ORIGINAL REPORT

Check for updates

RESPONSIVENESS OF PERFORMANCE-BASED OUTCOME MEASURES FOR MOBILITY, BALANCE, MUSCLE STRENGTH AND MANUAL DEXTERITY IN ADULTS WITH MYOTONIC DYSTROPHY TYPE 1

Marie KIERKEGAARD, PT, PhD[1], Émilie PETITCLERC, PT, MSc[2], Luc J. HÉBERT, PT, PhD[3], Jean MATHIEU, MD, MSc[2] and Cynthia GAGNON, OT, PhD[2]

From the [1]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, and Function Area Occupational Therapy & Physiotherapy, Karolinska University Hospital, Stockholm, Sweden, [2]Faculty of Medicine and Health Sciences, Centre de recherché Charles Le-Moyne, Université de Sherbrooke, Sherbrooke, and [3]Faculty of Medicine, Rehabilitation (Physiotherapy) and Radiology/ Nuclear Medicine Department, Université Laval, Québec City, Québec, Canada

Objective: To assess changes and responsiveness in outcome measures of mobility, balance, muscle strength and manual dexterity in adults with myotonic dystrophy type 1.

Design: A 9-year longitudinal study conducted with 113 patients.

Methods: The responsiveness of the Timed Up and Go test, Berg Balance Scale, quantitative muscle testing, grip and pinch-grip strength, and Purdue Pegboard Test was assessed using criterion and construct approaches. Patient-reported perceived changes (worse/stable) in balance, walking, lower-limb weakness, stair-climbing and hand weakness were used as criteria. Predefined hypotheses about expected area under the receiver operating characteristic curves (criterion approach) and correlations between relative changes (construct approach) were explored.

Results: The direction and magnitude of median changes in outcome measures corresponded with patient-reported changes. Median changes in the Timed Up and Go test, grip strength, pinch-grip strength and Purdue Pegboard Test did not, in general, exceed known measurement errors. Most criterion (72%) and construct (70%) approach hypotheses were supported. Promising responsiveness was found for outcome measures of mobility, balance and muscle strength. Grip strength and manual dexterity measures showed poorer responsiveness.

Conclusion: The performance-based outcome measures captured changes over the 9-year period and responsiveness was promising. Knowledge of measurement errors is needed to interpret the meaning of these longitudinal changes.

Key words: myotonic dystrophy; psychometrics; rehabilitation; responsiveness; outcome measures.

Correspondence address: Marie Kierkegaard, Function Area Occupational Therapy & Physiotherapy, A6: U1 Karolinska University Hospital, SE-171 76, Stockholm, Sweden. E-mail: marie.kierkegaard@sll.se, marie.kierkegaard@ki.se

Myotonic dystrophy type 1 (DM1) is a slowly progressive, inherited, neuromuscular disease characterized by muscle weakness and myotonia. Besides impairments in muscle function, other body functions and structures are affected, and DM1 is therefore recognized as a multi-system disorder. Clinical phenotypes of DM1 can be described based on age at onset, symptoms, and the unstable expansion of trinucleotide (CTG) repeats causing the disease (1). In the adult-onset phenotype, clinical signs generally appear between the ages of 20 and 40 years, whereas the late-onset phenotype appears at an older age (>40 years) with fewer and milder symptoms.

No curative treatment is currently available for DM1, although potential treatments are being developed (2). Concerns have been raised about outcome measures with good measurement properties to assess the effects of interventions (3). Outcome measures that reflect mobility, balance, muscle strength and manual dexterity will be needed to assess endpoints in future trials. There is a lack of knowledge, however, of the measurement properties, especially concerning responsiveness, of such measures in the DM1 population (4).

Responsiveness can be defined as "the ability of an instrument to detect change over time in the construct to be measured" (5), and thus refers to the validity of a change score (6). Two different approaches are proposed for the assessment of responsiveness. The criterion approach is used if a gold standard is available, e.g. a global rating scale capturing the impression of change in the construct of interest (7). *A priori* hypotheses should be made about the direction and strength of the relationship between changes in the criterion and changes in the outcome measures. Such relationships can, if the criterion is a dichotomous variable, be assessed with receiver operating characteristic (ROC) curve analyses. The size of the area under the curve (AUC) in ROC analyses indicates the ability of the outcome measure to distinguish between those who have/have not changed according to the criterion (7). The construct approach relies on hypothesis testing about expected directions and magnitudes of correlations between change scores in outcome measures measuring similar or different constructs (7). It is important that there is a rationale for these *a priori* formulated hypotheses, e.g. previous research findings and/or clinical experience.

