Feasible and clinical relevant outcome measures for adults with mitochondrial disease

Peggy M.J. Berge, Daphne M. Maas, Mirian C.H. Janssen, Jan T. Groothuis

Abstract

There is no consensus on clinical outcome measures that reflect function, activities and participation which are suitable for adults with mitochondrial diseases (MD). The aim of this study was to determine feasible and clinically relevant outcome measures for patients with MD. In 156 adult patients with MD, endurance, balance, strength and mobility tests were evaluated. All tests showed a negative deviation to healthy reference values. Balance tests were feasible and significantly correlated with clinical severity. The Astrand cycle test was not feasible in 55%, whereas the feasibility of the 6 min walking test is unclear in patients with MD.

1. Introduction

Mitochondrial diseases (MD) are the most prevalent inherited metabolic diseases, with an incidence of approximately 1 in 5000 live births [1]. They originate in mitochondrial dysfunction, often as a consequence of mutations in the mitochondrial (mtDNA) or nuclear DNA. Since the mitochondria are responsible for energy production, a defect can cause clinical manifestations in any organ or tissue throughout the human body. The organs that are most dependent on the production of energy by the mitochondria, such as the brain and skeletal muscles, are most frequently and most severely affected [2]. Currently, no treatment is available, and therefore the focus of treatment of patients with MD is on support, and improving function, activities and participation with the help of physical and occupational therapy and adequate nutrition [3]. Since the most common symptoms of MD are muscle weakness, pain and fatigue, physical therapy evaluation focusses on strength, range of motion, balance and coordination, posture, motor function and endurance [4].

2. Methods

Adults (≥ 18 years) diagnosed with MD caused by mutations in mtDNA or nuclear DNA, who participated in a multidisciplinary admission called “mitostreet” within the Radboud Centre for Mitochondrial Medicine (RCMM) between January 2015 and March 2020 were included in this study.
This study is approved and received granted permission by the local ethical committee (CMO-file number 2017–3687). All subjects participated voluntarily and had signed written informed consent.

2.1. Procedure

Mitostreet is a 4 day admission within the RCMM at the Radboud university medical center, Nijmegen, the Netherlands, in which patients’ organ systems, functioning, activities and participation are structurally evaluated and monitored. Consultation of different medical specialties took place, including physical therapy. The physical therapist assessed the admitted patients with MD in a standardized protocol on different domains. Mini-BESTest a number of patients, it was not possible to perform all tests due to physical reasons and time constraints. The reasons for not performing the test were listed in the medical record.

2.2. Instruments

**Endurance** was measured using three different tests:

- Åstrand cycle test is a submaximal cycle ergometer aerobic fitness test involving the relation between heart rate and percentage of maximal aerobic capacity to calculate maximal oxygen consumption (VO₂max (ml/kg/min)) [6]. Patients were asked to maintain a cycling rate of 55–60 rpm during six minutes. The goal is to obtain two consecutive heart rate values exceeding 120 bpm, during the fifth and sixth minute. After warming up, the workload is adjusted to bring the heart rate to within the desired range. During the last three minutes, the workload was kept constant in order to achieve a steady-state heart rate.

- Six minute walking test (6MWT) is a submaximal aerobic fitness test to assess walking endurance and aerobic capacity. Patients walk in a self-selected past as far as they can in 6 min, measured in meters [7].

- The 30 s sit to stand test (30SSTS) can measure functional mobility lower extremity strength beside endurance. Patients are encouraged to complete as many full stands as possible within 30 s without using their arms [8].

**Balance** was measured using three different tests:

- Berg Balance Scale (BBS) is an objective test to assess static balance and gait risk in adults during a series of predetermined tasks, consisting of 14 items rated on a scale from 0 to 2, maximal 28 points [9].

- Performance-Oriented Mobility assessment (POMA) assess two domains, namely balance and gait. Balance includes sitting and standing balance, and gait includes symmetry and continuity. Gait has 12 and balance 16 items rated on a scale from 0 to 2, maximal 28 points [10].

- Mini Balance Evaluation Systems Test (Mini-BESTest) assess four domains, anticipatory posture, reactive posture control, sensory orientation and gait. A total of 14 items rated on a scale from 0 to 2, maximal 28 points [11].

