.2	-1.4	-0.1	to	-2.7	-13%	0.03
30SSS 2nd Trial	22	13.1	6.3	-1.0	0.2	to	-2.1	-7%	ns
30SSS 3rd Trial	6	15.7	3.1	-1.0	1.1	to	-3.1	-6%	ns
Average 30SSS	34	9.2	7.3	-1.8	-0.7	to	-2.9	-16%	0.003
Best 30SSS	34	10.1	7.3	-2.4	-1.1	to	-3.7	-19%	0.00
Worst 30SSS	34	8.9	7	-1.3	-.15	to	-2.4	-13%	0.03
Total stands 30SSS	34	20.2	18.8	-7.1	-3.2	to	-11	-26%	0.001

SARA: severity of ataxia rating scale, QMT: quantitative muscle test, 6MWT: six-minutes walking test, 10-mWT: ten meter walk test (comfortable speed), 10-mW/RT: ten meter walk/run test (as fast as possible), 30SSS: 30 seconds sit and stand test, N= sample size per outcome measure, SD: standard deviation, Mean change: mean change from corresponding baseline values, 95%CI: 95 percent confidence intervals around the estimated mean change, % from baseline: mean change/baseline value * 100, ns: not significant.

3.4. DISCUSSION

Assessing functionality in people with DM1 is essential to monitor natural disease progression and the possible effect of any intervention. The OMMYD consortium suggests five functional outcome measures (FOM) considered suitable for DM1. This study explores the feasibility, reliability and validity of three of these outcomes: the 10-mWT; the 10-mW/RT; and the 30SSS. The 6MWT has been explored before (39) so, for this case study, it is presented as a reference; finally, the Nine-hole Peg test (9HPT) has been included as an outcome measure in PHENO-DM1 study also, however the exploration of upper extremity outcomes has been planned as an independent project.

By the time of this study, few differences on the FOM operational methodology and the last version of the FOM standard operational procedures (SOP) agreed by the OMMYD consortium were identified; these variants were: [1] two trials of the 6MWT if possible instead of one; [2] to keep the 30 meters length corridor for the 6MWT if possible instead of 25 m; and, [3] a firm start for the 10-mWT instead of a flying start. Still we do not expect these variants to impact significantly on the comparability of these results neither to interfere on the study conclusions. The findings of thesis will be presented to the OMMYD consortium for consideration before final OMMYD SOPs get submitted for publication.

FOM validity and reliability:

This study provides a cross-sectional analysis of FOM in DM1 and validates their use in this population, proving them to be feasible, reliable and sensitive to change. This sample size and severity distribution allows for an extrapolation of results when comparing to other studies and reliable estimates (100). The cohort included participants with a wide spectrum of disease burden including wheelchair users and with MIRS from I to V. However, not all the participants were able to complete all the assessments three times as expected.

Significant differences exist between sexes as expected from a healthy population with men scoring better than women (83, 101, 102), as also shown in other progressive neurological disorders (103). This correlates with the

differences in lower limb muscle strength plus it is known that the step length of men tends to be longer so influencing the walking-test scores (104, 105). Correcting for height could reduce the level of difference in walking tests between the sexes; however, one aim of this study is to present reference values of a representative DM1 sample including categorization in the sexes with their natural characteristics (106). These differences do not necessarily indicate a more severe phenotype in women. In fact, in DM1 it has been reported that men more frequently than women have muscular weakness and disability (107). In this particular sample if we consider the SARA test and the MIRS classification as disease-severity parameters, the disease-severity distribution between the sexes shows no difference.

Mean values are comparable to other relevant physically impaired disorders (84, 108). However due to the possible variability in methodologies, cautious comparison should be made. The healthy reference values most commonly used for the ten meter walk test (10-mWT) and the ten meter walk/run test (10-mW/RT) are those established by Bohannon et al. in 1997; however, their estimations come from a 7.62 m timed length as the acceleration and deceleration phases were included in the 10-metre-length corridor (109). Hence, our results (10-mWT 1.4 m/s (SD3.8) and 10-mW/RT 2.5 m/s to 2.6 (SD5.4)) should not erroneously be compared to a healthy population in their 40s and 50s. They perform better compared to the ones published by Hammaren et al. from a DM1 sample of 10 participants (10-mWT = 10.2 (SD1.7), range 6.2-12.3 and 10-mW/RT= 7.7 (SD1.6), range 4.7-9.8) (44). We explored the relationship between participants' scores in these two tests and their perceived capability in running as scored in the DM1-ActvC questionnaire (Figure 3. 1). These interesting results suggest that, as a group, participants can distinguish and report their capability to perform a fast pace test. For this study, participants completed the DM1-ActivC before performing the FOM. However, the DM1-ActivC questionnaire has been questioned as the responses can be influenced by the participants' daily life challenges or experiences, more than a real capacity. Certainly, the scoring of activities such as vacuuming and running might be influenced by the participant's own experience or real need to perform them as a daily life task. It will be

interesting to ask participants to repeat the questionnaire after performing the tests and assess any impact of their recent experience of being challenged to run or walk as fast as possible.

Normative data for the 30 seconds sit and stand test (30SSS) for a population between 20 to 80 years old range from 13 to 15 for women and from 14 to 17 in men (82, 110). In our population only participants performing the third trial accomplished these scores as a group. The averages presented on the first trial will be below the cut-off values predicted for a population between 60 and 70 years old (110, 111). Variations of this test (e.g. 5 times sit-to-stand, or time-up-and-go) have also good normative data for reference and good reliability; however, this does not allow a flooring effect and would have excluded all our participants not capable of standing up from the chair as required in these tests, with the 30SSS allowing a score of 'zero' for these cases and increasing the chances to quantify disease progression.

The correlation scores prove once more the influence of muscle strength and balance (SARA) on the ability to walk and to stand up from a chair in DM1 (39, 44, 45, 112, 113). The two walking tests (i.e. 6MWT and 10-mWT) maintain similar correlation trends among all tests. By showing similar correlation scores to those from the 6MWT, these tests have shown the same level of strength when assessing strength and disease severity as measured by SARA and PRO. Once more, knee extensors and ankle dorsiflexors strength have shown significant impact on test performance (44, 45).The correlation levels between the FOM and the mobility (MDHI subscale), ability to perform activities (MDHI subscale) and daily life activities participation (DM1-ActivC) scores corroborate that these FOM measures participants reported daily life performance and disease severity (overall MDHI). Interestingly the 10-mW/RT was not the FOM with the highest correlation values when is expected to be the most challenging test of the four, however this might be associated to the fact that a number of severe participants did not perform the test. Another explanation could be the measurement precision accomplished by each test. Still, the correlation values are moderate or strong for all FOM. This is the first time that the MDHI and DM1-ActivC questionnaires are tested against relevant tests of functionality. Associating physiological changes

with health-related quality of life enriches the evidence base complementing an outcome as a tool to assess meaningful interventions (114).

Grip strength is an outcome measure in DM1 with proven reliability and responsiveness, and has been suggested as disease severity surrogate marker capable of detecting real disease progression over time (27, 28, 115-118). In this study, associations to grip strength are presented as additional correlation test to an outcome representing disease progression but with no expected associated causality to the outcome measures (i.e. grip strength measures hand-muscle strength and is not expected to impact on the performance of the FOM).

The 30SSS ICC values were higher than previous reliability reports (from ICC=0.84 to ICC=0.92) (119, 120). For the 10-mWT and the 10-mW/RT there was a high relative reliability demonstrated with ICC values comparable to or surpassing what has been shown in DM1 in previous studies and in other neurological disorders (44, 121-123). Still, the variability between study methodologies when conducting these tests has been highlighted before so enhancing the need for uniformity in the protocols to allow validity comparability of data (124, 125). Due to the sample size involved in this study and the difficulty of collecting this amount of data with a standardized methodology in this population, it is recommended to consider using the same methodology for the 10-mWT and 10-mW/RT for any future clinical trial in DM1.

Sensitivity and specificity:

The ROC curve analysis showed fair and good levels when discriminating between participants according to outcomes representing participants with a more severe phenotype. The reasoning behind the selected outputs to compare to were: [1] a MIRS score of IV implies mild proximal weakness and a MIRS of V implies a severe proximal weakness, and it has been suggested that DM1 patients with MIRS ≥ IV are more likely to fall and have less balance confidence than those with MIRS ≤ III (43); and [2] a SARA score of 8 has been demonstrated as a good predictor of a patient's need for a walking-device and that those patients with scores ≥8 and no walking-device have a higher risk of severe falls (42). It is not suggested that these FOM are good predictors of falls:

for this conclusion to be valid it will require a direct correlation to falls history or observed falls. However, these outputs (i.e. SARA and MIRS) were selected as evidence-based reference cut-offs to identify more severe participants and with associated issues such as falling and ability to walk independently that can impact on daily life activities and the quality of life (43, 45, 126). Ambulation ability predicted by gait speed has shown to be a reliable method in other neurological population (127, 128).

Variability between tests:

When performed more than once, the 10-mWT and the 10-mW/RT showed good test-test reliability. When testing with functional outcomes, the need for at least one practice trial before establishing the most appropriate results for analysis has been previously reported not only in healthy populations but also in DM1 and other diseases with motor impairment and fatigue (39, 129-134).

In the particular case of DM1, not only the well-known learning effect is visible but also the possible combination with the natural lack of motivation predominantly present in these patients (135). Similar findings by Kierkegaard et al. for the following FOM: 6MWT, a variant of the 10-mWT, the timed-stands test (TST) and the time up-and-go (TUG) test where at least half of the participants performed their best test at either the second or the third trial. Due to the high level of agreement between the second and the third trial observed for the 10-mWT and the 10-mW/RT it is valid to assume that two trials of these tests will be enough to provide a valid and reliable score. Kierkegaard et al. suggested that there might be no need of repeated trials in the 10-mWT but that there is for additional tests such as the 6MWT, TUG and TST; with their results the question remains whether it would be better to report the first, the best or the average of all performed (129). In this study we have attempted to respond to this question as explained in the following sections.

This study section also provides information about the feasibility of performing all these tests three times plus the 6MWT once. It is fair to say that it is feasible to perform these three tests at least once as even after performing the 6MWT ≥96% completed all three. However, due to disease-associated limiting factors such as

fatigue, pain and poor balance an overlong examination of these patients is discouraged and a careful consideration for reducing the number of assessments or visit length is recommended. The 10-mWT and the 10-mW/RT had a good retention rate with ≥95% of performers having completed it by the second trial. Between the first two trials of these tests we also allow ≥80% of the participants to perform their best attempt. Due to the reduced variability of these two tests between the second and third trials this all adds evidence for the proposal of performing these tests only twice. The 30SSS had a lower retention rate; however, the more trials were requested the more chance participants had to perform their best trial; this correlated to the MIRS level. These findings could justify the attempt to perform the 30SSS test three times as long as the participant is willing to continue and the examiner considers it safe to continue. In this study the participants were always asked how they felt to continue and/or if they felt they could do a better trial; this allowed participants to continue and self-challenge to perform their best whenever possible.

Timed ten-meter walking tests and sit-and-stand task assessments have been correlated to balance gold standard scales (i.e. Berg Balance Scale) in DM1 and other neuromuscular disorders (44, 136, 137). In fact, the 10-mWT and the 30SSS have been recommended to supplement the ceiling effect commonly observed on the Berg Balance Scale when assessing walking ability affected by impaired balance (137). The 30SSS test has previously been correlated to fatigue (110, 138). In this study, fatigue impact can be suggested by the reduced compliance from trial to trial associated to participants' reported fatigue and by the moderate and strong correlation shown with the MDHI-fatigue subscale (table 3.4). Finally, 30SSS can be significantly influenced by the level of cognitive functioning (139). The possibility to assess other symptoms related to DM1 disease burden such as balance, fatigue or cognitive impairment could be seen as strength for interventions expecting a multisystem effect.

The results of the SEM (SEM%) estimations were: [1] 0.53 sec (5%) for the 10-mWT; [2] 0.44 sec (8%) for the 10-mW/RT; and [3] 1.4 times (10%) for the 30SSS. The 10-mWT and 10-mW/RT SEM% values are similar to those established by Flansbjer et al. for a post-stroke population of a mean age of 58

(SD6.4) years old with a 5.7% and 7.9% estimated change for each test (140). A SEM score for the 10-mWT and the 10-mW/RT was reported before for a sample of 10 DM1 patients, with an estimated SEM of 0.6 and 0.4, respectively (44). And the 30SSS are close to those established for patients with hip osteoarthritis (1.27 times) (79).

Longitudinal analysis:

The last section of the study aimed to document the progression of the disease as assessed by changes in strength and FOM. When analysing responsiveness to natural disease progression, the mean change after 12 months from baseline indicates a clinically relevant disease progression as measured by the progression detected in all FOM (6MWT included), the SARA as a disease severity parameter and knee extensor's strength, and a muscle group with the highest intra-rater QMT consistency. All FOM had at least three scores significantly higher than the SEM% which validates this change as real clinical change at group level. This compares to a longitudinal analysis testing upper limb performance test that only detect significant change in two of the four tools tested after nine years of follow-up (118). The scores with the highest change in percentage (%) from baseline can be suggested as the best score to report and/or include for analysis. The 10-mWT and 10-mW/RT had very similar results when considering the average of the performed trials or when considering only the best or the worst one. The 30SSS, on the other hand, showed a significantly higher sensitivity to change when the total stands achieved or the best trial score are considered. This adds up to the proposed encouragement for three trials in the 30SSS to elicit the participants' best score.

Based on the results of this study, preliminary estimates and power calculations for a randomised controlled clinical trial and a one arm observational study using these outcome measures as primary endpoint can be calculated. For a two-arm study, a sample size of at least 65 participants per group would be required to allow the detection of a 20% difference or change in the FOM performance with 90% power (99, 141). For a one-arm study, a cohort of 40 patients is enough. This assumes that patient populations are similar to this study, and that the variations of FOM values are not higher. Both the Optimistic and the PhenoDM1

studies have recruited more participants per study arm, so would be likely to generate statistically significant results. However an important point to analyse in the future is how much change is really needed for it to be considered a relevant change for the patient. In this case we estimated for a 20% difference between groups or change in time to allow at least a change of two times the expected SEM% for each FOM. These findings are just estimates and will need further investigation as every trial has unique characteristics to be added to the estimations.

Limitations from this study are: [1] these assessments were part of a day-long study visit, which may have contributed to fatigue; [2] each examiner might have different criteria when deciding to stop a test, which have not been accounted for as a reason for discontinuing the trials; [3] the reduced sample at follow-up in comparison to the baseline sample size. The final analysis including the whole sample at follow-up and will compensate for some of these limitations. This would reduce intra-examiner' variability and might identify significant correlations to other variables not considered at this stage. The difference in progression between the mild (or late-onset) and the classic (or adult-onset) phenotypes has been detected before and there is an encouragement to compare progression with these phenotypes as independent subgroups but for this, we would need the bigger sample.

3.5. CONCLUSIONS

Based on a systematic review published in 2017 about muscle and performance-based assessment instruments in DM1 this is the first time that a full analysis of reliability, validity and responsiveness has been made for the 10-mWT, the 10-mW/RT and the 30 SSS(142).

Overall, this study has defined a baseline and twelve-month follow-up reference suitable for future studies interested in assessing functionality in DM1. It has identified possible limitations of the assessments like the variability between each subject's ability to complete the full set of assessments when part of a long and complex study visit.

It is recommended to follow this study's methodology when considering any of these FOM for clinical research to allow an appropriate comparison for these reference values. When performing the 10-mWT and 10-mW/RT it is suggested to perform two trials and record the best score or the average of the two trials. For the 30SSS the attempt of three trials is encouraged, giving a chance for the best performance to occur at any of the attempts and report the score of the best performance.

For interventions offering improvement in DM1 patients' functionality and/or strength, it is encouraged to perform these three tests together. Even though they all assess strength and disease severity, they seem to complement each other as 30SSS most probable provides more information about balance and fatigue to the table, whereas the other two tests (10-mWt and 10-mW/RT) directly assess the capability and confidence to walk, and walk or run at maximal speed. This FOM battery may make the 6MWT redundant in order to assess strength, walking capability and disease severity; plus, the feasibility of performing the 10-mWT for more than one trial is better than performing more than one trial of the 6MWT.

A cohort of 34 DM1 participants has shown a significant increase in disease severity and decline in FOM scores above the minimum expected which should be considered as real change and not an error in the measurement. However, to better understand the progression of the disease functional-impairment with these FOM, further studies need to evaluate the natural FOM progression as whole cohort and in the mild and classic phenotypes as independent subgroups. Also, the use of repeated measures with more than two follow-up visits in the observational study would enrich final conclusions.

CHAPTER 4. EXPLORING WEARABLE ACCELEROMETRY-BASED ACTIVITY MONITORS IN MYTONIC DYSTOPHY TYPE 1

4.1. INTRODUCTION

Over the last decades, new technology has been refined to monitor activity of daily living with potential use in research and health-care practice for a variety of diseases including neuromuscular disorders (143, 144). Habitual physical activity (HPA) refers to any activity performed in the natural environment and it measures someone's *participation* and functionality in daily life (19, 145). One of the most advanced ways for assessing HPA levels is by quantifying objectively the body acceleration when moving (accelerometry). Hence, accelerometers have been used to estimate functioning, disability and health by assessing movement quality and movement persistence in daily life activity (52, 146, 147). However the understanding of these tools' applicability in diseases like DM1 is in its early stages and there is still work to do before presuming validity and reliability for any clinical trial.

The aim of this study was to explore wearable accelerometry-based technology in Myotonic Dystrophy type 1 (DM1) to assess habitual physical activity patterns. It will do so by addressing the following questions: (1) **Part I:** What sorts of activity monitor devices have been used previously to measure HPA in DM1 and similar neuromuscular disorders; and: Which methodology have these studies followed; (2) **Part II:** Is it valid to assess ambulation and other functional activities in DM1 with accelerometry; and: Is there a location on the body for these devices to assess walking activities in DM1 better; and (3) **Part III:** What would an accelerometry-based device tell us about the HPA patterns in the fatigued DM1; How does it differ from other fatigued cohorts and from the non-affected by DM1; and: Does HPA in DM1 change over time?

4.2. METHODS

4.2.1. Part I: Systematic Review [Appendix B]

Aiming to identify previously used activity monitors and assessment methodologies in DM1, a systematic review was performed by searching for any published study reporting the use of an activity monitor assessing HPA in neuromuscular disorders (148). For the purposes of the publication, patient-reported outcomes, assessing HPA in neuromuscular disorders, were also included; however, for this thesis, only the search part related to activity monitors will be presented.

Search Methodology

An initial literature search was performed through the electronic databases: EMBASE, MEDLINE and PsychINFO including the terms “physical activity”, “free living activity” and “daily life activity” in combination with “neuromuscular disease(s)”, “neuromuscular disorder(s)”, “muscular dystrophy” and “muscle disease(s)”. The initial search was performed and the following selection process followed a systematic search methodology. The titles and abstracts of all retrieved references were screened excluding. All papers that did not fulfil the inclusion criteria or had an evident exclusion criterion were excluded. Publications selected as possible for reporting the use of HPA measures in NMD were proof reviewed by two other independent researchers (SC and JN) and only those with common consensus were included for analysis.

Inclusion and Exclusion Criteria

The search included publications between 1996 and March 2016 (time of the literature search).

Papers included for analysis fulfilled the following criteria: (1) including participants with a progressive neuromuscular disorder; and (2) meeting the definition of habitual physical activity as in daily life (145, 149). Publications were excluded when: (1) not published in English; or (2) single cases, reviews, conference abstracts or pre-clinical studies.