A longitudinal natural-history study of functioning, disability and health in people with the adult-onset and late-onset phenotypes of DM1 was initiated in 2002 at the Saguenay Neuromuscular Clinic (Quebec, Canada), providing an opportunity to explore the responsiveness of various outcome measures. Thus, the aims of the present study were to describe and assess changes and responsiveness in outcome measures for mobility, balance, muscle strength and manual dexterity in adults with DM1.

METHODS

Participants

Adults with a genetically DNA-confirmed diagnosis of DM1 and presenting with the adult-onset or late-onset phenotype were recruited from the Saguenay Neuromuscular Clinic (Quebec, Canada). As part of a funded longitudinal study, 200 patients with DM1 were included at baseline (T1) (8–18) and were invited to attend a follow-up 9 years later (T2) (19–20). At T2 59 of the 200 patients had died, 8 had left the region, 4 were excluded due to severe cognitive impairment, and 14 declined to participate (for personal reasons or lack of interest). Thus, 115 patients with DM1 were included at T2 (i.e. 58% of participants from T1). All patients provided signed informed consent before enrolment at both baseline and follow-up.

Protocol

The same procedures for assessments were used at both time-points. One physiotherapist assessed all patients at T1 and 2 others performed the assessments at T2. All assessors had received formal training in the administration of the outcome measures. Study-specific global rating of change scales (GRSs) were administered by 1 of 3 nurses at T2. The longitudinal study was approved by the Ethics Review Board of the Centre Intégré Universitaire de Santé et de Services Sociaux du Saguenay-Lac-St-Jean (Quebec, Canada) and procedures were conducted in accordance with the Declaration of Helsinki.

Clinical assessments

For descriptive purposes data on sex, age, phenotype and muscular impairments were collected. The latter were assessed with the DM1-specific muscular impairment rating scale (MIRS) which is a 5-graded scale ranging from no muscular impairment (grade 1) to severe proximal weakness (grade 5) (21). Mild muscular impairment was defined as MIRS grades 1–3 and severe as MIRS grades 4–5.

The Timed Up and Go (TUG) test was used for assessing mobility and balance (22). Following standardized instructions, the time taken to stand up from a chair, walk 3 m, turn, walk back and then sit down, was recorded. Patients were instructed to walk at their preferred self-selected speed. They performed 3 trials, and data from the second trial were used in the analyses in accordance with the original test administration (22). The measurement error, i.e. the standard error of measurement (SEM) and smallest detectable change (SDC) of TUG in DM1 are reportedly 0.7 and 1.9 s, respectively (23).

The Berg Balance Scale (BBS) was used for assessing balance (24). Following standardized instructions, the performance of 14 tasks of varying difficulty was assessed and scored on a 5-graded ordinal scale. A sum score, which can range from 0 to 56 points, was calculated, higher scores indicating better performance/balance. Measurement error of the BBS in the DM1 population is unknown as there are no published reliability studies.

Quantitative muscle testing (QMT) was performed to assess maximal isometric muscle strength using make tests with a calibrated handheld dynamometer (Microfet – 2, Hoggan Health Industries, Salt Lake City, UT, USA). Following a standardized protocol (18) maximal isometric contractions of 4 lower-limb muscle groups, i.e. hip flexors, knee extensors, knee flexors and ankle dorsiflexors of the right and left legs, were assessed. The length of the corresponding lever arm (in m) between the point of application of the dynamometer and the relevant joint centre was measured in order to calculate the maximal isometric torque in Newton-metres (Nm). Two trials were performed for each muscle group, unless they differed more than 10%, in which case a third trial was performed. The mean of the 2 trials (or the 2 nearest trials if 3 trials were made) was calculated for each muscle group and side. To reduce the number of explored variables, the mean of both sides for each muscle group was then calculated. In addition, a lower-limb composite score was created to take into account that each muscle group contributes to a certain extent to each related criterion used in the assessment of responsiveness. The lower-limb composite score was calculated by adding the scores from each muscle group and side, thus representing a lower-limb total muscle force (25). There are no test–retest reliability studies on maximal isometric lower-limb muscle strength using make tests with the Microfet – 2 in DM1. Thus, the measurement errors are unknown.

A Jamar™ dynamometer (Sammons Preston, Chicago, IL, USA) was used for assessing grip strength. Following standardized procedures (26) the maximal grip score, in kg, was recorded. Participants performed 3 trials with each hand. The mean score of all trials was calculated and used in analyses. For right and left hand together, the SEM and SDC for maximal grip strength in DM1, measured with the Grippit®, are reportedly 1.4 and 3.8 kg, respectively (27).