**Strength** was measured using four different tests:

- Jamar handgrip test is used to measure grip strength in Newton using a hand-held dynamometer. Patients are asked to squeeze as hard as possible with the device in their hand [12].

- Pinch test (two-point pinch) is used to measure strength in Newton of the index finger and thumb. Patients pinch as hard as possible on the device [13,14].

- We used Hand-held dynamometry (HHD) and Manual Muscle Testing (MMT) to assess the strength of the biceps and knee flexors as representation of upper and lower extremity strength. For HHD muscle strength measurement (in Newton) the physical therapist holds the hand-held dynamometer (Citec, CT3002) and gives resistance, the patient is instructed and encouraged to push as hard as possible against the device [15,16]. MMT is a standardized set of assessments to measure muscle strength and function against specific criteria with a Medical Research Council Scale (MRC-scale ranging from 0 to 5) [17].

**Mobility** was measured using the motor function measure (MFM) which consist of 32 task items, rated on a 4 point likert scale, in 3 dimensions that provide a detailed profile of the physical impairment: D1, standing and transfers; D2, axial and proximal motor function; and D3, distal motor function [18]. Items in this test, especially stand and transfers, overlap with balance tests.

**Genotype** of mtDNA was assessed in terms of heteroplasmy levels in leukocytes, urine epithelial cells and in buccal mucosa [19]. Genotype of nuclear DNA was determined by Polymerase Chain Reaction (PCR) and sequence analysis of the gene, or by whole exome sequencing.

**Disease severity** was assessed by the Newcastle mitochondrial disease adult scale (NMDAS). The NMDAS is a validated method for both measuring and monitoring disease manifestation and clinical features in patients with MD and is established out of three sections [19].

**Section I: Current functioning**, consists of ten questions using clinical history to assess functioning of individual organ systems. **Section III, Current clinical assessment**, consists of nine questions using clinical history to assess functioning of individual organ systems. **Section III, Current clinical assessment**, consists of ten questions, a current clinical and neurological examination and three tests on cognition: a symbol search test, a reading test and a speed of comprehension test. All questions can be scored from 0 to 5, with 0 meaning no involvement and 5 meaning severe involvement.

2.3. Reference values

Reference values were used to assess whether patients with MD deviate from the healthy population. Literature search using Pubmed was performed for reference values in healthy individuals, corrected for age and sex. Reference values for Astrand [20], 6MWT [7], HHD [15,16], Jamar [12] and pinch [13] were found. Healthy individuals are able to perform 20 or more complete stands during the 30STS [8]. Maximum scores for MFM (100% [18]), MMT [21], BBS (58 points [22]), POMA (28 points [23]) and Mini-BESTest (28 points [24]) were used.

2.4. Statistical analysis

Feasibility was calculated as percentage of patients able to perform the test. Ad hoc, we divided patients in two disease severity groups (NMDAS ≤10 and > 10) to assess differences in deviation from reference values with an independent t-test. Correlation between NMDAS and deviation from reference value was calculated with Pearson’s correlation coefficient. Correlations were considered as weak (>0.3), moderate (>0.5) or strong (>0.7). Ad hoc, we divided patients into three groups of percentage blood heteroplasmy (not available, 0–20%, 21–69%, and 70–100%) or mutation type (m.3243A > G and other mutation). All analyses were performed using IBM SPSS Statistics v25 (SPSS Inc., Chicago, Illinois, United States). A p-value of <0.05 was considered to represent statistical significance. Bonferroni’s correction was used to correct for multiple comparisons, in this case p-value of <0.01 was considered statistically significant. Clinical relevance of the tests was suggested if there was a significant correlation between disease severity (NMDAS) and the deviation from the reference values.