Data extraction

At first, a selection of papers fulfilling the inclusion criteria was made and papers were evaluated independently with the collaboration of two independent

reviewers (SC and JN) identifying the following study variables: (1) sample size and age distribution; (2) study design and follow-up duration; (3) study aim; (4) study primary outcome; and (5) any HPA-related results reported. If any information was not reported or specified it was recorded as not available.

4.2.2. Part II: Validity of an accelerometry-based device and site of placement in DM1, a cross-sectional study.

When performing the systematic review two studies were identified as ongoing at that time but using accelerometry-based activity monitors in DM1. One of these was the OPTIMISTIC study (NCT02118779), which used an ankle-worn accelerometer (GENEActiv) for two weeks after each study visit (150, 151). To test the concept of using GENEActiv devices with DM1 patients, this study explored the use of GENEActiv on the ankle as against the wrist in a DM1 group and compared it to a healthy cohort.

The study was covered under the ethical approval of the PHENO-DM1 study (NCT02831504) by The Newcastle and North Tyneside Ethics committee (Re: NE/15/0178).

Sample

Participants were recruited as part of the ongoing DM1 natural history study, PHENO-DM1. This cohort includes 30 patients recruited to one of the sites (Royal Victoria Infirmary - Newcastle Upon Tyne NHS Foundation Trust) and assessed at baseline. Selected participants were all genetically confirmed DM1 adults able to consent and to perform the functional assessments independently. Walking assistive devices and orthosis were permitted. Patients were classified as mild if they met two of the three following criteria: 1) First symptoms reported at the age of 40 or older; 2) 200 or fewer CTG repeats as mutation length; and 3) a score of 1 or 2 on the Muscular Impairment Rating Scale (MIRS). This DM1 cohort was compared against a healthy-control group formed by adult volunteers from Newcastle University (students and staff). The collection of data from the healthy volunteers was covered under internal ethical approval of the university.

Study assessments

All participants wore four accelerometers at the same time: one on each limb (right wrist, right ankle, left wrist and left ankle) while performing different functional tasks. Functional assessments performed were performed in the following order: (1) stand still for a minimum of ten seconds; (2) six minutes walking test (6MWT); (3) ten meters walking test (10-mWT); and (4) ten meters walk/run test (10-W/RT). The aim of these ordered study assessments were to obtain accelerometry data representative of different walking paces. The protocol for these tests has been described in the previous chapter.

The following time points were recorded: (1) time when devices were placed on; (2) time participant started to perform each task; (3) time participant stopped performing any task; (4) resting periods, either sitting or standing; and (5) time when devices were taken off.

GENEActiv

The GENEActiv (Activinsights Ltd., Cambridgeshire, UK) is a tri-axial, ±6 g seismic acceleration sensor. It is portable device that measures 36 cm × 30 cm × 12 cm and weights 16 g. GENEActiv offers a near body temperature and light sensor to allow identification of wear and non-wear time. Wrist-worn GENEActiv has demonstrated strong validity against indirect calorimetry for both physical activity and sedentary behaviour (152, 153).

The unit of measurement presented is the Euclidean Norm Minus One (ENMO – mg), which not only considers the raw data of the three planes of acceleration provided by the device but also systematically includes gravity into its algorithm (as in $(x^2 + y^2 + z^2)^{1/2} – 1$) making it more reliable for dynamic physical activity estimations (150, 154, 155).

GENEActiv devices were configured to their maximum sampling frequency of 100 Hz. Downloaded (.bin files) were converted to 1s epochs and imported into a

custom-built Excel spreadsheet. Based on the assessor records, start and finish time for each of the functional tasks and the closest 10 seconds were plotted to identify the real start point as the point with a visible increment in value and continued until the recorded finish time. For each functional task's set of data, the mean value per second was calculated and then multiplied by 60 to obtain a value per minute for each task (156).

Statistics

Normality distribution was tested for each individual set of data with the Shapiro-Wilk test. Intraclass Correlation Coefficients (ICC) estimates and their 95% confident intervals were calculated based on a mean-rating absolute-agreement, 2-way random model to measure reliability within each accelerometer and between accelerometers (62). ICC measures the degree of correlation and the agreement between measurements provided by the device. Pearson correlation coefficients (r) are also reported for each functional test comparing an all-seconds sequence along the test. Bland-Altman plots were used to examine the agreement level between ankle measurements and wrist measurements for each functional task (64, 157). Sex (male/female) and age (years) were tested on an adjusted model as possible confounders as these factors were different between groups and might impact on the outcomes tested.

All statistical analyses were performed using SPSS version 23 (*IBM Corporation, SPSS Inc, Chicago, IL*) and only results with p-values ≤0.05 have been described as statistically significant.

4.2.3. Part III: Habitual Physical Activity in DM1, a longitudinal study.

This part of the study aimed to test the feasibility of the OPTIMISTIC protocol: an ankle worn GENEActiv to quantify habitual physical activity (HPA) levels in fatigued DM1 adults. This study has two parts: (1) a cross-sectional study comparing HPA patterns from a fatigued DM1 group against a healthy-volunteers group and a chronic fatigued group (CFS); and (2) a 16-month follow-up of the same DM1 group.

This study is covered under the ethical approval of the OPTIMISTIC trial approved by The Newcastle and North Tyneside Ethics committee (Re:13/NE/0342).

Sample

The fatigued DM1 cohort was recruited to the OPTIMISTIC multicentre study and these are participants from two of the four centres (Newcastle and Nijmegen). The OPTIMISTIC study is composed of four follow-up visits (including the baseline) and for each visit, participants will be requested to wear the GENEActiv device for 14 consecutive days, 7 of which will be selected for analysis (69). The healthy cohort corresponds to the same cohort used for the GENEActiv ankle-wrist validation study (part II). For this part, volunteers were asked to wear an ankle device for 7 consecutive days. Finally, the chronic-fatigued-syndrome (CFS) sample belongs to a population attending the Radboud University Medical Centre in Nijmegen, The Netherlands. CFS patients were screened along the same period of the DM1 cohort and by one of the OPTIMISTIC partners, following the same baseline protocol. Age, gender, BMI and fatigue severity were collected for each participant at baseline. Additional outcomes of disease severity were collected for the DM1 group at each study visit and considered for the longitudinal analysis. For the longitudinal analysis, the DM1 sample was subdivided into a: (1) treatment allocation group; and (2) classic and mild (or late onset) phenotype as explained earlier (158); allowing for exploration of any potential differences in these subgroups either at baseline or with time.

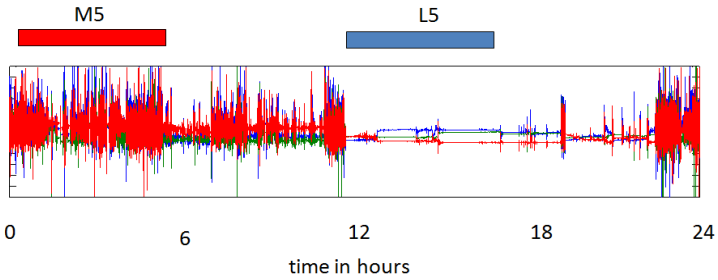
GENEActiv

HPA was measured by GENEActiv accelerometer worn on a daily basis in the participants' normal environment (159).

Instructions on how to use the GENEActiv were provided whenever the device was allocated to the participant and it was placed as applicable at the end of the visit. Participants were instructed to wear the GENEActiv monitor around the ankles of their non-dominant side (per-protocol) continuously for seven days (for the healthy cohort) and 14 days (for the DM1 cohort and the chronic fatigued cohort). The device is waterproof but participants were advised to take them off

when showering or taking a bath. Participants were expected to return the device by post when finished.

GENEActivs were configured with a sampling frequency of 50 Hz for 15 days. In this case, the system Matlab (Matlab v 12.0, Mathworks, Natick, MA) was initially used to filter the information. Seven consecutive days were expected to be extracted from the returned device. Data were considered valid when these



reported a wearing time of a minimum of 23 hours per day and a minimum of five full days had been extracted for the healthy controls and seven days for the DM1 and the CFS groups (160). The selected unit measure was ENMO (Euclidean Norm Minus One - mg), as in $(x^2 + y^2 + z^2)^{1/2} – 1$), and this was summarized as average ENMO (mg) values over 24 hours over the 7 days, the average of the most active 5 hours of the 7 days (M5 ENMO) and the average of the least active 5 hours of the 7 days (L5 ENMO) as shown in the figure (150, 151, 161).

Measures of fatigue severity

The Checklist of Individual Strength (CIS) is a 20-item long questionnaire that measures four different dimensions that impact on someone's perceived fatigue levels and are factors that may influence fatigue perception. The four dimensions include: (1) fatigue severity; (2) concentration problems; (3) reduced motivation; and (4) reduced physical activity (162). This outcome measure has shown very good internal consistency and reliability and with moderate to high correlation strength with other known fatigue scales. When assessing fatigue severity, a cut-off score of 35 has previously been used to distinguish those severely fatigued (163). All the participants of these studies completed a CIS questionnaire before wearing the accelerometer.

Measures of disease severity

Patient-reported disease severity was considered, based on scores from two disease-specific questionnaires: (1) the Myotonic Dystrophy Health Index (MDHI), a 114-item (symptoms prevalence and life-impact) questionnaire (81); and (2) the DM1-ActivC scale, a 20-item Rasch-built scale that measures the patient's participation in daily life activities (73). The Muscular Impairment Rating Scale (MIRS) was considered as the clinical disease severity outcome. MIRS is a disease-specific 5-point scale with which the clinician assesses and scores the weakness and impairment progression of the muscle affection (74).

Statistics

Descriptive statistics (mean and standard deviation [SD]) were used to describe the sample's characteristics. Baseline characteristics were compared between groups. An intraclass correlation coefficient (ICC) was estimated for each of the functional test results by comparing the average acceleration value per minute of each accelerometer. HPA progression over time was performed by a paired t-test between each time point and baseline for each individual group (164). Comparison between groups was tested with a non-parametric Mann-Whitney test at each time point. All statistical analyses were performed using SPSS version 23 (IBM Corporation) and only results with p-values ≤0.05 are presented as statistically significant. Sex (male/female) and age (years) were tested on an adjusted model as possible confounders as these factors were different between groups and might impact on the outcomes tested. A percentage of the standard error of the mean (SEM%) was assessed as a measure of within subject's deviation over seven days and it was calculated as followed: SEM%= (standard deviation x √(1-reliability)) x 100. The standard error of the mean was estimated from the standard mean ENMO change across seven consecutive days. This measurement can be used to determine the limit for the smallest change expected to claim a real change for a group of subjects with these characteristics following an intervention (165).

4.3. RESULTS

4.3.1. Part I: Systematic Review

The literature search retrieved 1,070 published titles and abstracts matching the search criteria. Eighty-nine full papers were selected by the first reviewer (CJM) and after detailed reviews by two independent researchers (SC and JN), 22 (166-187) were selected for analysis. Appendix B includes a summary of the protocols used by each of the identified papers recognized as assessing habitual physical activity in a form of neuromuscular disease. Sixteen of these studies (166, 168, 170, 172-179, 181-184, 187) quantitatively assessed activity levels as opposed to qualitative assessment (i.e. patient-reported outcome). Authors of six of these papers were contacted to complete information regarding the methodology used. Of the 22 papers, only three presented a clear attempt to systematically validate their utilized tool in their populations of interest (172, 178, 184).

There was only one study identified exclusive to DM1 participants; the study used the StepWatch activity monitor (SAM) for 7 consecutive days around the ankle (126), aiming to investigate the incidence of falls and stumbles in DM1. This cardinal study reported for the first time an increased falls rate in patients with DM1 when compared to healthy controls and that ankle strength correlates best with gait speed (r=0.92, p<0.001). Kalkman et al. **(182)** studied a large mixed neuromuscular diseased sample including a myotonic dystrophy group. They used the actometer Actilog V3.0 worn on the ankle for 12 consecutive days, showing a correlation between the actometer reports and the functional impairment scores obtained from the Sickness Impact Profile (SIP) questionnaire; however, this did not correlate with reported fatigue.

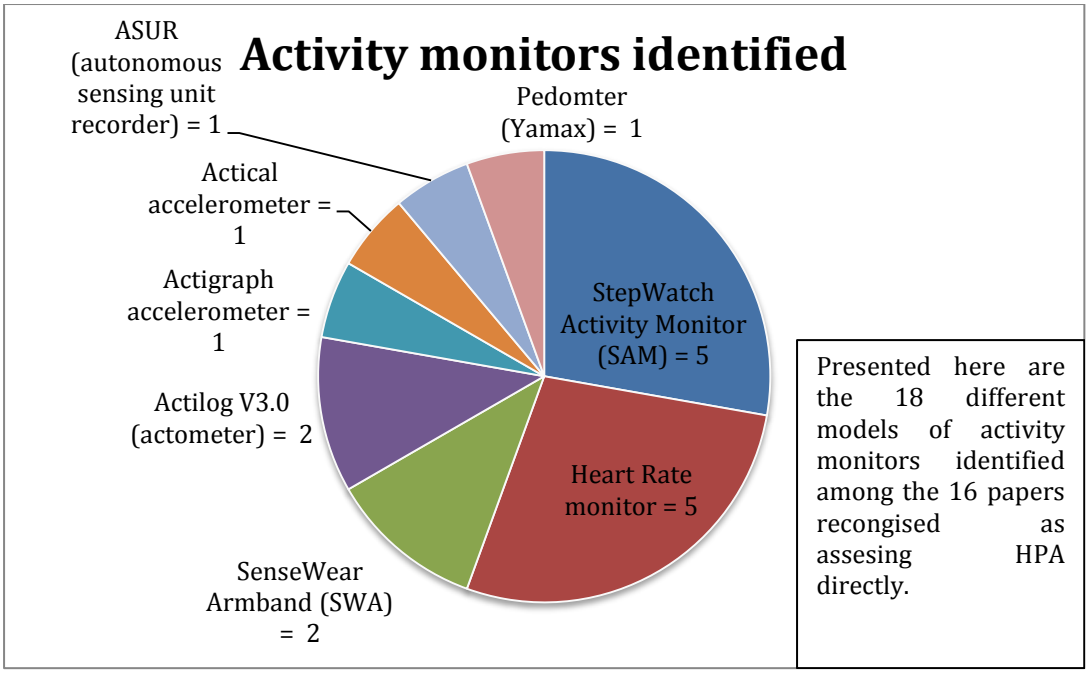


Figure 4. 1 Activity monitors (by models) identified in the systematic search

Over the last decade, there has been a notable increase in the use of accelerometry-based HPA studies. At the time of this review, the most popular activity monitor used in neuromuscular diseases was the StepWatch Activity Monitor (SAM), reported in five out of the 16 papers followed by heart rate monitors (Figure 4. 1). The time period of activity monitoring ranged from two to twelve days with three as the most common. The protocol of 'during waking hours' was more common than the 24-hours wear-time rule (Figure 4. 2). Four studies did not provide full details of the model used or the location where the device was placed (168, 173, 174, 179) and the full data collection protocols (data criteria for analysis, definition of non-wear episode and the processing of missing data) were only identifiable in three of the papers (170, 172, 181). Four papers reported the use of a patient-reported outcome (i.e. activity diary) as a quality control method (173, 178, 179, 184). McDonald et al. was the only study assessing issues of feasibility, with an explanation of lost data from heart rate monitors due placement issues (188).

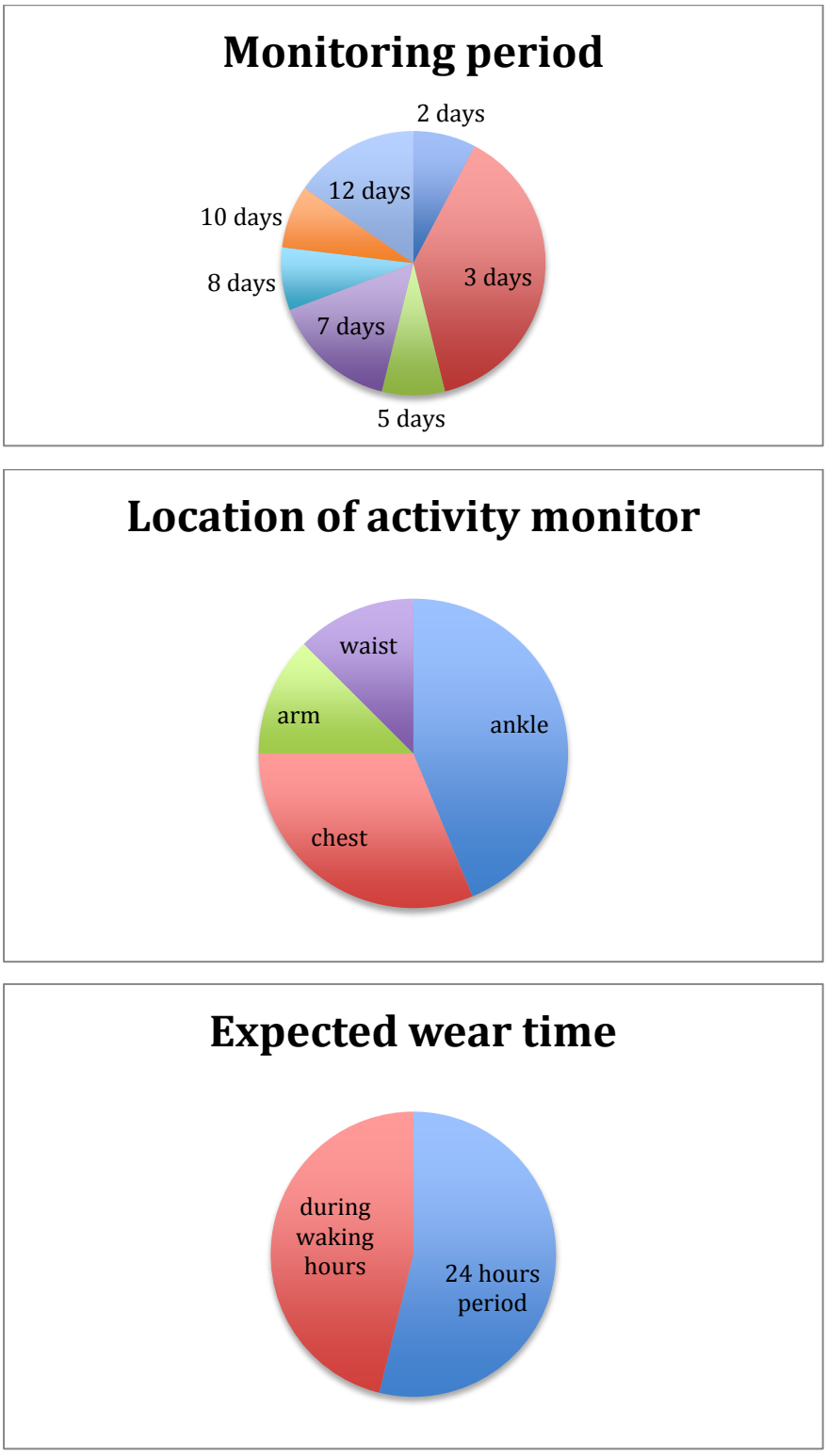


Figure 4. 2 Monitoring protocols reported when assessing habitual physical activity with activity monitors: monitoring periods, body location and expected wearing (active monitoring) time; identified in the systematic search.