A B&L pinch gauge (B&L Engineering, Tustin, CA, USA) was used to assess pinch-grip strength. Following standardized procedures (26) the maximal pinch-grip score, in kg, was recorded. Participants performed 3 trials with each hand. The mean score of all trials was calculated and used in analyses. For right and left hands together, the SEM and SDC for maximal pinch-grip strength in DM1, measured with the Grippit®, are reportedly 0.7 kg and 1.9 kg, respectively (27).

The Purdue Pegboard Test (PPT) (model 32020, Lafayette Instrument Co., Lafayette, IN, USA) was used for assessing manual dexterity. Following standardized instructions (28), the number of pegs placed on the pegboard in 30 s was recorded for right and left hands, respectively. The participant performed 2 trials with each hand. The mean score of all trials was calculated and used in analyses. For right and left hands together, the SEM and SDC of PPT in DM1 are reportedly 0.7 and 1.9 pegs, respectively (27).

Study-specific GRSs were used to collect information on perceived changes in recent years in balance, walking, lower-limb weakness, stair-climbing and hand weakness. Patients rated their perceived change on 5-graded scales, ranging from a lot worse to a lot better.

Rationale for criterion approach

Patients' answers on the 5-graded GRSs were dichotomized into stable or worse and used as criteria. For each outcome 1–3 patients (1–3%) were excluded, as they reported a change for the better and it was not possible to create a distinct group for analysis. The GRSs on perceived change in balance, walking and lower-limb weakness were regarded by the research team as capturing similar constructs as measured with TUG, BBS and QMT lower-limb composite score. The stair-climbing GRS was considered to reflect a similar construct to that captured by the QMT lower-limb composite score and QMT knee extensor strength. The hand weakness GRS was regarded as reflecting a construct similar to that captured by the outcome measures for manual dexterity and grip strength. We made *a priori* hypotheses about AUCs for change scores (Table I) in the various outcome measures to distinguish patients who had become worse according to the criterion, i.e. their answers on the GRSs We also hypothesized that the AUC for change in BBS would be greater than for change in TUG concerning those worse in balance, and vice versa for walking. These hypotheses were based on published literature (25, 29–31) and clinical experience.

Rationale for construct approach

The research team considered TUG and BBS to measure similar constructs, i.e. mobility and balance, and that relative changes in these outcome measures should therefore be at least moderately correlated ($r \geq 0.6$). Although lower-limb muscle strength is important for mobility and balance, relative changes in the composite score were thought to correlate less than moderately ($r \leq 0.4$) with relative changes in TUG and BBS. It was further hypothesized that correlations between relative changes in TUG and BBS and relative changes in QMT measures would be greatest with changes in QMT knee extensor strength, followed by QMT foot dorsiflexor strength. Manual dexterity, as measured with PPT, and grip strength measures, were considered to capture slightly diverging constructs, and thus correlations between relative changes would be less than moderate ($r \leq 0.4$). Grip and pinch-grip strength measures were considered to capture a similar construct, and relative changes in these outcome measures would thereby be highly correlated ($r \geq 0.7$). Hypotheses were based on published literature (25, 29–31) and clinical experience.

Statistical analyses

Descriptive statistics were used to present the data. Median values were chosen, since not all data were normally distributed. Change scores in the various outcome measures were calculated by subtracting T1 scores from T2 scores, except for TUG, where T2 scores were subtracted from T1 scores. Thus, negative change scores indicated deterioration. Changes were expressed in absolute (s, points, Nm, pegs and kg) and relative (percentage) values.

Responsiveness was evaluated both with a criterion approach and a construct approach (5, 7). Descriptive statistics and ROC curve analyses with calculations of AUCs were used for the criterion approach. *A priori* hypotheses were made as specified above in the rationale for the criterion approach. Spearman's correlation analyses were used for the construct approach. The correlations between relative changes in outcome measures were hypothesized to be in the positive direction and the magnitudes to be as stipulated above in the rationale for the construct approach. All analyses were performed using SPSS software (version 22).

RESULTS

Descriptive results

Two of the 115 patients were excluded; one due to inability to perform the physical tests, and the other when data analyses showed that this patient was a severe outlier in all QMT measures (showing improvement) and, at the same time, reported changes for the worse in all GRSs, thereby having a major inappropriate effect on the results. Thus, 71 women and 42 men were included. Mean age at follow-up was 52 years (range 29–85 years). Eighty-nine (79%) patients had adult-onset phenotype and 24 (21%) late-onset. At baseline, 47 (42%) patients were classified as having mild muscular impairment, i.e. MIRS grades 1–3, and at the follow-up 9 years later 29 patients were so classified (26%). Statistical analyses (independent *t*-tests and χ^2 tests) of T1 baseline data between the 113 participants in the study and the 87 lost to follow-up/excluded showed no significant differences in phenotype, number of CTG repeats or sex. However, there were significant between-group differences in age and MIRS classification. Participants lost to follow-up were a mean of 8 years older at T1 and fewer were classified as MIRS 1–3.