3. Results

Data from 156 patients with MD was obtained of which 60% was female, the average age was 42 years. In 62% of the study population the m.3243A > G mutation was the cause of MD. Patients characteristics are shown in Table 1.
were not able to perform the 30SSTS, due to exhaustion from previous
dently for 6 min (10%, n = 6), other reasons were not stated. The 30SSTS
son for not performing the 6MWT was not being able to walk indepen-
perform an Åstrand test and 28 patients performed both tests. The rea-
was also not able to perform a 6MWT. In total, we successfully
Additionally, out of these 86 patients more than half (62%, n = 53)
of betablockers (3%, n = 3), and 4) unknown (78%, n = 67).
were: 1) not being able to cycle against resistance (13%, n = 11),
Ástrand test or calculate the maximal oxygen consumption (VO2Max)
people was able to perform a 30SSTS (39%, n = 61) performed POMA and Mini-BESTest
41%, n = 64). The main reason for not performing all balance tests
to time restraints within the physical therapy consultation. The majority of patients performed HHD(62%, n = 93) and in a vast major-
Some patients endured pain during one or both of these tests (15%,
especially HHD caused pain and pressure on the muscles
POMA and Mini-BESTest score compared to NMDAS<10. Other tests showed no signifi-
other mutation
m.3243A > G
m.9155A > G
m.13513G > A
m.7898 T > C
MELAS
m.3243A > G
m.3271 T > C
m.3291 T > C
m.9176 T > C
CPEO
m.3243A > G
Twinkle
POLG1
mtDNA deletion
MERRF
m.3243A > G
m.8334A > G
m.8363G > A
NARP (m.8993 T > G)
Myopathy
m.3243A > G
m.10015 T > C
m.12287 T > C
OPA1
Other
m.3243A > G
OPA1
m.7471insC
MTAT1
m.7405A > G
gtpbp-3
NDUFS7

3.1. Feasibility

Many patients (55%, n = 86) were not able to perform an Åstrand cycle test. The most frequent reasons for not being able to perform the Åstrand test or calculate the maximal oxygen consumption (VO2Max) were: 1) not being able to cycle against resistance (13%, n = 11), 2) terminating prematurely due to exhaustion (6%, n = 5), 3) use of betablockers (3%, n = 3), and 4) unknown (78%, n = 67). Additionally, out of these 86 patients more than half (62%, n = 53) was also not able to perform a 6MWT. In total, we successfully performed a 6MWT in 61 patients (39%), 33 patients after not able to perform an Åstrand test and 28 patients performed both tests. The rea-

3.2. Comparison to healthy reference values

All results from the performed tests showed a negative deviation compared to healthy reference values, age and sex-matched where applicable (Table 2). The 6MWT (−24.1%) and 30SSTS (−39.5%) had both a significant (p < 0.01) larger percentage negative deviation from the healthy reference value compared to the Åstrand cycle test (−12.2%). Pinch test had for both the right (−54.0%) and left (−56.1%) hand a signifi-
cantly larger negative deviation (p < 0.01) from healthy reference values compared to the Jamar test. HHD of the biceps muscle deviated significantly (p < 0.01) more from healthy reference value than HHD of the knee flexor muscles. Manual strength testing of the biceps muscle and knee flexor muscle was on average 0.2–0.4 points lower than the maximum score of 5.

3.3. Comparison with NMDAS

Patients with NMDAS>10 had a significantly (p < 0.01) lower BBS, POMA and Mini-BESTest score compared to NMDAS<10. Other tests showed no significant difference between NMDAS<10 and NMDAS>10.

3.4. Correlation with NMDAS

BBS, POMA and Mini-BESTest had significant weak correlations with NMDAS (r: −0.349, p = 0.01; r: −0.425, p = 0.01, respectively) whereas the Mini-BESTest had a significant strong correlation with NMDAS (r: −0.766, p < 0.01) (Table 4, Fig. 1). Comparably, knee flexion strength measured with MMT had significant weak correlations with NMDAS (r: −0.300, p = 0.01) (Table 4). There were no significant correlations between Åstrand cycle test, 6MWT, 30SSTS, Jamar, pinch test and MFM with NMDAS (Table 4, Fig. 1).

3.5. Correlation with NMDAS in mutation groups

Mini-BESTest had a significant strong correlation with NMDAS for ‘m.3243A > G mutation’ (r: −0.895, p < 0.01) but not for ‘other muta-
tion’ (r: −0.546, p = 0.03). MFM had a significant weak correlation for ‘m.3243A > G mutation’ with NMDAS (r: −0.314, p = 0.01) but not for ‘other mutation’ (Table 4). Other tests showed no significant correla-
tions in both mutation groups (Table 4).