4.3.2. Part II: Validity of an accelerometry-based device and site of placement in DM1, a cross-sectional study.

4.3.2.1. Accelerometers' validity

Data from 30 DM1 participants and from 14 healthy volunteers were cleaned and collected. Data from 2 accelerometers from the healthy group were lost due to machine error. Table 4. 1 presents each group's demographics. There was a significant age difference between groups. As a whole sample, with the exception of demographic values and the accelerometer output data from ankle-worn accelerometers during the six-minutes walking test (6MWT) and the left ankle ten-meters walk test (10-mWT) outputs, all other accelerometer variables showed a not-normal distribution. When normality tests were done individually per sample, the healthy control group showed a normal distribution in all variables except three: right wrist ten-meter walk/run test (10-mW/RT) and both ankle reports when standing still. On the other hand, the DM1 group only remained normally distributed on the demographic variables and on the reports of the right ankle 10-mW/RT.

Table 4. 1 Characteristics of the sample groups: DM1 patients and healthy-control volunteers.

	DM1	**Healthy-controls**
No. (No of males)	30 (20 m)	14 (6 m)
Age – years (min-max)	47.8 (25-72)	32.3 (23-47)*
Height – cm (SD)	171 (7.9)	167.7 (11.2)
BMI (SD)	25.3 (4.8)	24.1 (3.9)

* Difference between groups is statistically significant

The mean acceleration values for each of the performed tests and the standing-still position are presented for each group in Table 4. 2. These values come from the average-per-minute estimations performed for each different test. All tests showed significant differences between groups.

Table 4. 2 Mean acceleration values for each of the performed tests and standing-still position for each group.

Test	Location	DM1		Healthy controls		*p* value
		Mean (SD)	95% CI	Mean (SD)	95% CI	
6MWT	Left ankle	47.6 (16.5)	41.5-53.7	74.4 (15.5)	65.1-83.8	< .001
	Right ankle	47.8 (16.6)	41.6-54	75.5 (14.4)	66.4-84.7	< .001
	Left wrist	21.3 (15.6)	15.5-27.1	40.6 (20.7)	27.5-53.7	< .001
	Right wrist	20.3 (13.8)	15.2-25.5	41.9 (24.5)	26.3-57.4	.002
10-mWT	Left ankle	37.4 (9.5)	33.8-41	46.0 (6.7)	41.8-50.2	.001
	Right ankle	37.6 (8.8)	34.2-40.9	48.1 (9.7)	41.9-54.3	.001
	Left wrist	13.2 (3.7)	11.8-14.6	17 (5.6)	13.5-20.5	.031^
	Right wrist	12.1 (3.2)	10.9-13.3	18.1 (8.1)	13.0-23.2	.001
10-mW/RT	Left ankle	101.6 (50)	82.1-121	183.6 (36)	160.9-206	< .001
	Right ankle	98.1 (44.3)	80.9-115	176.5 (32)	155-198	< .001
	Left wrist	99.3 (14.5)	66.8-132	210.6 (35)	189.5-232	< .001
	Right wrist	94 (80.5)	62.8-125	214.9 (42)	188-241.9	< .001
Stand Still	Left ankle	1.2 (1.3)	0.7-1.7	1.9 (1.9)	0.7-3.1	*ns*
	Right ankle	1.3 (1.2)	0.8-1.7	1.7 (1.7)	0.6-2.7	*ns*
	Left wrist	2.3 (2.3)	1.5-3.2	2.7 (1.9)	1.5-4.0	*ns*
	Right wrist	2.3 (1.6)	1.7-3.0	2.5 (1.5)	1.4-3.4	*ns*

Mean (SD) and 95% Confidence intervals of the accelerometer output (mg) recorded by each device during each activity performed and compared between groups. *ns* not statically significant. 6MWT (six minutes walking test), 10-mWT (10 meters walk test), 10-mW/RT (10 meters walk/run test). ^when adjusted to age and gender the significance level changed to .08 and estimated marginal means of 16.7 for healthy controls and 13.4 for the DM1 group.

Data from each joint was analysed independently to compare the differences from test to test. The mean differences from test (or activity) to test were all significant for all tested accelerometer placement sites. However, the distinction between the results from the two walking tasks 10-mWT (comfortable speed) and 6MWT (sub-maximal speed) was clearer on the Healthy Control group with no overlap of the percentile estimations. When dividing DM1 into classic and mild phenotypes, the significance between the mean values of the healthy and the mild phenotype groups disappears. Figure 4. 3 exemplifies the 6MWT and the 10m-W/RT cases for each joint. On these graphs there is a clear difference between the acceleration levels reached by the ankle and wrist when walking (6MWT) and

how these change when speed is increased to running (10-mW/RT). This order swap between wrist and ankle was not the case on the classic phenotype; in this sub-group, ankles remained the fastest joint even when running.

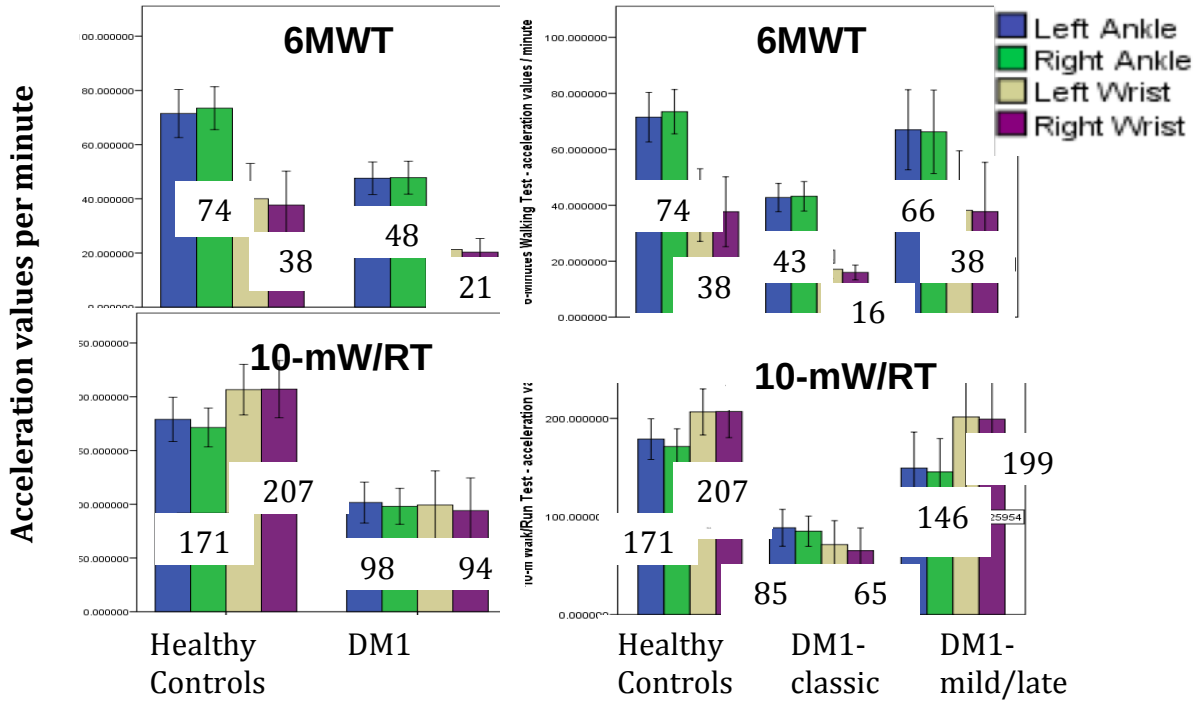


Figure 4. 3 Bar charts: accelerometry values per minute reached on average by each joint and compared between the groups.

Figure 4.3 Bar chart with standard error bars representing the accelerometry values per minute reached on average by each joint and compared between the groups: 1) left side: healthy controls vs. DM1 patients; and 2) right side: healthy controls vs. DM1 patients subdivided into classic and late (mild) onset. Tests exemplified here correspond to the 6MWT (six minutes walking test) graphs above and the 10-mRT (10 meters walk/run test) on the graphs below. Values on y-axis represent the average of acceleration counts estimated per minute for each activity.

4.3.2.2. Inter-accelerometer reliability

The intra-class correlation (ICC) of acceleration per second was also estimated from the data obtained at each second of the 6MWT and from the per-minute estimations for each one of the functional tests. ICCs ranged from .44 (6MWT at every second) to .97 (6MWT per minute). The lowest ICC values were for the 6MWT per second scores and the widest 95% CI were for the activity of standing still (Table 4. 3).

Table 4. 3 Intraclass Correlation Coefficient (ICC) test between accelerometers at each performed test.

Activity	Group	ICC	95% CI
6MWT per second	Healthy Control	0.440	0.28-0.56
	DM1	0.510	0.43-0.6
6MWT per minute	Healthy Control	0.940	0.85-0.98
	DM1	0.972	0.95-0.99
10-mWT	Healthy Control	0.848	0.6-0.96
	DM1	0.854	0.74-0.93
10-mW/RT	Healthy Control	0.762	0.38-0.93
	DM1	0.960	0.93-0.98
Standing Still	Healthy Control	0.726	0.24-0.93
	DM1	0.504	0.13-0.74

Intraclass Correlation Coefficient (ICC) between accelerometers at each of the performed tests for each group. 6MWT (six minutes walking test), 10-mWT (10 meters walk test), 10-mW/RT (10 meters walk/run test).

4.3.2.3. Inter-accelerometer (joint-placement position) reliability

Table 4. 4 presents the Bland-Altman Plot estimations obtained from the difference in mean values of one accelerometer to another and the mean and standard error between both. One sample T-test showed significant differences ($p < 0.05$) between wrist and ankle accelerometers in all tests except for running; these come together with a strong linear regression score. Wrist and wrist results for the 10-mWT were also statically significant but these do not have a significant coefficient of regression. Bland-Altman Plots present a Y-axis representing the difference from the mean estimated value (red line) and an X-axis with the mean acceleration values spread within the 95%CI (blue lines). Figure 4.4 represent four different Bland-Altman plots with different distributions as an example. Two of these compare the same test (6MWT) between wrist-wrist results and ankle-ankle results. And two compare wrist vs. ankle for the walking test and the walk/run results, this latter shows a better spread distribution of plots along the estimated mean. Table 4. 4 also presents the correlation values and significance between each of the accelerometer's data.

Table 4. 4 Inter-accelerometer (joint placement position) reliability testing.

			Bland-Altman Plot estimations				Correlation	
Activity	**Joints**	**Group**	Difference in Mean	*Sig.*	95% Interval of agreement	**Coefficient of regression**	r	*Sig.*
6MWT	Wrist-Ankle	Healthy Control	31.8	<.05 *	23.1-40.5	<0.001	0.83	<.001
		DM1	26.3	<.05 *	23.2-29.4	<0.001	0.85	<.001
		All	27.8	<.05 *	24.7-31.1	<0.001	0.91	<.001
	Wrist-Wrist	Healthy Control	2.4	0.4	(-)16.5-21.2	ns	0.93	<.001
		DM1	0.99	0.2	(-)7.3-9.2	ns	0.87	<.001
		All	1.4	0.2	(-)10.5-13.2	ns	0.95	<.001
	Ankle-Ankle	Healthy Control	(-)1.84	0.1	(-)9.7 - 6	ns	0.92	<.001
		DM1	(-)0.2	0.7	(-)6-5.6	ns	0.98	<.001
		All	(-)0.67	0.2	(-)7.2-5.8	ns	0.99	<.001
10m-WT	Wrist-Ankle	Healthy Control	28.3	<.05 *	20.8-36.7	<0.001	0.65	<.001
		DM1	24.2	<.05 *	10.5-37.9	<0.001	0.57	0.001

		All	23.4	<.05 *	12.6-38.1	<0.001	0.66	<.001
	Wrist-Wrist	Healthy Control	(-)0.6	<.05 *	(-)9.1-7	ns	0.79	<.05
		DM1	(-)1.1	<.05 *	(-)5.7-3.4	ns	0.67	<.001
		All	(-)0.97	<.05 *	(-)6.8-4.8	ns	0.76	<.001
	Ankle-Ankle	Healthy Control	(-)1.2	-0.2	(-)15.4-12	ns	0.53	ns
		DM1	(-)0.2	0.8	(-)8-7.6	ns	0.83	<.001
		All	(-)0.46	0.6	(-)10.3-9.4	ns	0.86	<.001
10-mW/RT	Wrist-Ankle	Healthy Control	(-)24.6	0.06	(-)105.3-56.1	ns	0.24	ns
		DM1	2.3	0.8	(-)81.6-86.2	ns	0.85	<.001
		All	(-)5.8	0.4	(-)91.3-79.7	ns	0.5	<.001
	Wrist-Wrist	Healthy Control	(-)6.7	0.5	(-)64.6-51.2	ns	0.64	0.03
		DM1	5.3	0.2	(-)39.7-50.2	ns	0.92	<.001
		All	1.7	0.7	(-)48-51.3	ns	0.76	<.001
	Ankle-Ankle	Healthy Control	7.5	0.3	(-)35.1-50.1	ns	0.61	ns
		DM1	3.5	0.2	(-)22.3-29.4	ns	0.98	<.001
		All	4.6	0.1	(-)26.1-35.2	ns	0.6	<.001

Standing Still	Wrist-Ankle	Healthy Control	(-)1.4	<.05 *	(-)5.7-2.8	ns	0.27	ns
		DM1	(-)1.2	<.05 *	(-)5.8-3.5	0.01	0.5	0.005
		All	(-)1.2	<.05 *	(-)5.7-3.3	<0.001	0.2	ns
	Wrist-Wrist	Healthy Control	0.3	0.99	(-)4-4.6	ns	0.55	0.09
		DM1	(-)0.003	0.99	(-)5.3-5.3	ns	0.43	0.02
		All	0.08	0.9	(-)4.9-5.1	0.04	0.1	ns
	Ankle-Ankle	Healthy Control	0.2	0.5	(-)3.8-4.1	ns	0.07	0.8
		DM1	(-)0.1	0.5	(-)1.5-1.3	ns	0.58	0.001
		All	(-)0.02	0.9	(-)2.4-2.3	ns	0.7	<.001

The first section of this table presents the Bland–Altman plot estimations assessing the level of agreement between accelerometer outputs (ENMO-mg) obtained from two different monitors either placed on the wrist(s) or placed on the ankle(s). *When the differences between the mean are statistically significant, a Bland-Altman plot is not valid as there is no agreement between these two sets of data. The second section of this table presents the coefficient of regression estimated between the difference between the means and mean estimated value, when this value is statically significant (<0.05) this indicates a distribution with a trend that does not support reliability between devices. Finally, the third section of the table presents the Spearman correlation (rho) results. ns = not statistically significant.

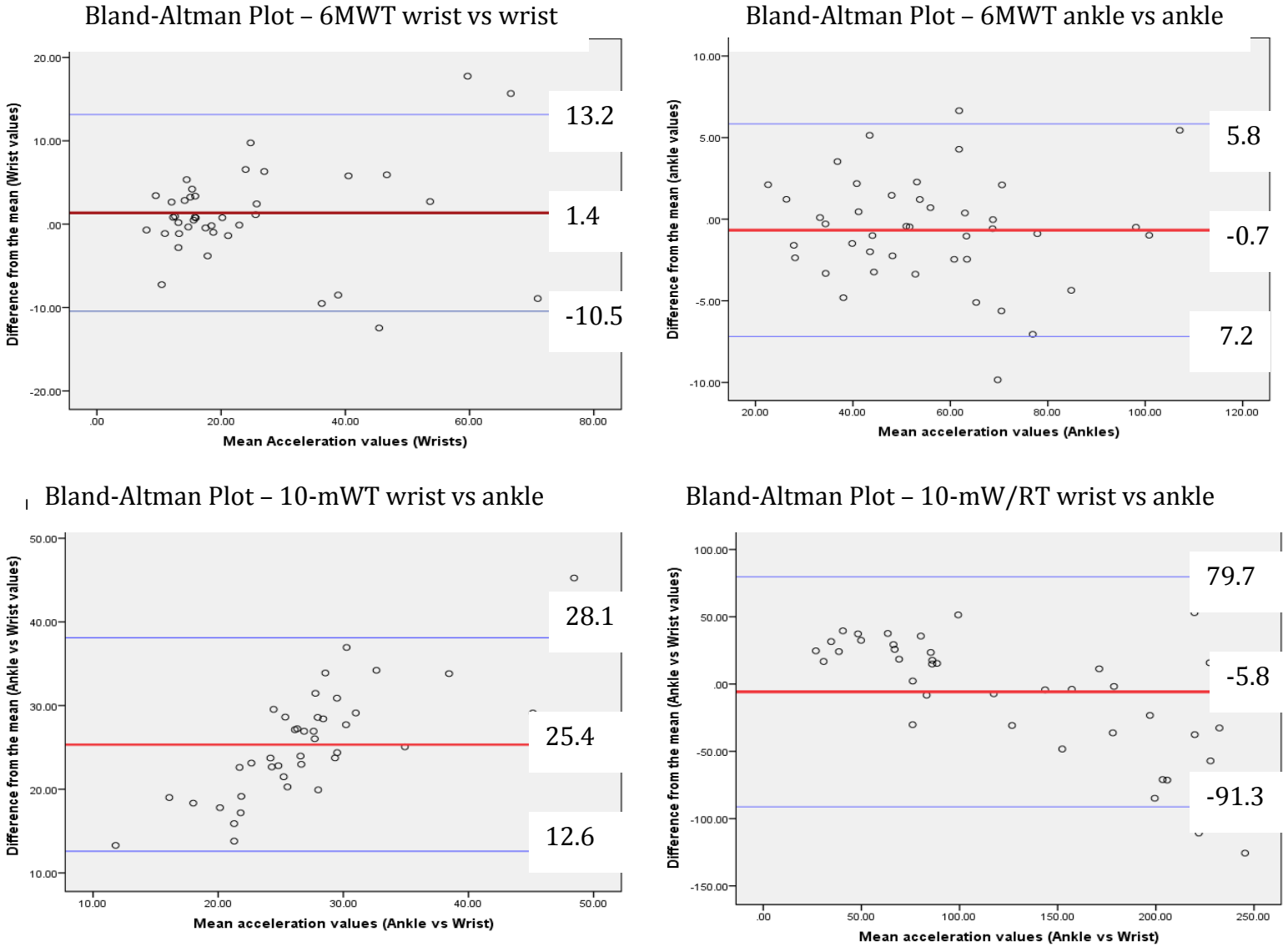


Figure 4. 4 Four different examples of Bland-Altman Plots.

Figure 4.4. Bland–Altman plots representing the agreement between accelerometer outputs (ENMO-mg) from two different monitors either placed on the wrist(s) or placed on the ankle(s). The red line represents the mean value obtained from the difference between each accelerometer's set of data and the second one. The blue lines represent the 95% limits of agreement (±1.96 SD) around this difference from the mean. 1) Plots above present the results from the 6MWT plotting, with left wrist vs. right wrist on the left hand side and left ankle vs. right ankle on the right hand side; all these have an even spread within the intervals of agreement and with no visible linear trend. The plots of the ankle vs. ankle plot maintains a similar distribution all along the graph, showing no visible impact of the increment of speed detected by the accelerometers. 2) Plots below present the results from computing ankle vs. wrist in two different tests, the 10-mWalkT on the left hand side and the 10-mWalk/RunT on the right hand side; both show a visible linear trend of distribution along the graph but only the 10-mWalkT regression coefficient was statistically significant (<0.05).