Results of outcome measures at T1 and T2, and absolute and relative changes, are presented in Table II. Due to transport problems 7 patients were not evaluated

Table I. Criterion approach hypotheses about size of the area under curve (AUC) to distinguish patients who had changed to the worse according to the criteria, i.e. their answers on the global rating of change scales in balance, walking, lower-limb (LL) weakness, stair-climbing and hand weakness

	Balance AUC	Walking AUC	LL weakness AUC	Stair-climbing AUC	Hand weakness AUC
Timed Up and Go	≥0.7	≥0.6	≥0.6		
Berg Balance Scale	≥0.7	≥0.6	≥0.6		
QMT, LL composite score	≤0.6	≤0.6	≥0.7	0.5–0.7	
QMT, knee extensors				0.5–0.7	
Grip strength					≥0.7
Pinch-grip strength					≥0.7
Purdue Pegboard					≥0.6

QMT: quantitative muscle testing.

Table II. Results from outcome measures at baseline (T1) and the 9-year follow-up (T2)

Outcome measure	Median (IQR)	Min	Max
Timed Up and Go (*n* = 98)			
T1, s	9.5 (8.6–10.5)	7.1	18.8
T2, s	9.1 (7.4–11.5)	4.9	26.3
Change T1–T2, s [%]	0.4 [4.6] (–1.0–1.8 [–11.2 to 18.2])	–16.0 [–156.1]	4.3 [43.8]
Berg Balance Scale (*n* = 76)			
T1, points	56.0 (53.5–56.0)	21.0	56.0
T2, points	55.0 (46.0–56.0)	5.0	56.0
Change T2–T1, points [%]	–1.0 [–1.8] (–7.0–0.0 [–13.9 to 0.0])	–49.0 [–88.9]	2.0 [3.7]
QMT LL comp score[a] (*n* = 102)			
T1, Nm	390.9 (310.6–474.5)	156.7	723.2
T2, Nm	302.5 (237.7–393.4)	57.1	743.2
Change T2–T1, Nm [%]	–100.2 [–24.3] (–133.3 to –48.8 [–35.7 to –12.5])	–323.0 [–63.5]	101.5 [29.0]
QMT, hip flexors[b] (*n* = 105)			
T1, Nm	50.8 (39.8–62.2)	19.7	137.5
T2, Nm	52.5 (41.0–71.2)	14.0	121.6
Change T2–T1, Nm [%]	1.8 [3.4] (–5.5–12.3 [–12.6 to 23.0])	–47.0 [–53.5]	44.9 [143.9]
QMT, knee extensors[b] (*n* = 104)			
T1, Nm	79.5 (59.6–98.9)	18.9	165.3
T2, Nm	53.4 (38.9–72.1)	4.6	153.8
Change T2–T1, Nm [%]	–25.2 [–33.6] (–37.8 to –10.3 [–43.5 to 17.7])	–80.0 [–86.3]	16.0 [22.3]
QMT, knee flexors[b] (*n* = 106)			
T1, Nm	48.9 (40.0–63.1)	20.3	114.4
T2, Nm	31.9 (23.2–43.1)	4.5	88.4
Change T2–T1, Nm [%]	–16.5 [–33.9] (–25.9 to –8.1 [–50.2 to –19.0])	–55.4 [–81.5]	16.7 [23.4]
QMT, dorsiflexors[b] (*n* = 105)			
T1, Nm	15.1 (11.0–19.6)	2.2	38.7
T2, Nm	6.8 (4.5–10.0)	0.5	29.5
Change T2–T1, Nm [%]	–8.1 [–56.4] (–11.4 to –4.9 [–67.4 to –40.1])	–25.3 [–90.9]	2.6 [20.8]
Grip strength[b] (*n* = 105)			
T1, kg	11.2 (5.2–17.8)	0.8	51.5
T2, kg	8.9 (5.0–15.6)	0.8	51.5
Change T2–T1, kg [%]	–1.0 [–13.3] (–2.7–0.7 [–30.4 to 11.1])	–24.2 [–81.8]	6.8 [220.0]
Pinch grip strength[b] (*n* = 109)			
T1 kg	5.7 (4.5–7.2)	2.2	13.7
T2, kg	4.0 (3.0–5.0)	0.8	12.5
Change T2–T1, kg [%]	–1.4 [–27.8] (–2.0 to –1.0 [–37.1 to –20.6])	–5.2 (–79.1)	0 (–0.4)
Purdue Pegboard[b] (*n* = 110)			
T1, pegs	12.3 (11.0–13.5)	5.0	16.3
T2 (pegs)	11.6 (9.3–13.3)	2.5	17.0
Change T2–T1, pegs [%]	–0.8 [–5.9] (–2.0–0.3 [–18.6 to 1.9])	–7.3 [–72.2]	3.3 [23.6]