3.6. Correlation with NMDAS in heteroplasmy groups

Mini-BESTest had significant (p < 0.01) moderate to strong correlations with NMDAS in not available, 0–20% and 21–69% blood heteroplasmy group, unfortunately a correlation could not be calculated in the 70–100% blood heteroplasmy group because of the low number of subjects. All other tests showed no significant correlations in all four blood heteroplasmy groups (Table 4).
Table 2

<table>
<thead>
<tr>
<th>Test</th>
<th>Subjects, n</th>
<th>Mean ± SEM</th>
<th>Deviation from the reference value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6MWT</td>
<td>70</td>
<td>92.7 ± 0.9</td>
<td>−7.3 ± 0.9</td>
</tr>
<tr>
<td>30STS</td>
<td>61</td>
<td>24.0 ± 1.7</td>
<td>2.0 ± 0.1</td>
</tr>
<tr>
<td>POMA</td>
<td>64</td>
<td>39.6 ± 2.4</td>
<td>7.3 ± 0.9</td>
</tr>
<tr>
<td>BBS</td>
<td>61</td>
<td>7.3 ± 0.9</td>
<td>0.2 ± 0.1</td>
</tr>
<tr>
<td>Mini-BESTest</td>
<td>64</td>
<td>26.1 ± 0.4</td>
<td>4.6 ± 0.5</td>
</tr>
<tr>
<td>HHD biceps</td>
<td>64</td>
<td>25.3 ± 2.6</td>
<td>4.6 ± 0.5</td>
</tr>
<tr>
<td>HHD knee</td>
<td>61</td>
<td>27.9 ± 2.5</td>
<td>4.6 ± 0.5</td>
</tr>
</tbody>
</table>

Values are mean ± SEM. Deviation from healthy reference values in percentage, except for POMA, BBS, Mini-BESTest and MMT. Åstrand in VO2max (ml/kg/min); 6MWT, 6-min walking test in meters; 30SSTS, 30 s sit to stand test in number of times; POMA, Performance-Oriented Mobility Assessment in points (maximal 28 points); BBS, Berg Balance Scale in points (maximal 28 points); Mini-BESTest, Mini Balance Evaluation Systems Test in points (maximal 28 points); HHD, Hand-held dynamometer in Newton; MMT, Manual muscle testing (MRC scale 0–5); MFM, Motor Function Measure. R, right side; L, left side.

4. Discussion

This is the first study that structurally investigated the feasibility and clinical relevance of functional outcome measures in daily practice in patients with MD. It demonstrates that patients with MD scored overall lower on all performed tests compared to healthy individuals. Balance tests (BBS, POMA, and Mini-BESTest) have the best feasibility, the Mini-BESTest has the strongest correlation with disease severity in patients with MD. Pinch, Jamar, HHD, MMT and MFM are feasible in patients with MD, however, they did not correlate to disease severity. The endurance tests (Åstrand, 6MWT and 30SSTS) were not feasible in most patients due to several, mostly MD related, reasons. Even though these tests demonstrated lower values compared to healthy individuals, the clinical relevance is not clear yet in patients with MD. Based on the results of the present study we propose a toolbox in patients with MD to minimally include the Mini-BESTest.

4.1. Feasibility

The Åstrand submaximal exercise cycle test was not feasible in more than half of the patients with MD. This is coherent to the feasibility of this test in patients with Parkinson's disease [6]. Even though the test was found to be safe to perform in patients with MD, Steel et al. already questioned the ability of patients with cardiac involvement and significant intellectual or physical disabilities to exercise at the level required for a valid test [5]. Studies that include VO2max testing in patients with MD [5,25,26] include fewer patients and also patients with less severe symptoms. This could explain why in this study the feasibility is low in patients with MD. In our study, heart rate did not increase enough to calculate VO2max in several patients who used beta blockers, and patients with lower extremity weakness were not able to cycle against required resistance. Due to the low feasibility of this test it is not suitable for clinical outcome measures in patients with MD.
In community dwelling elderly 43.5% was unable to perform a 6MWT [27], similar to our results (44.8% of patients). In patients with myopathy, dystrophy and proximal weakness the 6MWT was feasible, however in these studies performing a 6MWT was an inclusion criteria, suggesting a selection bias [28–30]. It must be noted that in our population in most cases the 6MWT was performed in patients who were not able to perform the Åstrand cycle test. Therefore it is not completely clear what the feasibility the 6MWT would have been if the test was performed in all included patients. Even though a large portion of the patients was unable to perform the 6MWT, the feasibility is higher than that of the Åstrand cycle test. Therefore it is suggested that the 6MWT is a better outcome measure to indicate endurance in patients with MD.