4.3.3. Part III: Habitual Physical Activity (HPA) in DM1, a longitudinal study.

4.3.3.1. Cross-sectional analysis

Data from 91 participants, belonging to one of the three groups: healthy volunteers (N=19), chronic fatigue syndrome (CFS) group (N= 12) and severely fatigued DM1 patients (N=60), were included in this analysis. **Table 4. 5** presents and compares the three cohorts selected for HPA analysis in severely fatigued populations such as DM1. There was an initial 10% to 12% of data lost from the accelerometers registered for each group, either to lost data, problems when setting up the device, not enough monitoring time completed or devices lost/not returned; only those with complete information at the baseline were included for analysis. All variables except age had a non-normal distribution hence these were always considered for non-parametric tests. These populations were non-matched and the DM1 cohort mean age was older than the healthy volunteers. Both fatigued populations (CFS and DM1-fatigued participants) differ significantly in all outcomes from the healthy volunteers with the exception of the average acceleration values of the least active 5 hours of the day (L5-ENMO) and the activity spread (SD) over the first 5 consecutive days of monitoring. According to the CIS reported outcome, the DM1 and CFS cohorts reported significantly lower activity levels (i.e. higher scores on the reduced physical activity CIS subscale) than the healthy controls. The CFS cohort reported the highest scores (i.e. higher levels) for perceived fatigue and concentration disturbances. According to the accelerometer results, DM1 demonstrated the lowest objectively measured physical activity levels. The CFS group had an average of 62(±8) counts (i.e. accelerometry counts per minute) when only considering the period awake, and the DM1 group had only 51(±21) counts.

To confirm the reliability of the GENEActiv at daily-life bases, this was compared to a previously validated actometer (or accelerometer) used as a standard at the Radboud fatigue clinics when screening patients (the white actometer). Reports from both accelerometers (or actometers) correlated significantly with each other (Table 4. 6). The CFS group only showed significant correlation values between the white

actometer daily outputs and the reports of the most active 5 hours (M5 ENMO) from the black accelerometer (GENEActiv). Bland-Altman plotting was computed to analyse the intra-accelerometry reliability but the difference identified between the means was statistically significant and with a non-significant regression coefficient value (p = 0.084). Accelerometer reports, correlated to the CIS-reported outcomes, were estimated but due to the sample size, the significance in correlation was only detectable for the DM1 cohort and for the whole sample as a cohort. There was a strong, significant and consistent correlation between both accelerometers (black and white) and the CIS-reported reduced fatigue levels (Table 4. 7). Patient-reported outcomes correlated significantly to the accelerometer reports of activity (except for the L5 ENMO values) with strong values (r >0.45) for the CIS-reported activity levels and the MIRS and DM1-ActivC disease severity outcomes (Table 4. 8).

Table 4. 5 Cross-sectional study sample demographics presenting objectively measured activity levels (accelerometer reports) and the patient-reported outcome Checklist of Individual Strength (CIS) results.

	Healthy Controls		**CFS**			**DM1**			
Sample (N)	19		12			60			
Gender (% males)	37%		42%			62%			
Variable	**mean (SD)**	**95% CI**	**mean (SD)**	**95% CI**	sig. vs Healthy	**mean (SD)**	**95% CI**	sig. vs Healthy	sig. vs CFS
Age	34.5 (10.7)	29.4-39.7	42.2 (13)	33.9-50.5	ns	47.3 (10.3)	44.6-49.9	<0.001	ns
CIS-perceived fatigue	20.4 (7.4)	16.8-24	51.4 (30.1)	49.5-53.4	<0.001	43.9 (6.2)	42.3-45.6	<0.001	<0.001
CIS-concentration problems	11.6 (6.1)	8.7-14.5	30.8 (6.6)	26.6-34	<0.001	18.6 (7.3)	16.7-20.5	<0.001	<0.001
CIS-reduced motivation	9.4 (3.2)	7.8-10.9	18 (5.1)	14.8-21.2	<0.001	17.9 (4.6)	16.7-19.1	<0.001	ns
CIS-reduced physical activity	9.4 (4.9)	7-11.7	15.7 (3.5)	13.4-17.9	<0.001	16.9 (4.6)	15.7-18.1	<0.001	ns
Overall PA levels measured by ActivLog (Counts over 24 hours)	N/A		180.6 (35.9)	157.8-203.4	N/A	138.5 (56.1)	111.5-165.5	N/A	ns^
Overall PA levels measured by GENEActiv (ENMO over 24 hours)	44.8 (12.7)	38.7-51	28.5 (8.3)	23.2-33.7	<0.001	20.9 (11.5)	12-23.9	<0.000	0.03
Most active 5 hours of the day (GENEActiv - ENMO)	118.1 (37)	100.2-135.9	61.5 (18.2)	50-73.1	<0.001	53.5 (34.8)	44.5-62.5	<0.001	0.02
Least active 5 hours of the day (GENEActiv - ENMO)	4 (1.6)	3.3-4.8	4.2 (2)	2.8-5.7	ns	4 (1.1)	3.7-4.3	ns	ns
SD over 5 consecutive days	7.8 (8.8)	3.3-12.3	7.5 (3.5)	5.3-9.7	ns	8.7 (9.6)	6.1-11.1	ns	ns

Demographics of the three groups involved in the cross sectional study. CFS = Chronic Fatigue Syndrome; DM1 = myotonic dystrophy 1; CIS = Checklist of Individual Strength; PA = physical activity; ENMO = the Euclidean Norm Minus One (mg). ^when corrected to age significance value changed to a borderline significant 0.052 and estimated marginal means of 178.4 for the CFS group and 139.5 for the DM1 group.

Table 4. 6 Comparison between two different accelerometer models.

	GENEActiv (*black accelerometer*)				
	ENMO 24hrs		**M5 ENMO**		
	Spearman's r rho	Sig.	Spearman's r rho	Sig.	Group
ActivLog (*white accelerometer*)	ns		0.83	0.001	**CFS**
	0.9	<0.001	0.97	<0.001	**DM1**
	0.83	<0.001	0.84	<0.001	**Whole Sample**

ENMO: Euclidean norm minus one, M5 ENMO: most active five hours of the day, CFS: chronic fatigue syndrome, DM1: Myotonic Dystrophy type 1.

Table 4. 7 Correlation values between objectively measured activity levels (accelerometer reports) and the patient-reported outcome Checklist of Individual Strength (CIS).

		HPA levels measured with ActivLog	**HPA levels measured with GENEActiv**	**M5 ENMO**
Checklist Individual Scale (CIS)		Spearman's r rho		
DM1	CIS – perceived fatigue severity score	ns	(-)0.28*	(-)0.3*
	CIS – concentration problems score	ns	0.36*	0.31*
	CIS – reduced motivation score	ns	(-)0.36*	(-)0.34**
	CIS – reduced physical activity score	(-)0.58**	(-)0.55***	(-)0.56***
WHOLE SAMPLE	*CIS – perceived fatigue severity score*	ns	(-)0.49***	(-)0.53***
	CIS – concentration problems score	0.41*	ns	ns
	CIS – reduced motivation score	ns	(-)0.54***	(-)0.54***
	CIS – reduced physical activity score	(-)9.45*	(-)0.64***	(-)0.64***

CIS: checklist of individual strength questionnaire (presented by domains), HPA: habitual physical activity, ENMO: Euclidean norm minus one, M5 ENMO: most active

5 hours of the day (average over seven days). *sig.≤ 0.05 ** sig.≤0.01 ***sig.≤0.001 ns: not-statistical significant.

4.3.3.2. Longitudinal analysis

Sixty severely fatigued DM1 patients during follow-up examinations were included in this analysis, 56 completed the first 10 months of the overall study which represents the active phase of the study and 51 completed the whole study. Age was the only variable normally distributed. Thirty-three participants of this sample (66%) belong to the group allocated to receive cognitive-based therapy (CBT) as part of the overall study they belong to (OPTIMISTIC study). Hence, all estimations were done for the whole sample as one cohort and divided, based on the allocated study arm. Sample demographics in detail are presented Table 4. 9. The average of the five least active hours of each day (L5 ENMO) was the only variable differing between groups at the baseline.

When analysed over time CIS-perceived fatigue showed an improvement in both groups at month 10 and month 16 after randomization, but with a larger change in the CBT group and when samples were compared with each other the mean was significantly higher in the CBT group at each follow-up visit. The CBT group showed a consistently significant change in other reported outcomes such as the MDHI score and CIS – motivation and activity level scores. Accelerometry reports showed a significant difference between groups in habitual physical activity (HPA) levels (22.6±8.6 vs. 15±5.1) and on the M5 ENMO (61.1±27.1 vs. 37.6±15.1) 10 months after randomization, which represents the end of the active intervention. However, these values do not differ significantly from the baseline for either of the groups (Table 4. 10).

Classic and mild phenotypes as defined previously in methodology were analysed as subgroups showing a constant difference between groups at each time point, with the exception of the M5 ENMO value at 10 months after

randomization. There were no significant changes between the baseline and any of the follow-ups in either of the groups (Table 4. 11).

Based on the standard deviation detected per participant over the data obtained per day for seven consecutive days, SEM and SEM% were estimated. The SEM (and SEM%) estimated for the average HPA values were 2.7 (ENMO-mg) or 13% of the overall HPA and 8.4 (ENMO-mg) or 17% of the most active five hours of the day (M5 ENMO).

Table 4. 8 Correlation values between objectively measured activity levels (accelerometer reports) and outcomes of disease-severity.

Disease Severity Scale	HPA levels	M5 ENMO	L5 ENMO
DM1-ActivC	0.49**	0.52**	*ns*
MDHI	(-)0.43**	(-)0.42**	*ns*
MIRS	(-)0.51**	(-)0.47**	*ns*
CIS – perceived fatigue	(-)0.28*	(-)0.3*	*ns*
CIS – concentration problems	0.36**	0.31*	*ns*
CIS – reduced motivation	(-)0.36**	(-)0.34**	*ns*
CIS – reduced physical activity	(-)0.55**	(-)0.56**	*ns*

* p < 0.05 / ** p <0.01 / *ns* not-significant

MDHI Myotonic Dystrophy Health Index / MIRS Muscular Impairment Rating Scale / CIS Checklist of Individual Strength

Table 4. 9 Longitudinal study sample demographics.

	Whole Sample			Care-as-usual Group			Intervention (CBT) Group			
				Sample (N)			27			33
				Gender (% males)			67%			58%
				MIRS			1 = 0			1 = 2 (6%)
							2 = 9			
							(33%)			2 = 8 (24%)
							3 = 9			3 = 13
							(33%)			(39%)
							4 = 9			4 = 10
							(33%)			(30%)
				Phenotype (% mild)			22%			21%
Variable	**mean (SD)**	**95% CI**	**min-max**	**mean (SD)**	**95% CI**	**min-max**	**mean (SD)**	**95% CI**	**min-max**	**sig. between groups**
Age	47.3 (10.3)	44.6-50	19-69	48.6 (10)	44.7-52.6	31-69	46.2 (10.5)	42.2-49.9	19-63	*ns*
DM1-ActiC centile score	63.7 (18)	59-68.3	31-100	62.3 (17.5)	55.4-69.3	39-100	64.7 (18.5)	58.2-71.3	31-100	*ns*
MDHI score	33.3 (18.3)	28.6-30	2-68.1	35.2 (16.4)	28.7-41.7	6.3-64.6	31.7 (19.8)	24.7-38.7	2-68.1	*ns*
CIS – perceived fatigue	44 (6.3)	42.3-45.6	35-56	44.5 (6.5)	41.9 -47	35-56	43.6 (6.1)	41.4-45.7	15.4-20.4	*ns*

CIS – concentration problems	18.6 (7.3)	16.7-20.5	5-35	19.4 (7.6)	16.4-22.4	5-35	17.9 (7.1)	15.4-20.4	5-29	*ns*
CIS – reduced motivation	17.9 (4.6)	16.7-19.1	7-26	18 (4.5)	16.3-19.8	9-25	17.8 (4.8)	16.1-19.5	7-26	*ns*
CIS – reduced physical activity	16.9 (4.6)	15.7-18.1	3-21	16.6 (5.4)	14.4-18.7	3-21	17.2 (3.9)	15.8-18.5	7-21	*ns*
HPA activity levels (average over 7 days)	21 (11.5)	18-24	6.1-75	21 (14.4)	15.3-26.6	6.5-75	21 (8.8)	17.8-24.1	6.1-37.2	*ns*
M5 ENMO (average of the most active 5 hours of each day)	53.5 (34.8)	44.5-62.5	11.8-225.4	52.2 (43.4)	35-69.4	12.2-225	54.5 (26.5)	45-64	12-101.6	*ns*
L5 ENMO (average of the least active 5 hours of each day)	4 (1.1)	3.7-4.3	1.9-9.6	4.4 (1.4)	3.9-5	2.8-9.6	3.6 (.63)	3.4-3.8	1.9-5	0.016

MIRS: muscular impairment rating scale, CIS: checklist individual scale, HPA: habitual physical activity, ENMO: Euclidean norm minus one (mg), M5 ENMO: most active 5 hours of the day, L5 ENMO: least active 5 hours of the day, SD: standard deviation, 95% CI 95% confidence interval, ns: non-statically significant.

Table 4. 10 Groups' progression over time.

Care-as-usual Group (N = 27)	Mean change over 5 months (N = 25)	SE mean	Sig. (2-tailed)	Mean change over 10 months (N = 26)	SE mean	Sig (2-tailed)	Mean change over 16 months (N = 23)	SE mean	Sig. (2-tailed)
DM1-ActiC centile score	+0.3	2.0	*ns*	-0.2	1.9	*ns*	-0.3	2.2	*ns*
MDHI score	-5.4	3.2	*ns*	-2.4	2.5	*ns*	-3.3	2.8	*ns*
CIS – perceived fatigue	-1.9 *	1.3	*ns*	-3.0 *	1.2	0.02	-3.4 *	1.6	0.04
CIS – concentration problems	+0.2	1.4	*ns*	+0.4	1.5	*ns*	+0.6	1.7	*ns*
CIS – reduced motivation	-0.2 *	0.9	*ns*	-0.6 *	1.1	*ns*	-1.4	1.2	*ns*
CIS – reduced physical activity	-0.1 *	0.8	*ns*	-0.5 *	0.9	*ns*	-0.2 *	0.9	*ns*
HPA activity levels (average over 7 days)	-1.2	1.1	*ns*	-1.4 *	1.3	*ns*	-2.5	1.4	*ns*
M5 ENMO (average of the most active 5 hours of each day)	-0.0	3.3	*ns*	-1.4 *	3.8	*ns*	-6.3	4.5	*ns*
L5 ENMO (average of the least active 5 hours of each day)	-0.6	0.3	*ns*	-0.6	0.3	*ns*	-0.7	0.4	*ns*

Intervention group (Cognitive Behavioural Therapy- CBT) (N = 33)	Mean change over 5 months (N=32)	SE mean	Sig. (2-tailed)	Mean change over 10 months (N = 30)	SE mean	Sig (2-tailed)	Mean change over 16 months (N = 28)	SE mean	Sig. (2-tailed)
DM1-ActiC centile score	+4.8	2.1	0.03	+2.7	1.9	*ns*	+3.0	2.4	*ns*
MDHI score	-7.7	2.5	0.00	-7.2	2.1	0.002	-8.5	2.7	0.004
CIS – perceived fatigue	-11.6 *	1.7	0.00	-10.5 *	1.8	0.00	-10.6 *	1.8	0.000
CIS – concentration problems	-1.6	1.0	*ns*	-1.1	1.0	*ns*	-1.1	1.4	*ns*
CIS – reduced motivation	-5.0 *	0.9	0.00	-3.5 *	0.8	0.00	-3.1	1.3	0.028
CIS – reduced physical activity	-5.6 *	0.9	0.00	-4.7 *	0.7	0.00	-4.6 *	0.8	0.000
HPA activity levels (average over 7 days)	+0.2	1.3	*ns*	+1.3 *	1.2	*ns*	+0.2	1.9	*ns*
M5 ENMO (average of the most active 5 hours of each day)	+2.2	3.9	*ns*	+5.4 *	4.1	*ns*	+1.1	5.9	*ns*
L5 ENMO (average of the least active 5 hours of each day)	+0.1	0.2	*ns*	+0.1	0.1	*ns*	-0.1	0.2	*ns*

MIRS: muscular impairment rating scale, CIS: checklist individual scale, HPA: habitual physical activity, ENMO: Euclidean norm minus one (mg), M5 ENMO: most active 5 hours of the day, L5 ENMO: least active 5 hours of the day, SD: standard deviation, 95% CI 95% confidence interval, ns: non-statically significant. * Significant difference between groups at this point (Mann-Whitney U Test).

Table 4. 11 Sample comparing classic and mild (late onset) phenotype subgroups at baseline and progression over time.

Time Point	GENEActiv report	Classic Mean (SD)	Mild Mean (SD)	sig. between groups
Baseline	HPA 24 hours	18.5 (9.4)	31 (4.3)	0.008
	M5 ENMO	45.8 (26.8)	81.2 (11.4)	0.002
	L5 ENMO	4 (1.3)	3.8 (.3)	*ns*
5 months	HPA 24 hours	18.8 (10.6)	26.2 (4.3)	0.026
	M5 ENMO	48.6 (32.7)	71.9 (15.2)	0.015
	L5 ENMO	3.8 (.9)	3.6 (.5)	*ns*
	sig. from baseline	*ns*	*ns*	
10 month	HPA 24 hours	18.9 (8.1)	26 (7.8)	0.031
	M5 ENMO	50.2 (25.4)	68.3 (26.4)	*ns*
	L5 ENMO	3.7 (.7)	3.9 (.4)	*ns*
	sig. from baseline	*ns*	*Ns*	
16 months	HPA 24 hours	16.6 (7.1)	33.2 (17.8)	0.002
	M5 ENMO	41.8 (20.5)	87.9 (48.7)	0.004
	L5 ENMO	3.6 (.5)	3.6 (.5)	*ns*
	sig. from baseline	*ns*	*Ns*	

HPA: habitual physical activity (24 hours average ENMO-mg over 7 consecutive days), M5 ENMO: most active five hours of the day (average ENMO over 7 consecutive days), L5 ENMO: least active five hours of the day (average ENMO over 7 consecutive days).

5

4.4. DISCUSSION

The systematic review revealed a clear increasing trend to utilize activity monitors in all forms of neuromuscular diseases ranging from more basic pedometers to

more sophisticated tri-axial accelerometers. However, when compared to other diseases with motor impairment, activity monitoring in neuromuscular disorders is still at an early stage of implementation and understanding. There is a clear need to encourage the introduction of these novel tools into research and clinical practice and to increase the awareness among health practitioners and researchers of their clinical usefulness (52). This could be improved by a conscious use of the existing tools by researchers and an appropriate selection of monitoring methodologies when implementing them to neuromuscular disorders. Byrom et al. (160) and Dilon et al. (189) proposed certain criteria to increase the level of confidence when assessing HPA with activity monitors by suggesting the use of triaxial accelerometry (like ActiGraph, RT3 and Actical) for at least 6 consecutive days of suggested 'wear time' and for at least 10 hours per day which was only followed in three of the identified studies (177, 183, 184). Common placement sites have been waist and hip (160). However distal joints have shown better compliance by the participants (160, 190). There was a clear lack of validity and reliability testing among the identified studies, which emphasizes the need to understand this technology better and to invest more in the investigation of its use and applicability in neuromuscular diseases. We suggest including the use of activity monitors in DM1 in future discussions related to the selection and standardisation of outcome measures in DM1 such as the OMMYD and the TREAT-NMD meetings (27, 144). Conclusions from this systematic review about an activity monitor of preference are not possible from this review as these were not compared with each other and only a few details regarding logistical issues were reported. This would require extensive reliability and validity work in the target population, which was not the aim to the literature review. Still, similar reviews have concluded that tri-axial accelerometry is feasible, and where raw data are accessible this should be investigated and adapted further for each cohort of interest (160, 191-193). In conclusion, there is a need to investigate this type of devices further and the pressure to identify feasible tools with valid outputs to measure physical activity parameters in DM1.