IQR: interquartile range; Min: minimum value; Max: maximum value; QMT: quantitative muscle testing; Nm: Newton meter; LL: lower-limb; [a]sum of values from right and left sides for the 4 tested muscle-groups, [b]mean value of right and left side.

at the neuromuscular clinic at T2, and thus lack TUG, BBS and QMT data. The BBS was added after the start of T1 and 30 patients were therefore not assessed with the scale. Other missing data in the various outcome measures are due to patients' incapacity to perform the tests due to severe muscle impairment. All but 2 outcome measures (TUG, QMT hip flexors), showed a median deterioration over time. A ceiling effect was found for BBS at both T1 and T2, where 59% and 41%, respectively, had the maximum score of 56.

Data were missing from 3–7 patients in the GRS (Fig. 1). Between 42% and 62% of the patients who answered the GRSs had experienced a change for the worse (Fig. 1). A majority of those who perceived themselves as stable had the late-onset phenotype.

Criterion approach results

The direction and magnitude of median absolute and relative changes in the various outcome measures
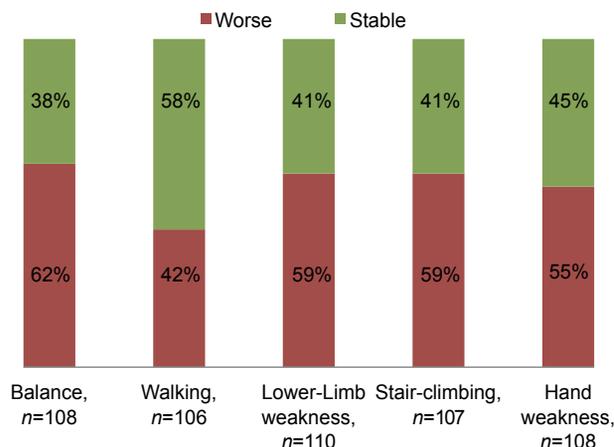
Fig. 1. Percentage of patients reporting being stable and worse on the global rating of change scales in balance, walking, lower-limb weakness, stair-climbing and hand weakness.

Table III. Absolute and relative median changes in outcome measures according to patients' answers on the global rating of change scales in balance, walking, and lower-limb (LL) weakness

	Timed Up and Go			Berg Balance Scale			QMT, LL composite score[a]		
	n	Median, s	Median, %	*n*	Median, p	Median, %	*n*	Median, Nm	Median, %
Balance									
A lot worse	23	−0.9	−8.9	23	−7.0	−16.1	28	−105.9	−29.1
A bit worse	32	0.0	−0.1	23	−1.0	−1.8	31	−100.8	−25.4
Stable	38	1.3	15.0	29	0.0	0.0	38	−100.9	−20.5
Better	3	1.9	19.5	1			3	−23.9	−8.2
Walking									
A lot worse	19	−1.5	−16.6	20	−6.5	−15.4	24	−103.9	32.6
A bit worse	11	−0.3	−3.4	9	−6.0	−10.7	12	−94.5	−25.7
Stable	61	1.0	12.2	45	0.0	0.0	59	−99.4	−21.2
Better	1			1			1		
Lower-Limb weakness									
A lot worse	24	−2.0	−20.1	21	−8.0	−16.1	28	−115.9	−32.3
A bit worse	27	0.3	1.5	22	−1.0	−1.8	29	−89.5	−24.5
Stable	44	1.0	10.7	33	0.0	0.0	42	−91.7	−20.8
Better	0			0			0		

QMT: quantitative muscle testing; s: seconds; p: points; Nm: Newton meter; SD: standard deviation; [a]sum of values from right and left sides for the 4 tested muscle-groups.

were, in general, in line with patients' self-reported changes in balance, walking, lower-limb weakness, stair-climbing and hand weakness (Tables III and IV, Figs 2 and 3). The median changes in TUG, grip strength, pinch-grip strength, and PPT for those a little/a lot worse in the various GRS did not, in general, exceed the measurement errors, i.e. reported SDCs. Most (72%) *a priori* criterion-approach hypotheses concerning AUCs were confirmed (Table V). However, the AUC for change in BBS was not greater than for change in TUG concerning balance, and there was no difference in AUCs between the outcome measures concerning walking. As shown by the AUCs, TUG, BBS and relative QMT measures could satisfactorily discriminate between stable and worse groups (Table V), although there was an overlap in change scores, meaning that the same change score could be attributed both to a patient who rated him/herself as worse and to a patient who rated him/herself as stable (Figs 2 and 3).