The feasibility of the 30SSTS cannot completely be assessed, because the most important reason for not performing the test was already experiencing fatigue from previous tests. This might indicate that patients are able to perform the test but are not able to perform the test when it is integrated with different test within one consultation. A possible alternative is a five-repetition sit-to-stand test which showed a higher test-retest reliability than the 30SSTS test in patients with dementia, stroke and cerebral palsy [31] and has been found feasible in a preliminary evaluation in 18 patients with MD [32]. The feasibility of the 30SSTS in patients with MD should be tested when there are more breaks between tests so patients are less exhausted. If these breaks do not increase feasibility it could be suggested to integrate the five-repetition sit-to-stand test instead.

Since muscle pain, fatigue and exercise intolerance are the most common symptoms in patients with MD [4], endurance test should cautiously combined and limited in quantity. We therefore suggest to use the 6MWT in combination with the 30SSTS as outcome measures of endurance in clinical evaluation of patients with MD.

BBS, POMA and Mini-BESTest are equally feasible in patients with MD, and all tests are previously used in comparable patient populations [11,33–35]. Similarly, the pinch and Jamar strength test are feasible in patients with MD and are both used previously in patients with MD, or comparable neuromuscular diseases [36–38]. MMT is used in medication trials for patients with MD [39] and in the present study found feasible as outcome measure in a clinical setting. We found that HHD was less feasible due to experienced pain caused by the applied pressure of the HHD on the muscles. In contrast with previous study in inflammatory myopathy that found equal feasibility for HHD and MMT [38]. Possible explanations for the difference in feasibility of DDH is that the experienced pain could be due to already existing muscle pain or low body weight in patients with MD. With MMT a lower pressure is applied on the muscle itself, thereby resulting in less or no experienced pain.

MFM is developed for patients with neuromuscular disease and in this study found feasible for patients with MD as well [18].

### 4.2. Correlation with NMDAS

This study shows that Åstrand cycle test does not correlate with disease severity, calculated with NMDAS, in patients with MD, although calculated VO2max was lower compared to healthy reference values. Previously, a correlation between muscle mutation load and VO2max, regardless of the mtDNA mutation in patients with MD was demonstrated [40,41]. However, there was no correlation between blood mutation load and VO2max, similar to the results found in this study. The remaining question is whether aerobic capacity determined using an Åstrand cycling test can be considered a useful outcome measure in patients with MD.

6MWT is used as a predictor of aerobic fitness and was found to have a moderate correlation with Åstrand cycle test in patients with musculoskeletal conditions [42]. Therefore, the 6MWT was recommended as a core set of reliable and valid measures to assess health related physical fitness [42]. However, in the current study there was no correlation...
found between disease severity and 6MWT. Similar, in patients with Parkinson's disease, there was no correlation found between 6MWT and the disease severity \([30]\). A possible explanation for the lack of correlation between disease severity and 6MWT in patients with MD is that aerobic capacity and endurance is only one single item within the NMDAS \([19]\). For this reason, it can be suggested to perform an endurance tests, such as a 6MWT, besides NMDAS to gain more complete insight in disease severity in patients with MD.

The BBS, POMA and Mini-BESTest are balance tests designed for elderly and patients with stroke to indicate fall risk \([43,44]\). For patients with stroke the fall risk can also indicate the severity of the stroke. All three test show reliability and validity to measure balance ability in these populations. However, the BBS showed a floor- and ceiling effect and a low validity for dynamic balance \([44]\). This is the first study to investigate the correlation between the disease severity and the BBS, POMA and Mini-BESTest. This study shows a significant correlation between all three balance tests and disease severity measured with NMDAS, with a weak correlation for the BBS and POMA and a strong correlation for the Mini-BESTest. Therefore these test can be stated as clinically relevant in patients with MD, with a preference for the Mini-BESTest.