This study explored the validity of the tri-axial accelerometer (GENEActiv) in DM1 participants. GENEActiv showed potential when assessing different walking paces in this disease, showing sensitivity to differentiate the affected from non-

affected and a difference in range values between walking speeds, while there was no difference in the standing-still position. In the walking tests, the results from the wrist demonstrated closer or overlapping 95% CI between the DM1 participants and the healthy participants which was not the case for the ankle data. This might suggest the ankle as a better option to distinguish walking periods of a DM1 participant with a comfortable speed, detecting outputs between 33 and 41 ENMO/minute. The findings of this study were not comparable to previously published GeneActiv wrist activity thresholds as these include oxygen consumption on their final scores (154). Table 4.2 proves the concept that each device records higher acceleration values with each activity intensity increment and these values differ in slower walkers and runners (e.g. DM1) from faster ones. There is a significant impact on accelerometer location when translating raw data into meaningful outputs. Ankle reports differ significantly from wrist ones when walking and this difference disappears when increasing speed into running; however, this effect is less evident in the DM1 cohort that can be explained by an impaired running capability as shown before. Unfortunately, there is a lack of GeneActiv ankle-based accelerometer studies that could enrich these findings but it has already been suggested that ankle-worn devices may have the highest correlation to actual physical activity energy expenditure (194, 195). Still, a crucial need in this field is the identification of intensity thresholds that can differentiate the movements recorded as mild, moderate or vigorous activity-intensities.

ICC was estimated with a two-way random model which assumes that each participant was measured by each accelerometer at the same time, and that the used accelerometers properly represent a larger population of similar devices (60, 62). There was an excellent level of agreement within each accelerometer's measurements for all waking (per minute) scores (61). Kayes et al. (156) performed a similar study, exploring the Actical accelerometer in people with multiple sclerosis, also finding low ICC scores for sedentary activities and higher scores for more vigorous and rhythmic activities such as the 6MWT; we have to consider that these results come from a different monitor that uses different interpretation outputs and based algorithm. ICC decreased for the 6MWT-per-second estimations, which might be explained by the variation in body position

from second to second. This effect disappears when summarised to counts per minute.

A Bland-Altman plot presents the average bias (or average of the differences) between one accelerometer report and another, and the closer to zero the result is the more intra-accelerometer reliability there is. These plots reveal good intra-accelerometer reliability when comparing accelerometers of the same limbs between each other but not when comparing upper limb (wrist) to lower limb (ankle). This reliability reduces with the increment of speed (i.e. running test). This is a common case scenario with accelerometers as these devices can exhibit a phenomenon where a speed increment could either emphasize walking disturbances that reflect differences between the limbs or due to a frequency-dependent filtering effect of accelerometers (196-198). When wrist values are plotted against ankle values, these (i.e. wrist) revealed a strong linear trend with a widening difference at the lowest and highest extremes of acceleration (Figure 4.4). These findings complement the finding of higher mean values recorded by the ankle compared to the wrist when walking and strongly suggest that algorithms created for activity outputs validated for the wrist cannot be translated directly to the ankle when referring to ambulation or standing-still positions. The ankle has already been suggested before as the most appropriate site for placement when aiming to classify different speeds in activity, in particular to distinguish low-speed walkers (190, 199).

The use of accelerometers to assess HPA was shown to be feasible but with an issue of lost data or devices not returned and an average of only 80% of returned devices suitable for data extraction. In the OPTIMISTIC study the percentage of devices returned and suitable for data extraction (i.e. non-faulty) was 84% from the total of patients attending the randomization visit, 80% from visit three and 86% from visit four. After the cleaning process and excluding accelerometers with incomplete information or extracted data not fulfilling the pre-defined inclusion criteria only 79%, 75% and 78% (respectively) from the total sample were suitable for analysis. From all patients still enrolled at visit 4 (i.e. 10 months after randomization and visit established as the time point to measure the effect on the primary outcome), only 58% had complete data for analysis from each of the

visits. These numbers compare with other epidemiological reports, emphasizing the need to estimate data lost from these devices when declaring sample size and the importance of searching data quality from those not lost, which was done in this study (200, 201). The assessors' experience with the device software certainly facilitated the process but it was recognised that there was a need for a period to become familiar with the data-handling process.

Finally, this study presents for the first time objectively and in detail HPA levels of a fatigued DM1 population. This study allowed for comparison with other relevant cohorts such as the CFS group, a disease characterised by altered daily life activity patterns (202, 203). This CFS population had an average of 62 counts/day when awake (measured by ActivLog), which is close enough to the previously reported average of 60 counts/day that helped define the cut-off value for ActivLog to determine active days from passive days in this population and establish treatment accordingly (203). This validates the CFS sample (even when small) as a good representation of the disease regarding activity patterns. Though reporting less severe levels of fatigue than the CFS group, the DM1 group showed the lowest active HPA levels. Not all CIS scores showed a strong correlation with the activity patterns, but this questionnaire is thought to assess factors that impact on fatigue and not the other way around. However, the strong correlation between the reported reduced activity and the low objective HPA levels justify the conclusion that participants report their activity levels quite close to reality. This is not always the case in pervasively passive populations (204, 205).

There was no statistically significant change in HPA levels over time for either of the subgroups analysed but there is a clear suggestion of a trend of progression that differs between the group under an intervention and the standard care one. This trend is more drastic in the M5 ENMO values. Additionally, after 10 months of study follow-up, at the visit that corresponds to the primary endpoint of the overall OPTIMISTIC study and the time when the intervention would be finished, there was a significant difference between allocation groups and the constant significant difference between mild and classic was not detectable for the M5 ENMO values. This could be explained by a pick effect of the intervention but to

confirm this conclusion, ANCOVA estimation, controlling for treatment allocation, should be performed. However, testing the effect of the intervention was not in the scope of this thesis.

Mild participants' HPA levels differ significantly from the classic onset phenotype, not reaching the levels of our healthy volunteers' cohort but scoring somewhere between the CFS group and the healthy volunteers. As suggested before, when considering the 'ideal DM1 patient' for a clinical trial the significant differences between phenotypes should be considered for stratification or inclusion criteria, as these can considerably impact the outcome measure results (15, 158).

SEM% estimation was performed using reports of the care-as-usual (control) group at each time point. The results showed that if an intervention aims to reflect its efficacy on HPA levels measured with a GENEActiv accelerometer placed around the ankle, the smallest difference expected to claim a real change for a group of subjects should be 13% for the average HPA levels and 17% for the M5 ENMO (206, 207).

As performed in the previous chapter (study 1), based on these results, an attempt of power and sample size estimations was performed. For a two-arm study, a sample size of at least 85 participants per group, or 52 participants cohort for a one-arm study, would be required to allow the detection of a 20% difference or change, in the average HPA levels with 90% power (21, 62). This also assumes that patient populations are similar to this study, and that there is no significant variation on the HPA mean and standard deviation values. In this case the 20% difference or change represents 1.5 times the estimated SEM% giving chance for a real clinical change over the expected systematic error. Still, the level of change reflecting a significant change in patients' life needs further investigation.

4.5. CONCLUSIONS

When interpreting reports of activity monitor data caution is needed as the majority of these outputs come from validation studies performed with healthy

volunteers. Identified outcome measures such as energy expenditure or physical activity levels in neuromuscular disorders (208-212) are not as straightforward as expected. In the case of DM1, gait abnormalities such as impaired balance and mobility, and reduced endurance, certainly impact on the outputs obtained from accelerometers and other activity monitors (213). Testing the validity and reliability of an activity monitor in DM1 patients before implementing the device in a clinical trial will help researchers understand the possible strengths and barriers when using these tools in this particular group.

Significant differences are apparent when comparing two different monitor placement-sites but when attached to the same body location (different sides) comparisons seem valid. This study demonstrated that GENEActiv worn around the ankle can provide reliable data about DM1 ambulatory patterns in clinical settings and activity levels from daily life monitoring. Using objective methods, DM1 fatigued patients recorded significantly lower activity level in their daily life not only when compared to a healthy cohort but also against a chronic fatigued population. This study provides a significant amount of data to generate reference values for other researchers interested in utilizing this tool or ENMO as a unit measure when assessing HPA in DM1 patients.

Identifying the appropriate HPA measurement tool, either objective, subjective or both, and understanding its deliverables in the best possible way will not only generate high-quality data but will allow an efficient investment of time and resources when investigating HPA in any neuromuscular disorder. Future studies are needed to understand the underlying source of HPA differences and identify if there are specific activity-types that could be approached for health improvement. The methodology used in this study to assess HPA in DM1 (ankle-worn tri-axial accelerometer summarised in ENMO per 24 hours obtained from seven consecutive days of ≥23-hours of wearing time excluding at least the first and last days of the real expected monitoring period) can be applied in future DM1 clinical trials and this study will provide them evidence justifying their tool and methodology selection plus reference data to compare to.

CHAPTER 5. DISCUSSION

Currently, there is no cure for DM1, but the potential of new treatment strategies is more visible today than just a couple of years ago (214, 215). Identifying appropriate outcome measures that are feasible, valid and reliable within the DM1 population is essential to monitor disease progression and the effect of any intervention. A potential target for future interventions in DM1 should be to improve these patients' performance and participation in daily living activities (18). Outcome measures (OM) can evaluate different aspects of the disease, from body function and structures to participation and involvement in daily life (216, 217). DM1 is a multi-systemic disorder affecting many functions and organs. DM1 is also highly variable from severely disabled infants to near-normal, minimally affected adults. This variability requires careful characterization and stratification of patients for research, including research into OM. Moreover, a single OM might be influenced by more than one of the phenotypic features present in a patient with DM1, e.g. performance in the 6-Minute Walking Test (6MWT) may be influenced not only by muscle strength and endurance, but also by lung capacity(218, 219), cardiac disease(220), metabolic or endocrine abnormalities(221), and impaired cognitive ability(222, 223), all of which can be found in DM1. Therefore, it would not be appropriate to conclude that a change in the 6MWT over time is due to a change in muscle strength exclusively, as long as the other parameters are not controlled for. Assessing disease with complex outcomes can be an advantage but to fully understand changes to function and well-being of a DM1 patient, usually a battery of complimentary tests, both technical (laboratory results, imaging) and functional outcomes are required. This thesis focuses on exploring the use of three functional outcome measures (FOM) established by the outcome measures in myotonic dystrophy type 1 (OMMYD) consortium in 2015 and an accelerometry-based activity monitor. A clearly defined methodology to use these outcome measures and knowing the expected margin of error is of extreme importance when deciding on tools and designing protocols for future clinical trials in DM1. The properties analysed to evaluate the integrity of the selected outcomes include feasibility, reliability, validity, variability and responsiveness. These properties will establish the degree of error and what outcome is produced when utilized in this specific population (224).

Due to the sample size and the distribution of phenotypes and severity in this study, we are confident to present reliable estimates for those adults from the DM1 population who are most likely approached for clinical trials. More than 60% of the participants from the OPTIMISTIC (NCT02118779) and PHENO-DM1 (NCT02831504) trials were recruited through the UK Myotonic Dystrophy Patient Registry. This self-initiated registry includes an element of motivation and ability and represents the first line of recruitment for clinical trials (75).

Both studies start with a cross-sectional study to obtain a “snapshot” of our sample and to verify or reject any hypothesis of group or subgroup differences. Correlation results reported at this point were to analyse associations of variables but not to conclude on any causality as a longitudinal analysis would be required to confirm this. The introduction of a test-retest design in study 1 was chosen because of a justified need to explore any learning and/or fatigue effects that might impact on the results. Finally, the longitudinal analysis was performed to identify any possible change in time that could be attributed to a real change in the participants’ characteristics and not to a systemic error and to support the responsiveness capacity of the outcome in question.

5.1. STUDY 1

The first set of outcome measures explored are three examples of outcome measures of functional capacity, describing each individual’s ability to execute an action in a specific standardised and controlled circumstance (216). Since the OMMYD consortium proposed a test battery of functional outcome measures (FOM) considered clinically relevant and robust enough for execution in clinical trials in adults with DM1, efforts have been invested into validating and exploring the feasibility of their use (28). The 6 minute walk test (6MWT) has been explored in adults with DM1 before showing high relative and absolute reliability (39). One other study reported preliminary exploratory results from the 10-mWT and from the 10-meter max speed walk test (not run) as possible outcomes measuring balance and risk of falling (43). However, to our knowledge, this is the first time that full validity, reliability and responsiveness in DM1 have been tested for the

following tests (142): [1] timed 10-meter walk test (10-mWT) – comfortable speed; [2] timed 10-meter walk/run test (10-mW/RT); and [3] 30-second sit and stand test (30SSS).

Despite proving once more an excellent test-re-test relative reliability (i.e. coefficients ranged from 0.98 to 0.99), the systematic variability identified between test trials (first and second) justifies the need of repeated trials when implementing these tests. The correlation levels shown with SARA, muscle strength and disease-specific patient reported outcomes, plus the sensitivity to distinguish participants with a reported ability to run from those that do not and to distinguish participants with proximal muscle weakness, validate these outcomes as suitable surrogate markers of disease severity. Finally, the mean change detected after one year of follow-up that was higher than the minimum expected standard error of measurement (SEM) validates their sensitivity to detect real change and suggests a significant natural disease progression measurable within 12 months.

This study suggests the following cut-off points to identify more significantly impaired patients: [1] 9 sec for the 10-mWT; [2] 6-7 sec for the 10-mW/RT; and [3] 11 stands for the 30SSS. These cut-off values could also be considered for clinical practice as potential identifiers of participants requiring special attention or interventions such as assistive-walking devices, orthotics and exercises. To reduce inter-test variability and increase the chances of addressing a real change over time we recommend considering for analysis the average of all completed trials for the 10-mWT, 10-mW/RT and 30SSS. However, in the longitudinal analysis, the biggest percentage of change over time was shown in the change of the total stands of the 30SSS and in the best-trial score for each of the FOM. So, considering both, average and best score could be an option. On the other hand, to suggest ‘total stands’ of 30SSS as the score to count, we would first recommend a standardised assessment methodology that would specify the input of encouragement to complete the three test-trials and a standard time of rest in between.

5.1.1. *6MWT alternative outcomes*

6MWT is the most commonly used outcome measure in neuromuscular diseases to date (225, 226). However, the originally-specified corridor length required for the 6MWT has been a barrier to implementation into clinical trials particularly when involving unexperienced sites (87, 227, 228). Additionally, there is a need for more than one single trial to reduce test-to-test variability (39, 129, 131, 133, 134, 229). Identifying outcome alternatives with more flexibility for different settings and suitable for trials involving long visits and additional assessments could improve the research experience for participants and researchers. With this study we suggest that the combination of the 10-mWT and the 30SSS can substitute for the 6MWT. By performing two trials of the 10-mWT and an attempt for three trials of the 30SSS important factors interfering in daily life participation in DM1, such as the ability to walk, balance and fatigue, can be addressed (3, 18, 43). Additionally, with this combination of tests we can assess balance (and possible risk of falling), which, in the case of DM1, will provide a wider picture of the disease burden in relation to functionality. Tyson et al. (84) performed a systematic review, aiming to identify and score psychometrically robust and clinically feasible walking-based outcome measures in neurological disorders and selected the 10-meter walk test as among the best three outcomes. In this case we do not have any data suggesting that the 10-mWT is better option than the 6MWT as both are feasible, valid and with same levels of reliability; both tests required more than one trial to increase reliability and both correlate in similar degree to muscle strength and patient reported outcomes (PRO). The advantages over the 6-minute walking test are: [1] a lower risk for a flooring effect, allowing for the inclusion of the most impaired patients able to walk short distances; and [2] a corridor length with more flexibility to be implemented in clinical settings and in research sites with variable building distributions.

5.1.2. *Assessing fatigue and endurance*

Fatigue and endurance can also be measured with the 6MWT (230). Shorter (in time) tests have been found appropriate to assess walking capability, but risk to miss the fatigue effect detectable by a reduction in speed in the last minutes of

the 6MWT (231). However, based on this study experience and previous reports in other diseases, the 30SSS, especially when tested with repeated trials, can be used for assessing fatigue and cardiovascular fitness but this will need to be explored further (232-236). In this study the examiners reported that the primary reason for stopping the 30SSS and not completing the 3 trials was participants' fatigue. Even though it was not established whether there was a standard rest of period between trials or between tests, all examiners were instructed to give sufficient resting time in sitting to the participant between tests and to guarantee a heart rate stabilization before starting the 30SSS (i.e. after the 6MWT). Between each 30SSS trial patients were asked for their approval to continue and when a participant reported feeling too tired to continue the test would stop. Finally, examiners have discussed the possibility of a visible impact of the 30SSS on participants' heart rate and a visible higher exertion than that perceived after the 6MWT. Upon additional analysis performed during this study, the number of stands completed at second 20 of the 30 seconds sit and stand test were recorded and the performance estimated for the last ten seconds had a mode of 80% and an average of 89% (median 87%; SD28%). This estimation and the direct correlation with perceived fatigue should be considered for future studies. One-minute-length sit and stand test has been used in chronic obstructive pulmonary disease to assess functionality and exercise performance (232, 236, 237); by eliciting more than one 30-seconds-length trial, it is possible to reach the fitness challenge threshold that would make this test a tool to evaluate exercise tolerance as such as functional capacity. In this case-scenario, considering the 'total stands' score as the outcome of analysis will make good sense. Similarly to 6MWT, the 30SSS can also be influenced by the motivation of the patient(238) so a standardisation of the encouragement lines given before and along the test most probably will increase accuracy. These findings suggest the need to explore further this test as a possible assessment measuring fatigue and fitness levels in patients with DM1. This test (i.e. 30SSS) requires less time and space than the 6MWT and can be performed in a variety of location conditions.

5.1.3. *Assessing balance and risk of falls*

Impaired balance and a high risk of falls affect the DM1 population and relate to disease severity and muscle weakness (42, 43, 45, 126). Identifying participants with a higher risk of falling may help clinicians to intervene and to identify risk factors that can be controlled or prevented. The ability to stand up from a chair is an important component of maintaining independence and is a movement that depends on stability and balance (238). Sit and stand tests like the 30SSS have been suggested as good predictors of fallers in populations with high risk of falls like DM1 (43, 44, 238, 239).

5.2. STUDY 2

The second study corresponds to an accelerometry-based device as a potential measurement tool of a proband's performance (148, 240). Performance describes what a person actually does in his/her daily life (19, 217). The aim of this study was to gain more insight into the validity of one of these devices, GENEActiv, when assessing functional tasks and when measuring habitual physical activity (HPA) in daily life.

5.2.1. Systematic Review

After identifying the current interest in the neuromuscular field to implement activity monitors into research and being aware of their potential as outcome measures that could address patients' participation in daily life, a systematic review was performed, aiming to collect and analyse all studies reporting HPA in neuromuscular disorders (143, 144, 241). Challenges in measuring HPA in individuals with physical impairment are how to measure activity in a reliable and valid fashion, ensuring that the data provided are representative of the targeted population's real performance. Accelerometry-based devices are in current use in clinical trials in DM1. The OPTIMISTIC trial included an ankle-worn GENEActiv as part of its outcome measures (69). However, to our knowledge, there is no previous publication reporting the validity and reliability of this monitor in DM1 (148). Never the less when placed around the ankle.