Construct approach results

Most (70%) *a priori* formulated construct approach hypotheses were confirmed (Table VI). Despite the predictions, the lowest correlations between QMT measures and TUG and BBS were found for foot dorsi-flexors. The correlation between re-

Fig. 3. Box plots showing relative (percentage) change in: (a) knee extensors strength and lower-limb composite score from quantitative muscle testing for stable and worse groups according to patients' answers on the global rating of change scale in stair-climbing; and (b) grip strength, pinch-grip strength and Purdue Pegboard test for stable and worse groups according to patients' answers on the global rating of change scale in hand weakness. Negative values indicate deterioration.

Fig. 2. Box plots showing relative (percentage) changes in Timed Up and Go, Berg Balance Scale and lower-limb composite score from quantitative muscle testing for stable and worse groups according to patients' answers on the global rating of change scales in balance, walking and lower-limb weakness. Negative values indicate deterioration.

Table IV. Absolute and relative median changes in outcome measures according to patients' answers on the global rating of change scales in stair-climbing and hand weakness

| | Stair-climbing | | | | | | Hand weakness | | | | | | | | |
| | QMT, knee extensors[a] | | | QMT, LL composite score[b] | | | Grip strength[a] | | | Pinch grip strength[a] | | | Purdue pegboard[a] | | |
	n	Median, Nm	Median, %	n	Median, Nm	Median, %	n	Median, kg	Median, %	n	Median, kg	Median, %	n	Median, pegs	Median, %
A lot worse	34	−29.2	−41.6	33	−100.8	−30.1	22	−1.1	−22.2	25	−1.7	−38.9	26	−1.0	−8.1
A bit worse	23	−29.8	−35.3	23	−104.8	−24.3	32	−1.0	−15.1	32	−1.4	−26.0	32	−1.0	−7.1
Stable	41	−17.6	−21.1	40	−63.1	−15.5	47	−1.0	−9.5	47	−1.5	−26.4	47	−0.5	−3.1
Better	1			1			0			0			0		

QMT: quantitative muscle testing; LL: lower-limb; Nm: Newton meter; SD: standard deviation.
[a]Mean value of right and left side, [b]sum of values from right and left sides for the 4 tested muscle-groups.

Table V. Criterion approach results from receiver operating characteristic curve analyses of change scores predicting patients classified as worse according to their answers on global rating of change scales in balance, walking, lower-limb (LL) weakness, stair-climbing and hand weakness

| | Balance | | Walking | | LL weakness | | Stair-climbing | |
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Timed Up and Go								
Change (s)	**0.8**	0.7–0.9	**0.8**	0.7–0.9	**0.7**	0.6–0.8		
Change (%)	**0.8**	0.7–0.9	**0.8**	0.7–0.9	**0.7**	0.6–0.8		
Berg Balance Scale								
Change (points)	**0.7**	0.6–0.9	**0.8**	0.7–0.9	**0.8**	0.6–0.9		
Change (%)	**0.7**	0.6–0.9	**0.8**	0.7–0.9	**0.8**	0.6–0.9		
QMT, LL composite score								
Change (Nm)	**0.6**	0.4–0.7	**0.5**	0.4–0.6	0.6	0.5–0.7	**0.6**	0.5–0.7
Change (%)	**0.6**	0.5–0.7	**0.6**	0.5–0.8	**0.7**	0.6–0.8	**0.7**	0.6–0.8
QMT, knee extensors								
Change (Nm)							**0.6**	0.5–0.7
Change (%)							**0.7**	0.6–0.8

| | Hand weakness | |
	AUC	95% CI
Grip strength		
Change (kg)	0.5	0.4–0.6
Change (%)	0.6	0.5–0.7
Pinch-grip strength		
Change (kg)	0.6	0.5–0.7
Change (%)	0.6	0.5–0.7
Purdue pegboard		
Change (pegs)	**0.6**	0.5–0.7
Change (%)	**0.6**	0.5–0.7

AUC: area under the curve; CI: confidence interval; QMT: quantitative muscle testing; comp: composite.
Hypothesis supported is shown in bold.