Tveter et al. \([42]\) found that the 30SSTS only displayed acceptable validity but not reliability in patients with various musculoskeletal conditions. However, their studied population was younger which meant a higher number of sit to stands, and they expected reliability to increase as the number of sit to stands decreases \([42]\). The current study in patients with MD showed no correlation with disease severity. However, patients with MD did show a large negative deviation from healthy reference values. This together with a presumably low reliability makes it still unclear whether the 30SSTS is clinically relevant in patients with MD.

### 4.3 Limitations/strengths of the study

Data collection was part of regular clinical patient care and therefore represented actual performance and capability. However, due to the retrospective nature and data collection being part of regular care, not all data was complete. Especially, not all reasons for not being able to perform tests were listed. The sequence of tests was standardized but included tests changed over time, which is common in clinical practice. Nevertheless, performance of all tests was standardized and assessed by trained an experienced physical therapists dedicated to MD patient care. A large population of patients with MD was included in this study. In almost all patients the mutation and heteroplasmy level was known and NMDAS was assessed in all patients, therefore the patient group was well characterized. Overall is this the first study that structurally investigated outcome measures in daily clinical practice in patients with MD on the physical therapy domain.

### 5. Conclusion

We demonstrate that especially balance tests are feasible and clinically relevant as outcome measures in daily clinical practice on the physical therapy domain in patients with MD. With a strong correlation for the Mini-BESTest to disease severity measured with NMDAS. The Åstrand cycle test is not feasible and feasibility of the 6MWT and 30SSTS is still unclear in patients with MD. Strength tests are feasible but do not convincingly correlate with disease severity. Therefore, we suggest more detailed investigation for optimal assessment of aerobic capacity and strength in patients with MD in both clinical practice and as potential outcome measure future studies and trials. Based on this study we would recommend a toolbox for clinical evaluation of patients with MD to minimally include the Mini-BESTest.
muscular diseases (EURO-NMD).

Center (NL-NMD) and the European Reference Network for rare neuro-

cides in the public, commercial, or not-for-pro

the main complaint of most patients with MD. Whether these tests can detect improvement or deterioration of disease

tance of other endurance tests, since decreased exercise tolerance is


[38] D.K. Van Der Stap, L.G. Rider, H. Alexanderson, A.M. Huber, B. Gualano, P. Gordon,


Based on current literature and symptoms of daily life in patients with MD we would recommend to also include the 6MWT, 30SSTS, Pinch test and MMT within the toolbox. Due to exhaustion of patients during testing it is recommended that sufficient resting time is taken between tests.

Future research should focus on repeated measures of all these tests in patients with MD to determine reliability and to give more insight in whether these tests can detect improvement or deterioration of disease severity. Research should also focus on the feasibility and clinical relevance of other endurance tests, since decreased exercise tolerance is the main complaint of most patients with MD.

Declaration of Competing Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: none.

Acknowledgements

The authors would like to thank Tamara Popping, Judith Kanters and Elke Buit for their contributions in the study, as well as all individuals participated in this study.

All authors of this publication are members of the Radboud Center for Mitochondrial Medicine (RCMM), Netherlands Neuromuscular Center (NL-NMD) and the European Reference Network for rare neuro-muscular diseases (EURO-NMD).

References


[7] C. Bérad, C. Payan, I. Hodgkinson, J. Herman, Group MCS, A motor function mea-


[15] R.W. Bohannon, Reference values for extremity muscle strength obtained by hand-


[5] E. Horst, Astrand Test: A Comparison between Patients with Parkinson's Disease and


[8] C.J. Jones, R.W. Bohannon, Reference values for extremity muscle strength obtained by hand-


[16] G. Meintjes, Performance-oriented assessment of mobility problems in elderly pa-


[22] F. Franchignoni, F. Horak, M. Godi, A. Nardone, A. Giordano, Using psychometric


[24] J.P. Pichon, N. Witting, J. Vissing, Decreased variability of the 6-minute walk test by

[25] W. Durfee, P. Iazzolo, Rehabilitation and muscle testing, Encyclopedia of Medical De-


[27] A.J. Videler, A. Beelen, G. Aufdemkampe, I.J. de Groot, M. Van Leemputte, Hand

[28] R. Whittaker, E. Ferenczi, D. Hilton-Jones, Myotonic dystrophy: practical issues relat-


[34] S. Tyson, L. Connell, How to measure balance in clinical practice. A systematic review of the psychometrics and clinical utility of measures of balance activity for neurolog-