5.2.2. *Validity of an accelerometry-based device in DM1*

Comparing acceleration scores with direct observation is an approach useful to establish criterion-validity of an activity monitor. GENEActiv placed at four different sites proved capable of capturing acceleration at different intensities of ambulation. Overall, all accelerometer site placements were significantly associated with the motion speeds performed at each task but wrist and ankle outputs cannot be translated to each other straightforwardly. Ankle-placed accelerometers performed better as sensitivity tools to differentiate ambulation speed between DM1 and healthy controls. The role of muscle weakness and impaired walking reflect on the reduced speedup of the upper limbs when compared to the ankles at the 10-meter walk/run test, more significantly when analysing the classic phenotype independently. These data cannot be translated into energy expenditure; for this, a comparison will be required against doubly labelled water ($^{2}H_2$ ^{18}O) method, in which after a dose of the liquid compound, the eliminated ^{2}H is subtracted from the amount of eliminated ^{18}O independently and equals the amount of CO_2 produced, representing the amount of energy expenditure from the moment of ingesting the liquid until the moment of elimination (209). Finally, due to the physiological variations in this population it will be invalid to employ predictive equations generated for other populations who are not at the same risk level.

5.2.3. *Habitual Physical Activity in DM1*

Byron et al. and Strath et al. (51, 160) proposed best practice guidelines when implementing activity monitors into research involving populations with physical impairments. They agree on the idea to avoid summary endpoints if these have not been validated before and to refer to raw data instead. The shorter the time sampling interval (epoch length or pre-defined activity bouts) the higher the chance to collect informative data from passive or slower participants. It is important to establish and employ standard methods for obtaining, cleaning and analysing data that, regardless of reducing the sample number, guarantee

reliable data. As we are dealing with an unknown sample regarding compliance and day distribution of activity patterns, in OPTIMISTIC we requested participants to wear the device full-time for two weeks. We excluded those not worn for at least 23 hours for a minimum of 7 consecutive days guaranteeing the inclusion of a weekday. This protocol showed feasibility with about 20% of data lost at every visit and a full study completion of only 58%. Lessons learned from this study may be used to develop a better-established logistics-system that could facilitate staff and study participants' management and understanding of the devices. In our systematic review we propose a guidance-checklist to follow whenever considering including an HPA measurement tool (including accelerometers) into a clinical trial or when reporting its use in neuromuscular disorders.

By combining previous experiences with activity monitors and the experience from the OPTIMISTIC trial, loss of data and participants' compliance are important issues to keep in mind prior estimating sample size, particularly if HPA as assessed by activity monitor is the primary outcome (51). There are, however, practices that presume better compliance such as: choosing distal joint areas (thigh, wrist or ankle) over centre-body areas like hips or chest (190, 242), a minimum of 7 days of requested monitoring to achieve at least 5 days of valid data and a minimum of 10 hours per day of wearable time.

5.3. BIAS MINIMIZATION

DM1 is a complex condition that presents a wide spectrum of severity, symptoms and impacts on functioning. An appropriate stratification of the population will help to enhance the clinical understanding of the natural disease progression (15, 118, 158, 243). The differences observed between the classic and the mild phenotypes were significant along all outcomes presented and their progression as sub-groups will be studied further. The differences between the sexes shown in all outcome measures should be noticed when comparing samples with different sex distributions (45, 118). Indeed, this suggestion has been made before, encouraging researchers to consider phenotype and sex separately when planning health standards of attention or when designing any future studies (15, 118, 158, 243). When designing a clinical trial protocol, especially randomized

controlled trials (RCTs), a plan for the minimization of potential bias should consider sex and the disease phenotype (118, 244).

5.4. STUDY LIMITATIONS

For a study searching for ideal outcome measures in a rare disease, few general limitations have to be mentioned (245). The healthy controls group differed in age and sex distribution from the DM1. However, even after including age and sex as possible confounders in the model, the differences between groups remained significant. This can be added to the fact that the milder phenotype prevailed among the older DM1 population with greater chances of showing a better performance, contrary to what is expected from a healthy population. Still, an ideal comparison between groups would come from age-matched groups.

Intra-rater reliability has been robustly assessed before (44, 142) and once more with this study, but inter-rater reliability has not been studied in DM1. However, publications in other neurologically impaired diseases have shown high inter-rater reliability scores for the walking tests in other neurological disorders (84, 125, 246). To properly assess these FOM feasibility and to influence properly on trials preference, an appropriate patients' feedback questionnaire or experience recollection would have been ideal to enrich information about feasibility. Enhancing the participant's clinical trials experience could improve study compliance and is one of the central pillars of quality in healthcare attention (247).

All these FOM were performed on the same day within a long study visit that involved other physical assessments and a large number of questionnaires to complete. On average these visits lasted between 4 and 6 hours, which most probably impacted on the participants' performance during the assessments. From the start of the study there was an agreement to retain the order of assessments for all participants: prioritizing cardio-respiratory assessments at the beginning of the visit, followed by patient-reported outcomes and finalized with FOM and strength assessments. Additionally, as an attempt to minimize variability due to fatigue and motivation, the time of assessments and the

sequence followed would be recorded for each participant and we would attempt to replicate the sequence on the follow-up visit.

Unfortunately, at the time of this thesis, not all of the PHENO-DM1 study follow-up visits had been carried out limiting the longitudinal study sample size. Having the full data of longitudinal analysis may provide additional information to effectively stratify data according to participants' sex and disease phenotype. The analysis of a second visit will allow the identification of cases with no significant measurable changes over time and most probably attained their own highest possible scores (i.e. ceiling effect) (248).

We have to be cautious when generalising the disease change over time as this longitudinal studies have 10 to 12 months between assessments, which may be too short or too long for a treatment to show any effect (245). Still, treatments are needed for long-term effects in addition to the short-term effects and knowing the progression trend of a disease-sample over a year will support the decision-making process of a change expected to reflect an intervention effect.

6. CONCLUSIONS AND CLINICAL IMPLICATIONS

The results presented in both of my sub-studies can be used to propose outcome measures that provide an objective picture capacity and performance of the DM1 patient. This study established the reliability and validity of a set of functional outcome measures and of an accelerometry-based activity monitor appropriate for use in mild to moderately affected, ambulant adults with DM1. Using valid and reliable outcome measures in this disease will provide a better understanding of its complexity and ultimately will support the identification of effective treatment opportunities.

6MWT continues to be an assessment with high correlation scores to other disease severity outcomes such as: muscle strength; the SARA and patient-reported outcomes. However, the three-studied FOM either independently or in combination seem to be good alternatives, as they show strong correlation values with disease severity outputs and, as shown in my study, have proven validity and reliability in patients with DM1.

Protocol recommendations derived from my study are: [1] consider including more than one assessment of functional capacity (i.e. FOM) and one measurement of performance (i.e. activity monitor) regardless of correlating between each other, these do not substitute each other and both types of outcomes are significant when explaining the level of disease impairment; [2] follow a standardized protocol that has been previously used in DM1 will allow intra-studies comparability; [3] perform FOM at least twice whenever possible and use either the average or the best score as the test results; [4] when using any activity monitor, consider a prior-protocol validation test if this has not been performed before; [5] identify the possible differences between the sexes and disease phenotypes (classic and mild), this is essential when studying disease progression of a sample with variability within the cohort; minimising a sample randomization for these factors (sex and phenotype) should be considered; and [6] there is minimal expected change of 5% (0.5 sec) for the 10-mWT, 8% (0.4 sec) for the 10-mW/RT and 10% (1.4 times) for the 30SSS, and 13% (2.7 mg ENMO) for the HPA (accelerometry) reports, that can be attributed to

measurement error and should be considered when designing a trial and when estimating the power and sample size.

The choice of outcome measure for future trials should be guided by the domain of the disease that an intervention is likely to impact on (245). When combined FOM and objectively measured HPA patterns in DM1, researchers can get a more complete picture of the real functionality of the patient including information about their capacity and their performance in daily life. An ideal intervention should impact on both aspects of functionality and translate into a better quality of life.

7. FUTURE DIRECTIONS

This study represents a substantial advance towards the standardization of disease-specific outcome measures in DM1. Questions remain about the design of a clinical trial protocol and the proper definition of outcome measures that will portray the participant's overall disease status and progression.

One of the first plans for the short-term future is to perform as similar study focus on the upper extremity functionality and strength; assessing the validity and reliability of the Nine-hole Peg test (9HPT) in order to complete the deep exploration of all OMMYD FOM.

Impaired balance with a risk of falling has been recognized as a frequent and common problem in adults with DM1 (42, 43, 45, 126). Still, the standards of care in the UK do not include a proper assessment and prevention of falls programme for DM1 patients (16) and is not yet recognized as a cohort at risk which may limit access to local fall services (249). There is a current need to perform a project aiming to increase the awareness of this issue within DM1 healthcare-personnel and researchers may be warranted. We recently developed a multinational survey to estimate the risk of falls and estimated an odds ratio of 1.6:1 of every DM1 adult <65 over every healthy adult >65 years old which represents the population with a higher risk of falling and the cut-off age to become eligible to enter falls clinics. By identifying an appropriate tool to predict falls that can be performed at clinic standards could facilitate the identification of patients that need attention and falls rehabilitation/education in a timely fashion. SARA and 30SSS are potential falls predictors; however, to validate these fully, we proposed a longitudinal recollection of falls and a study of associations with any gold standard balance assessment tool (234).

The use of accelerometers is lagging behind in the validation and harmonization efforts as the importance of these tools comes from the translation of raw data into clinical outputs with significant values. There is an initial need to identify physical activity levels and/or energy expenditure scores that can translate from the counts (ENMO/min) reported by GENEActiv. This will raise the interest of

researchers in implementing these devices into practice. In addition, these findings can encourage the experimentation with other types of activity monitors and compare their strengths and weaknesses to identify a valid but feasible and user-friendly device appropriate for DM1 trials.

Finally, but most importantly, there is the short-term plan of analyzing the follow-up data of the full cohort. With the whole sample's baseline and follow-up data, we aim to identify ceiling effect rates for each outcome and establish cut-off values that will predict this (and possibly to consider as exclusion criteria in clinical trials). The prediction is that very mild participants might not progress significantly enough to be suitable for effective interventions assessment. For this conclusion to be valid a subgroup analysis is required. Moreover, the OPTIMISTIC study will reveal whether cognitive behavioural therapy and exercise is of benefit to patients with DM1, and the PHENO-DM1 study may reveal blood biomarkers suitable to monitor disease progression align with functional changes.

REFERENCES

1. Chen H. Atlas of genetic diagnosis and counseling: Humana Press Totowa NJ; 2006.
2. Harper P. Myotonic dystrophy: OUP Oxford; 2009.
3. Bird TD. Myotonic dystrophy type 1. GeneReviews, University of Washington Seattle WA; 2015.
4. Norwood FLM, Harling C, Chinnery PF, Eagle M, Bushby K, Straub V. Prevalence of genetic muscle disease in Northern England: in-depth analysis of a muscle clinic population. Brain. 2009; 132(11): 3175-86.
5. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell. 1992; 68(4): 799-808.
6. Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, et al. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature. 1992; 355(6360): 545-6.
7. Harley HG, Rundle S, MacMillan J, Myring J, Brook J, Crow S, et al. Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. American Journal of Human Genetics. 1993; 52(6): 1164.
8. Lee KY, Li M, Manchanda M, Batra R, Charizanis K, Mohan A, et al. Compound loss of muscleblind‐like function in myotonic dystrophy. EMBO Molecular Medicine. 2013; 5(12): 1887-900.
9. Kim D-H, Langlois M-A, Lee K-B, Riggs AD, Puymirat J, Rossi JJ. HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. Nucleic Acids Research. 2005; 33(12): 3866-74.
10. Pettersson OJ, Aagaard L, Jensen TG, Damgaard CK. Molecular mechanisms in DM1—a focus on foci. Nucleic Acids Research. 2015; 43(4): 2433-41.
11. Wheeler TM, Leger AJ, Pandey SK, MacLeod AR, Nakamori M, Cheng SH, et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature. 2012; 488(7409): 111-5.

12. Kellerer M, Lammers R, Ermel B, Tippmer S, Vogt B, Obermaier-Kusser B, et al. Distinct alpha-subunit structures of human insulin receptor A and B variants determine differences in tyrosine kinase activities. Biochemistry. 1992; 31(19): 4588-96.

13. Gao Z, Cooper TA. Reexpression of pyruvate kinase M2 in type 1 myofibers correlates with altered glucose metabolism in myotonic dystrophy. Proceedings of the National Academy of Sciences. 2013; 110(33): 13570-5.

14. Thornton CA. Myotonic dystrophy. Neurologic Clinics. 2014; 32(3): 705-19.

15. 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. Journal of Neurology. 2015; 262(2): 285-93.

16. Turner C, Hilton-Jones D. Myotonic dystrophy: diagnosis, management and new therapies. Current Opinion in Neurology. 2014; 27(5): 599-606.

17. Lau J, Sy R, Corbett A, Kritharides L. Myotonic dystrophy and the heart: A systematic review of evaluation and management. International Journal of Cardiology. 2015; 184: 600-8.

18. Gagnon C, Mathieu J, Jean S, Laberge L, Perron M, Veillette S, et al. Predictors of disrupted social participation in myotonic dystrophy type 1. Archives of Physical Medicine & Rehabilitation. 2008; 89(7): 1246-55.

19. WHO. In: World Health Organization, editor. International Classification of Functioning, Disability and Health: ICF Short Version. Geneva, Switzerland: WHO Press; 2001.

20. Kohler F, Connolly C, Sakaria A, Stendara K, Buhagiar M, Mojaddidi M. Can the ICF be used as a rehabilitation outcome measure? A study looking at the inter- and intra-rater reliability of ICF categories derived from an ADL assessment tool. Journal of Rehabilitation Medicine. 2013; 45(9): 881-7.

21. Allet L, Bürge E, Monnin D. ICF: Clinical relevance for physiotherapy? A critical review. Advances in Physiotherapy. 2008; 10(3): 127-37.

22. Hammel J, Magasi S, Heinemann A, Whiteneck G, Bogner J, Rodriguez E. What does participation mean? An insider perspective from people with disabilities. Disability & Rehabilitation. 2008; 30(19): 1445-60.

23. Brockow T, Cieza A, Kuhlow H, Sigl T, Franke T, Harder M, Stucki G. Identifying the concepts contained in outcome measures of clinical trials on musculoskeletal disorders and chronic widespread pain using the International

Classification of Functioning, Disability and Health as a reference. Journal of Rehabilitation Medicine. 2004; 36(44 Suppl): 30-6.

24. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. Journal of Rehabilitation Medicine. 2002; 34(5): 205-10.

25. Kierkegaard M, Harms-Ringdahl K, Holmqvist LW, Tollbäck A. Perceived functioning and disability in adults with myotonic dystrophy type 1: a survey according to the International Classification of Functioning, Disability and Health. Journal of Rehabilitation Medicine. 2009; 41(7): 512-20.

26. Kierkegaard M. Disability and Physical Exercise in adults with Myotonic Dystrophy type 1: Inst för neurobiologi, vårdvetenskap och samhälle/Dept of Neurobiology, Care Sciences and Society; 2010.

27. Gagnon C, Meola G, Hébert LJ, Puymirat J, Laberge L, Leone M. Report of the first outcome measures in myotonic dystrophy type 1 (OMMYD-1) international workshop: Clearwater, Florida, November 30, 2011. Neuromuscular Disorders. 2013; 23(12): 1056-68.

28. Gagnon C, Meola G, Hébert LJ, Laberge L, Leone M, Heatwole C. Report of the second outcome measures in myotonic dystrophy type 1 (OMMYD-2) international workshop San Sebastian, Spain, October 16, 2013. Neuromuscular Disorders. 2015; 25(7): 603-16.

29. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. The Journal of Rheumatology. 1998; 25(2): 198-9.

30. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. The Journal of Rheumatology. 1997; 24(4): 799-802.

31. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials. 2007; 8(1): 38.

32. del Ser T, Steinwachs KC, Gertz HJ, Andres MV, Gomez-Carrillo B, Medina M, et al. Treatment of Alzheimer's disease with the GSK-3 inhibitor tideglusib: a pilot study. Journal of Alzheimer's Disease. 2013; 33(1): 205-15.

33. Eldar-Finkelman H, Martinez A. GSK-3 Inhibitors: Preclinical and Clinical Focus on CNS. Frontiers in Molecular Neuroscience. 2011; 4.

34. Gao X, Wang JY, Gao LM, Yin XF, Liu L. Identification and analysis of glycogen synthase kinase 3 beta1 interactome. Cell Biology International. 2013; 37(8): 768-79.

35. Jones K, Wei C, Iakova P, Bugiardini E, Schneider-Gold C, Meola G, et al. GSK3beta mediates muscle pathology in myotonic dystrophy. The Journal of Clinical Investigation. 2012; 122(12): 4461-72.

36. Heatwole CR, Eichinger KJ, Friedman DI, Hilbert JE, Jackson CE, Logigian EL, et al. An Open-Label Trial of Recombinant Human Insulin-Like Growth Factor-I/Recombinant Human Insulin-Like Growth Factor Binding Protein-3 (rhIGF-1/rhIGFBP-3) in Myotonic Dystrophy Type 1. Arch Neurol. 2011; 68(1): 37-44.

37. Pandey SK, Wheeler TM, Justice SL, Kim A, Younis HS, Gattis D, et al. Identification and characterization of modified antisense oligonucleotides targeting DMPK in mice and nonhuman primates for the treatment of myotonic dystrophy type 1. The Journal of Pharmacology & Experimental Therapeutics. 2015; 355(2): 329-40.

38. Arnold WD, Kline D, Sanderson A, Hawash AA, Bartlett A, Novak KR, et al. Open-label trial of ranolazine for the treatment of myotonia congenita. Neurology. 2017: 10.1212/WNL. 0000000000004229.

39. Kierkegaard M, Tollbäck A. Reliability and feasibility of the six minute walk test in subjects with myotonic dystrophy. Neuromuscular Disorders. 2007; 17(11): 943-9.

40. Pucillo EM, Dibella DL, Hung M, Bounsanga J, Crockett B, Dixon M, et al. Physical function and mobility in children with congenital myotonic dystrophy. Muscle & Nerve. 2017.

41. Pirpiris M, Wilkinson AJ, Rodda J, Nguyen TC, Baker RJ, Nattrass GR, et al. Walking speed in children and young adults with neuromuscular disease: comparison between two assessment methods. Journal of Pediatric Orthopedics. 2003; 23(3): 302-7.

42. DiPaolo G, Jimenez-Moreno C, Nikolenko N, Atalaia A, Monckton DG, Guglieri M, et al. Functional impairment in patients with myotonic dystrophy type 1 can be assessed by an ataxia rating scale (SARA). Journal of Neurology. 2017; 264(4): 701-8.

43. 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. Neuromuscular Disorders. 2014; 24(3): 207-15.

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

45. 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. Neuromuscular Disorders. 2015; 25(2): 141-8.

46. Andersson ÅG, Kamwendo K, Appelros P. Fear of falling in stroke patients: relationship with previous falls and functional characteristics. International Journal of Rehabilitation Research. 2008; 31(3): 261-4.

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

48. Aldehag AS, Jonsson H, Ansved T. Effects of a hand training programme in five patients with myotonic dystrophy type 1. Occupational Therapy International. 2005; 12(1): 14-27.

49. Whiteneck G, Dijkers MP. Difficult to measure constructs: conceptual and methodological issues concerning participation and environmental factors. Archives of Physical Medicine & Rehabilitation. 2009; 90(11): S22-S35.

50. Westerterp KR. Physical activity assessment with accelerometers. International Journal of Obesity & Related Metabolic Disorders. 1999; 23.

51. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the assessment of physical activity: clinical and research applications. Circulation. 2013; 128(20): 2259-79.