Table VI. Construct approach results presenting Spearman's correlation coefficients between relative (percentage) changes in outcome measures

	BBS	Hip flex	Knee ext	Knee flex	Dorsiflex	LL comp. score
Timed Up and Go	**0.7**	0.2	**0.5**	0.2	0.1	**0.4**
Berg balance scale (BBS)		0.2	**0.5**	0.3	0.1	**0.4**
QMT, hip flexors (hip flex)			0.3	0.4	0.2	0.6
QMT, knee extensors (knee ext)				0.5	0.4	0.8
QMT, knee flexors (knee flex)					0.3	0.8
QMT, dorsiflexors (dorsiflex)						0.5

	Pinch-grip strength			Purdue pegboard	
Grip strength	0.4			**0.1**	
Pinch-grip strength				**0.3**	

QMT: quantitative muscle testing; LL comp. score: lower-limb composite score.
Hypothesis supported is shown in bold.

lative changes in grip strength and pinch-grip strength was lower than anticipated.

DISCUSSION

This is the first study to assess, from a long-term perspective, the responsiveness of performance-based outcome measures in adults with DM1. Applying criterion and construct approaches, we explored predefined hypotheses about expected AUCs and correlations between changes in the various constructs. Our main findings are that promising responsiveness was found for TUG, BBS and QMT measures, i.e. the knee extensors and the lower-limb composite score, the latter when changes were expressed in relative terms. Grip strength and pinch-grip strength measures, and PPT showed poorer responsiveness.

The follow-up rate from baseline might be considered as rather low. There is, however, an increased mortality rate in this patient group and 30% of patients from baseline were deceased 9 years later. Thus, despite the loss to follow-up, we argue that the study is unique and important regarding both sample size and the follow-up period. Furthermore, that only 14 patients declined to participate must be regarded as a success.

Surprisingly, not all studied outcome measures showed median deterioration changes over the 9-year period. Even though DM1 is a slowly progressive disease, mobility, as measured with TUG, and QMT measures, is not likely to improve after such a long time. The way TUG is administered, i.e. participants must walk at a self-selected speed, adds variability, which probably influenced the result. Thus, it can be speculated that another method of administration, e.g. walking as fast as safely possible, might be better when using TUG in research or in the clinical setting. It is not known why there was no median decline in pure isometric hip flexor muscle strength in our study. All other QMT data showed large relative changes, in agreement with the previously reported force loss in muscle strength found over time in DM1 (29). The BBS descriptive data showed that this outcome measure had major ceiling effects at both time-points. This indicates an inability to detect changes over time for some patients, and BBS is therefore less suitable as an outcome measure in DM1. However, the scale might be useful in clinical practice to detect deterioration.

Criterion results discussion

The patients' self-rated changes on the GRS, which were used as criteria, were reflected accordingly by relative changes in the performance-based outcome measures. Most *a priori* hypotheses concerning AUCs

were confirmed, although some AUCs were below 0.7, which has been proposed as a threshold for adequate responsiveness (6, 7). Both TUG and BBS had AUCs indicating promising responsiveness in relation to patients' self-reported changes in balance, walking and lower-limb weakness. Thus, despite previously discussed shortcomings, these outcome measures appear to reflect the patients' perspective. The discriminative ability to distinguish those worse in stair-climbing was good for the 2 QMT measures, i.e. relative changes in lower-limb composite score and knee extensors. Thus, these QMT measures appear to give valuable information on functioning in DM1. The discriminative ability of grip strength and pinch-grip strength measures, and PPT to distinguish those who considered themselves as having worse hand weakness was, however, poorer.

The AUCs were, in general, larger for relative changes than for absolute outcome measures of muscle strength. Thus, it may be more relevant to report changes in relative scores (%) rather than in absolute (Nm, kg) when assessing muscle strength. However, when interpreting percentage changes one should bear in mind that a small change in a weak person will render a large relative change. In addition, one must consider where in the continuum of deterioration a person loses strength. That is, a change of 3% in muscles of normal strength does not have the same effect as a 3% loss in fairly weak muscles.

A responsive outcome measure should be able to distinguish clinically important changes from measurement error (6). It is therefore relevant to relate change scores to the SDC, which defines changes beyond measurement error (7). Our data suggest that TUG, grip strength and pinch-grip strength measures, and PPT, were unable to detect clinically important changes over time in our cohort since the changes did not exceed reported SDCs (23, 27). However, note that the sample sizes were small (10 and 16 persons) in the reliability studies from which these SDCs were derived and that another device (Grippit®) was used for grip strength measures. As for the other outcome measures, no data are available on measurement errors in the DM1 population for comparison.