52. Steins D, Dawes H, Esser P, Collett J. Wearable accelerometry-based technology capable of assessing functional activities in neurological populations in community settings: a systematic review. Journal of Neuroengineering & Rehabilitation. 2014; 11(1): 36.

53. Motl RW, Snook EM, Agiovlasitis S. Does an accelerometer accurately measure steps taken under controlled conditions in adults with mild multiple sclerosis? Disability & Health Journal. 2011; 4(1): 52-7.

54. Del Din SD, Godfrey A, Mazza C, Lord S, Rochester L. Free-living monitoring of Parkinson's disease: Lessons from the field. Movement Disorders. 2016; 1293-1313.

55. Fini NA, Holland AE, Keating J, Simek J, Bernhardt J. How is physical activity monitored in people following stroke? Disability & Rehabilitation. 2015; 37(17): 1717-31.

56. Keawutan P, Bell K, Davies PS, Boyd RN. Systematic review of the relationship between habitual physical capacity and motor capacity in children with cerebral palsy. Res Dev Disabil. 2014; 35(6): 1301-9.

57. Mercuri E, Mayhew A, Muntoni F, Messina S, Straub V, Van Ommen GJ, et al. Towards harmonisation of outcome measures for DMD and SMA within TREAT-NMD; report of three expert workshops: TREAT-NMD/ENMC workshop on outcome measures, 12th--13th May 2007, Naarden, The Netherlands; TREAT-NMD workshop on outcome measures in experimental trials for DMD, 30th June--1st July 2007, Naarden, The Netherlands; conjoint Institute of Myology TREAT-NMD meeting on physical activity monitoring in neuromuscular disorders, 11th July 2007, Paris, France. Neuromuscular Disorders. 2008; 18(11): 894-903.

58. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. American Journal of Health-System Pharmacy. 2008; 65(23).

59. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010; 19(4): 539-49.

60. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86(2): 420.

61. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. 2016; 15(2): 155-63.

62. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychological Reports. 1966; 19(1): 3-11.

63. Myles P, Cui J. I. Using the Bland–Altman method to measure agreement with repeated measures. British Jrnl Anaesthesia; 2007.

64. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986; 327(8476): 307-10.

65. Florkowski CM. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. The Clinical Biochemist Reviews. 2008; 29(Suppl 1): S83-7.

66. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. 1993; 39(4): 561-77.

67. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine. 1978; 8(4): 283-98.

68. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology. 2008; 61(2): 102-9.

69. van Engelen B. Cognitive behaviour therapy plus aerobic exercise training to increase activity in patients with myotonic dystrophy type 1 (DM1) compared to usual care (OPTIMISTIC): study protocol for randomised controlled trial. Trials. 2015; 16: 224.

70. Schillings M, Kalkman J, Janssen H, Van Engelen B, Bleijenberg G, Zwarts M. Experienced and physiological fatigue in neuromuscular disorders. Clinical Neurophysiology. 2007; 118(2): 292-300.

71. Kalkman JS, Schillings ML, Zwarts MJ, van Engelen BG, Bleijenberg G. The development of a model of fatigue in neuromuscular disorders: a longitudinal study. Journal of Psychosomatic Research. 2007; 62(5): 571-9.

72. Kalkman JS, Schillings ML, van der Werf SP, Padberg GW, Zwarts MJ, van Engelen BG, et al. Experienced fatigue in facioscapulohumeral dystrophy, myotonic dystrophy, and HMSN-I. Journal of Neurology, Neurosurgery & Psychiatry. 2005; 76(10): 1406-9.

73. Hermans MC, Faber CG, De Baets MH, de Die-Smulders CE, Merkies IS. Rasch-built myotonic dystrophy type 1 activity and participation scale (DM1-Activ). Neuromuscular Disorders. 2010; 20(5): 310-8.

74. Mathieu J, Boivin H, Meunier D, Gaudreault M, Begin P. Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. Neurology. 2001; 56(3): 336-40.

75. Wood L, Cordts I, Atalaia A, Marini-Bettolo C, Maddison P, Phillips M, et al. The UK Myotonic Dystrophy Patient Registry: facilitating and accelerating clinical research. Journal of Neurology. 2017; 264(5): 979-88.

76. Lin MR, Hwang HF, Hu MH, Wu HDI, Wang YW, Huang FC. Psychometric comparisons of the timed Up and Go, One-Leg stand, functional reach, and tinetti balance measures in community-dwelling older people. Journal of the American Geriatrics Society. 2004; 52(8): 1343-8.

77. Hammarén E, Kjellby-Wendt G, Lindberg C. Quantification of mobility impairment and self-assessment of stiffness in patients with myotonia congenita by the physiotherapist. Neuromuscular Disorders. 2005; 15(9): 610-7.

78. 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. Annals of Physical & Rehabilitation Medicine. 2010; 53(6): 387-98.

79. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. The Journal of Orthopaedic & Sports Physical Therapy. 2011; 41(5): 319-27.

80. Heatwole C, Bode R, Johnson NE, Dekdebrun J, Dilek N, Eichinger K, et al. Myotonic dystrophy health index: Correlations with clinical tests and patient function. Muscle & Nerve. 2016; 53(2): 183-90.

81. Heatwole C, Bode R, Johnson N, Dekdebrun J, Dilek N, Heatwole M, et al. Myotonic Dystrophy Health Index: Initial evaluation of a disease‐specific outcome measure. Muscle & Nerve. 2014; 49(6): 906-14.

82. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Research Quarterly for Exercise & Sport. 1999; 70(2): 113-9.

83. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. Age & Ageing. 1997; 26(1): 15-9.

84. Tyson S, Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. Clinical Rehabilitation. 2009; 23(11): 1018-33.

85. Enright PL. The six-minute walk test. Respiratory Care. 2003; 48(8): 783-5.

86. STATEMENT A. ATS committee on proficiency standards for clinical pulmonary function laboratories. Am J Respir Crit Care Med. 2002. pp. 111-7.

87. McDonald CM, Henricson EK, Han JJ, Abresch RT, Nicorici A, Elfring GL, et al. The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy. Muscle & Nerve. 2010; 41(4): 500-10.

88. Schmitz-Hübsch T, Du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, et al. Scale for the assessment and rating of ataxia Development of a new clinical scale. Neurology. 2006; 66(11): 1717-20.

89. Hermans MC, Hoeijmakers JG, Faber CG, Merkies IS. Reconstructing the Rasch-built myotonic dystrophy type 1 activity and participation scale. PloS One. 2015; 10(10): e0139944.

90. Heatwole C, Bode R, Johnson N, Quinn C, Martens W, McDermott MP, et al. Patient-reported impact of symptoms in myotonic dystrophy type 1 (PRISM-1). Neurology. 2012; 79(4): 348-57.

91. Barr AE, Diamond BE, Wade CK, Harashima T, Pecorella WA, Potts CC, et al. Reliability of testing measures in Duchenne or Becker muscular dystrophy. Archives of Physical Medicine & Rehabilitation. 1991; 72(5): 315-9.

92. Escolar D, Henricson E, Mayhew J, Florence J, Leshner R, Patel K, et al. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. Muscle & Nerve. 2001; 24(6): 787-93.

93. Escolar DM, Henricson EK, Pasquali L, Gorni K, Hoffman EP. Collaborative translational research leading to multicenter clinical trials in Duchenne muscular dystrophy: the Cooperative International Neuromuscular Research Group (CINRG). Neuromuscular Disorders. 2002; 12: S147-S54.

94. Florence JM, Pandya S, King WM, Robison JD, Signore LC, Wentzell M, et al. Clinical trials in Duchenne dystrophy: standardization and reliability of evaluation procedures. Physical Therapy. 1984; 64(1): 41-5.

95. Zar JH. Significance testing of the Spearman rank correlation coefficient. Journal of the American Statistical Association. 1972; 67(339): 578-80.

96. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. Journal of Clinical Epidemiology. 2000; 53(5): 459-68.

97. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. Journal of Chronic Diseases. 1986; 39(11): 897-906.

98. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Group CSCM, editors. Methods to explain the clinical significance of health status measures. Mayo Clinic Proceedings; 2002: Elsevier.

99. Martin F, Inoue E, Cortese ICM, de Almeida Kruschewsky R, Adonis A, Grassi MFR, et al. Timed walk as primary outcome measure of treatment response in clinical trials for HTLV-1-associated myelopathy: a feasibility study. Pilot & Feasibility Studies. 2015; 1.

100. Shoukri M, Asyali M, Donner A. Sample size requirements for the design of reliability study: review and new results. Statistical Methods in Medical Research. 2004; 13(4): 251-71.

101. Chetta A, Zanini A, Pisi G, Aiello M, Tzani P, Neri M, et al. Reference values for the 6-min walk test in healthy subjects 20–50 years old. Respiratory Medicine. 2006; 100(9): 1573-8.

102. Samson MM, Crowe A, de Vreede PL, Dessens JA, Duursma SA, Verhaar HJ. Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight. Aging (Milan, Italy). 2001; 13(1): 16-21.

103. Medijainen K, Pääsuke M, Lukmann A, Taba P. Functional Performance and Associations between Performance Tests and Neurological Assessment Differ in Men and Women with Parkinson's Disease. Behavioural Neurology. 2015; 2015.

104. Oberg T, Karsznia A, Oberg K. Basic gait parameters: reference data for normal subjects, 10-79 years of age. J Rehabil Res Dev. 1993; 30(2): 210-23.

105. Batista FS, Gomes GA, Neri AL, Guariento ME, Cintra FA, Sousa Mda L, et al. Relationship between lower-limb muscle strength and frailty among elderly people. Sao Paulo Medical Journal = Revista Paulista de Medicina. 2012; 130(2): 102-8.

106. Bendall MJ, Bassey EJ, Pearson MB. Factors affecting walking speed of elderly people. Age & Ageing. 1989; 18(5): 327-32.

107. Dogan C, De Antonio M, Hamroun D, Varet H, Fabbro M, Rougier F, et al. Gender as a Modifying Factor Influencing Myotonic Dystrophy Type 1 Phenotype Severity and Mortality: A Nationwide Multiple Databases Cross-Sectional Observational Study. PLoS One. 2016; 11(2).

108. Erdmann PG, Teunissen LL, van den Berg LH, Notermans NC, Schröder CD, Bongers BC, et al. Validity of the shuttle walk test as a functional assessment of walking ability in individuals with polyneuropathy. Disability & Rehabilitation. 2016: 1-7.

109. Bohannon RW. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. Age & Ageing. 1997; 26(1): 15-9.

110. Rikli RE, Jones CJ. Functional fitness normative scores for community-residing older adults, ages 60-94. Journal of Aging & Physical Activity. 1999; 7(2): 162-81.

111. Macfarlane DJ, Chou KL, Cheng YH, Chi I. Validity and normative data for thirty-second chair stand test in elderly community-dwelling Hong Kong Chinese. American Journal of Human Biology: the official journal of the Human Biology Council. 2006; 18(3): 418-21.

112. Solbakken G, Ørstavik K, Hagen T, Dietrichs E, Naerland T. Major involvement of trunk muscles in myotonic dystrophy type 1. Acta Neurologica Scandinavica. 2016; 134(6): 467-73.

113. Nitz JC, Burns YR, Jackson RV. Sit-to-stand and walking ability in patients with neuromuscular conditions. Physiotherapy. 1997; 83(5): 223-7.

114. Speight J, Barendse SM. FDA guidance on patient reported outcomes. British Medical Journal Publishing Group; 2010.

115. van Engelen BG, Eymard B, Wilcox D. 123rd ENMC International Workshop: management and therapy in myotonic dystrophy, 6-8 February 2004, Naarden, The Netherlands. Neuromuscular Disorders. 2005; 15(5): 389-94.

116. Torres C, Moxley RT, Griggs RC. Quantitative testing of handgrip strength, myotonia, and fatigue in myotonic dystrophy. Journal of the Neurological Sciences. 1983; 60(1): 157-68.

117. Nitz J, Burns Y, Jackson R. A longitudinal physical profile assessment of skeletal muscle manifestations in myotonic dystrophy. Clinical Rehabilitation. 1999; 13(1): 64-73.

118. Raymond K, Levasseur M, Mathieu J, Desrosiers J, Gagnon C. A 9-year follow-up study of the natural progression of upper limb performance in myotonic dystrophy type 1: A similar decline for phenotypes but not for gender. Neuromuscular Disorders. 2017; 27(7): 673-82.

119. Silva PF, Quintino LF, Franco J, Faria CD. Measurement properties and feasibility of clinical tests to assess sit-to-stand/stand-to-sit tasks in subjects with neurological disease: a systematic review. Brazilian Journal of Physical Therapy. 2014; 18(2): 99-110.

120. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Research Quarterly for Exercise & Sport. 1999; 70(2): 113-9.

121. Paltamaa J, West H, Sarasoja T, Wikström J, Mälkiä E. Reliability of physical functioning measures in ambulatory subjects with MS. Physiotherapy Research International. 2005; 10(2): 93-109.

122. Rossier P, Wade DT. Validity and reliability comparison of 4 mobility measures in patients presenting with neurologic impairment. Archives of Physical Medicine & Rehabilitation. 2001; 82(1): 9-13.

123. Horemans HL, Beelen A, Nollet F, Lankhorst GJ. Reproducibility of walking at self-preferred and maximal speed in patients with postpoliomyelitis syndrome. Archives of Physical Medicine & Rehabilitation. 2004; 85(12): 1929-32.

124. Graham JE, Ostir GV, Kuo Y-F, Fisher SR, Ottenbacher KJ. Relationship between test methodology and mean velocity in timed walk tests: a review. Archives of Physical Medicine & Rehabilitation. 2008; 89(5): 865-72.

125. Graham JE, Ostir GV, Fisher SR, Ottenbacher KJ. Assessing walking speed in clinical research: a systematic review. Journal of Evaluation in Clinical Practice. 2008; 14(4): 552-62.

126. Wiles CM, Busse ME, Sampson CM, Rogers MT, Fenton‐May J, van Deursen R. Falls and stumbles in myotonic dystrophy. Journal of Neurology, Neurosurgery, & Psychiatry. 2006; 77(3): 393-6.

127. Bowden MG, Balasubramanian CK, Behrman AL, Kautz SA. Validation of a speed-based classification system using quantitative measures of walking performance poststroke. Neurorehabil Neural Repair. 2008; 22(6): 672-5.

128. Plummer P, Behrman AL, Duncan PW, Spigel P, Saracino D, Martin J, et al. Effects of stroke severity and training duration on locomotor recovery after stroke: a pilot study. Neurorehabil Neural Repair. 2007; 21(2): 137-51.

129. Kierkegaard M, Petitclerc E, Hebert 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. Neuromuscular Disorders. 2017; 27(2): 153-8.

130. Motyl JM, Driban JB, McAdams E, Price LL, McAlindon TE. Test-retest reliability and sensitivity of the 20-meter walk test among patients with knee osteoarthritis. BMC Musculoskeletal Disorders. 2013; 14: 166.

131. Ziegler B, Rovedder PME, Oliveira CL, e Silva FdA, Dalcin PdTR. Repeatability of the 6-minute walk test in adolescents and adults with cystic fibrosis. Respiratory Care. 2010; 55(8): 1020-5.

132. Moller AB, Bibby BM, Skjerbaek AG, Jensen E, Sorensen H, Stenager E, et al. Validity and variability of the 5-repetition sit-to-stand test in patients with multiple sclerosis. Disability & Rehabilitation. 2012; 34(26): 2251-8.

133. Camarri B, Eastwood PR, Cecins NM, Thompson PJ, Jenkins S. Six minute walk distance in healthy subjects aged 55–75 years. Respiratory Medicine. 2006; 100(4): 658-65.

134. Jenkins S, Cecins NM. Six-minute walk test in pulmonary rehabilitation: do all patients need a practice test? Respirology (Carlton, Vic). 2010; 15(8): 1192-6.

135. Gallais B, Montreuil M, Gargiulo M, Eymard B, Gagnon C, Laberge L. Prevalence and correlates of apathy in myotonic dystrophy type 1. BMC Neurology. 2015; 15: 148.

136. Lecours J, Nadeau S, Gravel D, Teixera-Salmela L. Interactions between foot placement, trunk frontal position, weight-bearing and knee moment asymmetry at seat-off during rising from a chair in healthy controls and persons with hemiparesis. Journal of Rehabilitation Medicine. 2008; 40(3): 200-7.

137. Lemay JF, Nadeau S. Standing balance assessment in ASIA D paraplegic and tetraplegic participants: concurrent validity of the Berg Balance Scale. Spinal Cord. 2010; 48(3): 245-50.

138. Roldán-Jiménez C, Bennett P, Cuesta-Vargas AI. Muscular activity and fatigue in lower-limb and trunk muscles during different sit-to-stand tests. PloS One. 2015; 10(10): e0141675.

139. Blankevoort CG, van Heuvelen MJ, Scherder EJ. Reliability of six physical performance tests in older people with dementia. Physical Therapy. 2013; 93(1): 69-78.

140. Flansbjer UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. Journal of Rehabilitation Medicine. 2005; 37(2): 75-82.

141. Dupont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. Controlled Clinical Trials. 1998; 19(6): 589-601.

142. Symonds T, Campbell P, Randall JA. A review of muscle-and performance-based assessment instruments in DM1. Muscle & Nerve. 2017.

143. Steins D, Dawes H, Esser P, Collett J. Wearable acceloremtry-based technology capable of assessing functional activities in neurological populations in community settings: a systematic review. Journal of Neuroengineering & Rehabilitation. 2014; 11: 36.

144. Mercuri E, Mayhew A, Muntoni F, Messina S, Straub V, Van Ommen G, et al. Towards harmonisation of outcome measures for DMD and SMA within TREAT-NMD; report of three expert workshops: TREAT-NMD/ENMC workshop on outcome measures, 12th–13th May 2007, Naarden, The Netherlands; TREAT-NMD workshop on outcome measures in experimental trials for DMD, 30th June–1st July 2007, Naarden, The Netherlands; conjoint Institute of Myology TREAT-NMD meeting on physical activity monitoring in neuromuscular disorders, 11th July 2007, Paris, France. Neuromuscular Disorders. 2008; 18(11): 894-903.

145. Üstün TB, Chatterji S, Bickenbach J, Kostanjsek N, Schneider M. The International Classification of Functioning, Disability and Health: a new tool for understanding disability and health. Disability & Rehabilitation. 2003; 25(11-12): 565-71.

146. Rand D, Eng JJ, Tang P-F, Hung C, Jeng J-S. Daily physical activity and its contribution to the health-related quality of life of ambulatory individuals with chronic stroke. Health & Quality of Life Outcomes. 2010; 8(1): 80.

147. Motl RW, McAuley E, Snook EM, Gliottoni RC. Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. Psychology, Health & Medicine. 2009; 14(1): 111-24.

148. Jimenez-Moreno AC, Newman J, Charman SJ, Catt M, Trenell MI, Gorman GS, et al. Measuring Habitual Physical Activity in Neuromuscular Disorders: A Systematic Review. Journal of Neuromuscular Diseases. 2017; 4(1): 25-52.

149. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Reports. 1985; 100(2): 126.

150. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. Journal of Applied Physiology. 2014; 117(7): 738-44.

151. Menai M, Van Hees VT, Elbaz A, Kivimaki M, Singh-Manoux A, Sabia S. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. Scientific Reports. 2017; 8.

152. Rowlands AV, Olds TS, Hillsdon M, Pulsford R, Hurst TL, Eston RG, et al. Assessing sedentary behavior with the GENEActiv: introducing the sedentary sphere: Lippincott Williams & Wilkins; 2014.

153. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Medicine & Science in Sports & Exercise; 2011; 43(6): 1085-93.

154. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. Medicine & Science in Sports & Exercise. 2014; 46(9): 1816-24.

155. Van Hees VT, Gorzelniak L, Leon ECD, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PloS One. 2013; 8(4): e61691.

156. Kayes NM, Schluter PJ, McPherson KM, Leete M, Mawston G, Taylor D. Exploring actical accelerometers as an objective measure of physical activity in people with multiple sclerosis. Archives of Physical Medicine & Rehabilitation. 2009; 90(4): 594-601.

157. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999; 8(2): 135-60.

158. Jean S, Richer L, Laberge L, Mathieu J. Comparisons of intellectual capacities between mild and classic adult-onset phenotypes of myotonic dystrophy type 1 (DM1). Orphanet Journal of Rare Diseases. 2014; 9(1): 186.

159. Grant PM, Ryan CG, Tigbe WW, Granat MH. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. British Journal of Sports Medicine. 2006; 40(12): 992-7.

160. Byrom B, Rowe DA. Measuring free-living physical activity in COPD patients: deriving methodology standards for clinical trials through a review of research studies. Contemporary Clinical Trials. 2016; 47: 172-84.

161. Innerd P, Catt M, Collerton J, Davies K, Trenell M, Kirkwood TB, et al. A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study. Age & Ageing. 2015: afv062.

162. Worm-Smeitink M, Gielissen M, Bloot L, van Laarhoven H, van Engelen B, van Riel P, et al. The assessment of fatigue: Psychometric qualities and norms for the Checklist individual strength. Journal of Psychosomatic Research. 2017; 98: 40-6.

163. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. Journal of Psychosomatic Research. 1994; 38(5): 383-92.

164. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. Trials. 2011; 12(1): 264.

165. Bland JM, Altman DG. Statistics notes: measurement error. BMJ. 1996; 313(7059): 744.

166. Holtebekk ME, Berntsen S, Rasmussen M, Jahnsen RB. Phyisical Activity and Motor Function in Children and Adolescents with Neuromuscular Disorders. Pediatric Physical Therapy. 2013; 25: 415-20.

167. Ollivier KH, Hogrel JY, Gomez-Merino D, Romero NB, Laforet P, Eymard B, Portero P. Exercise Tolerance and Daily Life in McArdle's Disease. Muscle & Nerve. 2005; 31: 637-41.

168. Shrader JA, Kats I, Kokkinis A, Zampieri C, Levy E, Joe GO, et al. A randomized controlled trial of exercise in spinal and bulbar muscular atrophy. Ann Clin Transl Neurol. 2015; 2(7): 739-47.

169. Wintzen AR, Lammers GJ, van Dijk JG. Does modafinil enhance activity of patients with myotonic dystrophy?: a double-blind placebo-controlled crossover study. Journal of Neurology. 2007; 254(1): 26-8.

170. Davidson ZE, Ryan MM, Kornberg AJ, Walker KZ, Truby H. Strong correlation between the 6-minute walk test and accelerometry functional outcomes in boys with Duchenne muscular dystrophy. J Child Neurol. 2015; 30(3): 357-63.

171. Elliott SA, Davidson ZE, Davies PS, Truby H. Accuracy of Parent-Reported Energy Intake and Physical Activity Levels in Boys With Duchenne Muscular Dystrophy. Nutrition in Clinical Practice. 2015; 30(2): 297-304.

172. Jeannet PY, Aminian K, Bloetzer C, Najafi B, Paraschiv-Ionescu A. Continuous monitoring and quantification of multiple parameters of daily physical activity in ambulatory Duchenne muscular dystrophy patients. Eur J Paediatr Neurol. 2011; 15(1): 40-7.

173. Aitkens S, Kilmer DD, Wright NC, McCrory MA. Metabolic syndrome in neuromuscular disease. Archives of Physical Medicine & Rehabilitation. 2005; 86(5): 1030-6.

174. Kilmer DD, Wright NC, Aitkens S. Impact of a home-based activity and dietary intervention in people with slowly progressive neuromuscular diseases. Archives of Physical Medicine & Rehabilitation. 2005; 86(11): 2150-6.

175. Apabhai S, Gorman GS, Sutton L, Elson JL, Plotz T, Turnbull DM, et al. Habitual Physical Activity in Mitochondrial Disease. PLoS One. 2011; 6(7): e22294.

176. Phillips M, Flemming N, Tsintzas K. An exploratory study of physical activity and perceived barriers to exercise in ambulant people with neuromuscular disease compared with unaffected controls. Clinical Rehabilitation. 2009; 22: 746-55.

177. Wiles CM, Busse ME, Sampson CM, Rogers MT, Fenton-May J, van Deursen R. Falls and stumbles in myotonic dystrophy. Journal of Neurology, Neurosurgery & Psychiatry. 2006; 77(3): 393-6.

178. McDonald CM, Widman LM, Walsh DD, Walsh SA, Abresch RT. Use of step activity monitoring for continuous physical activity assessment in boys with Duchenne muscular dystrophy. Archives of Physical Medicine & Rehabilitation. 2005; 86(4): 802-8.

179. McCrory MAKHLCASKD. Energy expenditure, physical activity, and body composition of ambulatory adults with hereditary neuromusuclar disease. Am J Clin Nutr. 1998; 67: 7.

180. Anens E, Emtner M, Hellstrom K. Exploratory study of physical activity in persons with Charcot-Marie-Tooth disease. Archives of Physical Medicine & Rehabilitation. 2015; 96(2): 260-8.

181. Favejee MM, van den Berg LE, Kruijshaar ME, Wens SC, Praet SF, Pim Pijnappel WW, et al. Exercise training in adults with Pompe disease: the effects on pain, fatigue, and functioning. Archives of Physical Medicine & Rehabilitation. 2015; 96(5): 817-22.

182. Kalkman JS, Schillings ML, Zwarts MJ, van Engelen BG, Bleijenberg G. The development of a model of fatigue in neuromuscular disorders: a longitudinal study. Journal of Psychosomatic Research. 2007; 62(5): 571-9.

183. Voet N, Bleijenberg, G.; Hendriks J, de Groot I, Padberg GW, van Engelen BGM, Geurts ACH. Both aerobic exercise and cognitive-behavioral therapy reduce chronic fatigue in FSHD. Neurology. 2014; 83: 9.

184. Busse ME. Quantified measurement of activity provides insight into motor function and recovery in neurological disease. Journal of Neurology, Neurosurgery & Psychiatry. 2004; 75(6): 884-8.

185. Rosenberg DE, Bombardier CH, Artherholt S, Jensen MP, Motl RW. Self-reported depression and physical activity in adults with mobility impairments. Archives of Physical Medicine & Rehabilitation. 2013; 94(4): 731-6.

186. Hawker GA, Ridout R, Harris VA, Chase CC, Fielding LJ, Biggar WD. Alendronate in the treatment of low bone mass in steroid-treated boys with Duchennes muscular dystrophy. Archives of Physical Medicine & Rehabilitation. 2005; 86(2): 284-8.

187. Elsworth C, Winward C, Sackley C, Meek C, Freebody J, Esset P, et al. Supported community exercise in people with long-term neurological conditions: a phase II randomized controlled trial. Clinical Rehabilitation. 2011; 25(7): 588-98.

188. McDonald CM, Widman LM, Walsh DD, Walsh SA, Abresch RT. Use of step activity monitoring for continuous physical activity assessment in boys with Duchenne muscular dystrophy. Archives of Physical Medicine & Rehabilitation. 2005; 86(4): 802-8.

189. Dillon CB, Fitzgerald AP, Kearney PM, Perry IJ, Rennie KL, Kozarski R, et al. Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: A cross-sectional study. PloS One. 2016; 11(5): e0109913.

190. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. Medicine & Science in Sports & Exercise. 2013; 45(11): 2193-203.

191. Matthews CE, Hagstromer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Medicine & Science in Sports & Exercise. 2012; 44(1 Suppl 1): S68-76.

192. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. Eur J Cardiovasc Prev Rehabil. 2010; 17: 127-39.

193. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the Assessment of Physical Acitvity: Cinical and Research Applications: a scientific statement from the American Heart Association. Circulation 2013; 128: 2259-79.

194. Afaq S, Tan S-T, Afzal U, Loh M, Dimarco A, Kooner J, Chambers J. 117 Validation of Accelerometers for Measurement of Physical Activity Energy Expenditure in South Asians and Europeans. Heart. 2014; 100(Suppl 3): A66-A7.

195. Thaler-Kall K, Tusker F, Hermsdörfer J, Gorzelniak L, Horsch A. Where to wear accelerometers to measure physical activity in people? Studies in Health Technology & Informatics. 2012; 192: 1045-.

196. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. Medicine & Science in Sports & Exercise. 2010; 42(2): 368.

197. Park J, Ishikawa-Takata K, Tanaka S, Mekata Y, Tabata I. Effects of walking speed and step frequency on estimation of physical activity using accelerometers. Journal of Physiological Anthropology. 2011; 30(3): 119-27.

198. Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. Medicine & Science in Sports & Exercise. 2007; 39(4): 716-27.

199. Kim DY, Jung YS, Park RW, Joo NS. Different location of triaxial accelerometer and different energy expenditures. Yonsei Medical Journal. 2014; 55(4): 1145-51.

200. Lee I-M, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. British Journal of Sports Medicine. 2014; 48(3): 197-201.

201. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Medicine & Science in Sports & Exercise. 2012; 44(1 Suppl 1): S68.

202. Scheeres K, Knoop H, Bleijenberg G. Clinical assessment of the physical activity pattern of chronic fatigue syndrome patients: a validation of three methods. Health & Quality of Life Outcomes. 2009; 7(1): 29.

203. van der Werf SP, Prins JB, Vercoulen JH, van der Meer JW, Bleijenberg G. Identifying physical activity patterns in chronic fatigue syndrome using actigraphic assessment. Journal of Psychosomatic Research. 2000; 49(5): 373-9.

204. Meeus M, Van Eupen I, Willems J, Kos D, Nijs J. Is the International Physical Activity Questionnaire-short form (IPAQ-SF) valid for assessing physical activity in Chronic Fatigue Syndrome? Disability & Rehabilitation. 2011; 33(1): 9-16.

205. Liu SH, Eaton CB, Driban JB, McAlindon TE, Lapane KL. Comparison of self-report and objective measures of physical activity in US adults with osteoarthritis. Rheumatology International. 2016; 36(10): 1355-64.

206. Harvill LM. Standard error of measurement. Educational Measurement: Issues & Practice. 1991; 10(2): 33-41.

207. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. Journal of Clinical Epidemiology. 1999; 52(9): 861-73.

208. Aitkens S, Kilmer DD, Wright NC, McCrory MA. Metabolic syndrome in neuromuscular disease. Archives of Physical Medicine & Rehabilitation. 2005; 86(5): 1030-6.

209. Klein P, James W, Wong W, Irving C, Murgatroyd P, Cabrera M, et al. Calorimetric validation of the doubly-labelled water method for determination of energy expenditure in man. Human Nutrition Clinical Nutrition. 1984; 38(2): 95-106.

210. Ollivier K, Hogrel JY, Gomez‐Merino D, Romero NB, Laforêt P, Eymard B, et al. Exercise tolerance and daily life in McArdle's disease. Muscle & Nerve. 2005; 31(5): 637-41.

211. Elliott SA, Davidson ZE, Davies PS, Truby H. Accuracy of parent-reported energy intake and physical activity levels in boys with duchenne muscular dystrophy. Nutrition in Clinical Practice. 2015; 30(2): 297-304.

212. Kilmer DD, Wright NC, Aitkens S. Impact of a home-based activity and dietary intervention in people with slowly progressive neuromuscular diseases. Archives of Physical Medicine & Rehabilitation. 2005; 86(11): 2150-6.

213. Barak S, Wu SS, Dai Y, Duncan PW, Behrman AL. Adherence to accelerometry measurement of community ambulation poststroke. Physical Therapy. 2014; 94(1): 101-10.

214. Nikolenko N, Jimenez-Moreno A, Gomes T, Wood L, Moat D, Newman J, et al. Myotonic dystrophy trial readiness. Neuromuscular Disorders. 2017; 27: S3.

215. Gourdon G, Meola G. Myotonic Dystrophies: State of the Art of New Therapeutic Developments for the CNS. Frontiers in Cellular Neuroscience. 2017; 11.

216. WHO. International Classification of Functioning, Disability and Health: ICF: World Health Organization; 2001.

217. Holsbeeke L, Ketelaar M, Schoemaker MM, Gorter JW. Capacity, capability, and performance: different constructs or three of a kind? Archives of Physical Medicine & Rehabilitation. 2009; 90(5): 849-55.

218. Fowler R, Jenkins S, Maiorana A, Gain K, O'Driscoll G, Gabbay E. Measurement properties of the 6‐min walk test in individuals with exercise‐induced pulmonary arterial hypertension. Internal Medicine Journal. 2011; 41(9): 679-87.

219. Bittner V. Role of the 6-minute walk test in cardiac rehabilitation. Cardiac Rehabilitation. 2007: 131-9.

220. Bellet RN, Adams L, Morris NR. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness—a systematic review. Physiotherapy. 2012; 98(4): 277-86.

221. Stewart T, Caffrey DG, Gilman RH, Mathai SC, Lerner A, Hernandez A, et al. Can a simple test of functional capacity add to the clinical assessment of diabetes? Diabetic Medicine: a journal of the British Diabetic Association. 2016; 33(8): 1133-9.

222. Bautmans I, Lambert M, Mets T. The six-minute walk test in community dwelling elderly: influence of health status. BMC Geriatrics. 2004; 4: 6.

223. Baldasseroni S, Mossello E, Romboli B, Orso F, Colombi C, Fumagalli S, et al. Relationship between cognitive function and 6-minute walking test in older outpatients with chronic heart failure. Aging Clinical & Experimental Research. 2010; 22(4): 308-13.

224. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. Developing a protocol for observational comparative effectiveness research: a user's guide: Government Printing Office; 2013.

225. McDonald CM, Henricson EK, Abresch RT, Florence J, Eagle M, Gappmaier E, et al. The 6‐minute walk test and other clinical endpoints in duchenne muscular dystrophy: Reliability, concurrent validity, and minimal clinically important differences from a multicenter study. Muscle & Nerve. 2013; 48(3): 357-68.

226. Florence J, Van Der Ploeg A, Clemens P, Escolar D, Laforet P, Rosenbloom B, et al. TP 1.01 Use of the 6min walk test as an endpoint in clinical trials for neuromuscular diseases. Neuromuscular Disorders. 2008; 18(9): 738-9.

227. Dunn A, Marsden D, Nugent E, Van Vliet P, Spratt N, Attia J, et al. Protocol variations and six-minute walk test performance in stroke survivors: a systematic review with meta-analysis. Stroke Research & Treatment. 2015; 2015.

228. Singh SJ, Puhan MA, Andrianopoulos V, Hernandes NA, Mitchell KE, Hill CJ, et al. An official systematic review of the European Respiratory Society/American Thoracic Society: measurement properties of field walking tests in chronic respiratory disease. The European Respiratory Journal. 2014; 44(6): 1447-78.

229. Wu G, Sanderson B, Bittner V. The 6-minute walk test: how important is the learning effect? American Heart Journal. 2003; 146(1): 129-33.

230. Montes J, McDermott M, Martens W, Dunaway S, Glanzman A, Riley S, et al. Six-Minute Walk Test demonstrates motor fatigue in spinal muscular atrophy. Neurology. 2010; 74(10): 833-8.

231. Andersen LK, Knak KL, Witting N, Vissing J. Two-and 6-minute walk tests assess walking capability equally in neuromuscular diseases. Neurology. 2016; 86(5): 442-5.

232. Vaidya T, de Bisschop C, Beaumont M, Ouksel H, Jean V, Dessables F, et al. Is the 1-minute sit-to-stand test a good tool for the evaluation of the impact of pulmonary rehabilitation? Determination of the minimal important difference in COPD. International Journal of Chronic Obstructive Pulmonary Disease. 2016; 11: 2609-16.

233. Zanini A, Aiello M, Cherubino F, Zampogna E, Azzola A, Chetta A, et al. The one repetition maximum test and the sit-to-stand test in the assessment of a specific pulmonary rehabilitation program on peripheral muscle strength in COPD patients. International Journal of Chronic Obstructive Pulmonary Disease. 2015; 10: 2423-30.

234. Fry DK, Pfalzer LA. Reliability of four functional tests and rating of perceived exertion in persons with multiple sclerosis. Physiotherapy Canada. 2006; 58(3): 212-20.

235. Ozalevli S, Ozden A, Itil O, Akkoclu A. Comparison of the Sit-to-Stand Test with 6min walk test in patients with chronic obstructive pulmonary disease. Respiratory Medicine. 2007; 101(2): 286-93.

236. Reychler G, Boucard E, Peran L, Pichon R, Le Ber-Moy C, Ouksel H, et al. One minute sit-to-stand test is an alternative to 6MWT to measure functional exercise performance in COPD patients. The Clinical Respiratory Journal. 2017.

237. Beaumont M, Boucard E, Peran L, Pichon R, Liistro G, Le Ber-Moy C, et al. Comparison between six minutes walking test and sit-to-stand test in COPD patients. Eur Respiratory Soc; 2015.

238. Lord SR, Murray SM, Chapman K, Munro B, Tiedemann A. Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. The Journals of Gerontology Series A, Biological Sciences & Medical Sciences. 2002; 57(8): M539-43.

239. Applebaum EV, Breton D, Feng ZW, Ta A-T, Walsh K, Chassé K, et al. Modified 30-second Sit to Stand test predicts falls in a cohort of institutionalized older veterans. PLoS One. 2017; 12(5): e0176946.

240. Holtebekk ME, Berntsen S, Rasmussen M, Jahnsen RB. Physical activity and motor function in children and adolescents with neuromuscular disorders. Pediatric Physical Therapy. 2013; 25(4): 415-20.

241. Cervantes CM, Porretta DL. Physical activity measurement among individuals with disabilities: A literature review. Adapted Physical Activity Quarterly. 2010; 27(3): 173-90.

242. Fairclough S, Noonan R, Rowlands A, Van Hees V, Knowles Z, Boddy L. Wear compliance and activity in children wearing wrist and hip mounted accelerometers. Medicine & Science in Sports & Exercise. 2016; 48(2): 245-53.

243. Gallais B, Gagnon C, Mathieu J, Richer L. Cognitive decline over time in adults with myotonic dystrophy type 1: A 9-year longitudinal study. Neuromuscular Disorders. 2017; 27(1): 61-72.

244. Wade A, Pan H, Eaton S, Pierro A, Ong E. An investigation of minimisation criteria. BMC Medical Research Methodology. 2006; 6(1): 11.

245. Coster WJ. Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies. The American Journal of Occupational Therapy. 2013; 67(2): 162-70.

246. Geroin C, Mazzoleni S, Smania N, Gandolfi M, Bonaiuti D, Gasperini G, et al. Systematic review of outcome measures of walking training using electromechanical and robotic devices in patients with stroke. Journal of Rehabilitation Medicine. 2013; 45(10): 987-96.

247. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open. 2013; 3(1): e001570.

248. Austin PC, Brunner LJ. Type I error inflation in the presence of a ceiling effect. The American Statistician. 2003; 57(2): 97-104.

249. NICE. Falls: the assessment and prevention of falls in older people (Clinical Guidance CG21): National Institute for Health & Care Excellence; 2004. www.nice.org.uk/guidance/CG21 accessed on March 20th of 2017. This guideline has been updated and replaced by NICE guideline CG161.