Construct results discussion

Most of our *a priori* construct approach hypotheses were confirmed. However, we did not exceed the cut-off point of >75% of the results for correspondence with the hypotheses, which is proposed to indicate positive responsiveness (6, 32). There is, however, no consensus or guidelines about how many hypotheses are to be tested or if all hypotheses are regarded as equally important. Hence, this cut-off point must be

considered with caution. Furthermore, some of our hypotheses were based on the literature and clinical experience of relationships between outcome measures and not on the relationship between change scores. Correlation coefficients in responsiveness studies are often lower than those found in construct validity studies (32). That the relative change in QMT of knee extensors correlated highly with changes in mobility and balance outcomes and the QMT lower-limb composite score, highlights the importance of this muscle group in DM1. Thus, knee extensor muscle strength may be significant for activity and participation in DM1. Furthermore, it seems clinically relevant to measure strength in these muscles to identify patients who are at risk of loss of functioning.

Study limitations

The use of GRSs, capturing patients' self-rated perceived changes, as gold standards or anchors in responsiveness studies is not uncommon (33–36). The proper function of these scales, however, places a cognitive demand on the participant, as he/she must recall and quantify status at previous time-points (37). Our choice of GRS as criteria might therefore be seen as a limitation as there is a cognitive decline over time in DM1 (19, 38). Even if it is important to address and respect the individual's own opinion, we acknowledge that clinician-rated GRS would have added valuable information. Unfortunately, we lacked such data. Furthermore, it could arguably be difficult for people with slowly progressive diseases, such as DM1, to perceive the gradual decline in physical functioning. Living with DM1 is a process of constant change and adjustment. The response-shift phenomenon, which includes a change in internal standards, values and conceptualization (39, 40), is part of that process. Thus, the response-shift phenomenon might have influenced patients' ratings of perceived changes in recent years.

Another possible limitation is that the responsiveness of grip strength measures might have been influenced by the measurement instrument used. The standard Jamar dynamometer and the B&L pinch gauge are not sensitive enough to register subtle changes, the smallest possible change being 0.5 kg. Thus, digital measurement instruments might be more responsive and possibly better for assessing grip and pinch-grip strength over time, both in research and in clinical practice.

Conclusion

We found that the performance-based outcome measures captured changes over the 9-year period. Promising responsiveness was found in the studied outcome measures of mobility (TUG), balance (BBS) and some QMT muscle strength measures when changes were expressed in relative terms. Poorer responsiveness was found in measures of hand weakness (grip and pinch-grip strength) and manual dexterity (PPT). However, further knowledge is needed on measurement errors of measures in the DM1 population to fully interpret the meaningfulness of the longitudinal changes. Finally, it is necessary also to assess responsiveness over a shorter time.

REFERENCES

1. De Antonio M, Dogan C, Hamroun D, Mati M, Zerrouki S, Eymard B, et al. Unravelling the myotonic dystrophy type 1 clinical spectrum: a systematic registry-based study with implications for disease classification. Rev Neurol (Paris) 2016; 172: 572–580.
2. Turner C, Hilton-Jones D. Myotonic dystrophy: diagnosis, management and new therapies. Curr Opin Neurol 2014; 27: 599–606.
3. Blonsky K, Monckton D, Wieringa B, Schoser B, Day JW, Engelen B. 2010 Marigold therapeutic strategies for myotonic dystrophy. Neuromuscul Disord 2012; 22: 87–94.
4. Symonds T, Campbell P, Randall JA. A review of muscle, and performance-based assessment instruments in DM1. Muscle Nerve 2017; 56: 75–85.
5. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737–745.
6. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34–42.
7. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
8. Laberge L, Prevost C, Perron M, Mathieu J, Auclair J, Gaudreault M, et al. Clinical and genetic knowledge and attitudes of patients with myotonic dystrophy type 1. Public Health Genomics 2010; 13: 424–430.
9. Laberge L, Veillette S, Mathieu J, Auclair J, Perron M. The correlation of CTG repeat length with material and social deprivation in myotonic dystrophy. Clin Genet 2007; 71: 59–66.
10. Gagnon C, Mathieu J, Noreau L. Life habits in myotonic dystrophy type 1. J Rehabil Med 2007; 39: 560–566.
11. Gagnon C, Mathieu J, Jean S, Laberge L, Perron M, Veillette S, et al. Predictors of disrupted social participation in myotonic dystrophy type 1. Arch Phys Med Rehabil 2008; 89: 1246–1255.
12. Laberge L, Dauvilliers Y, Begin P, Richer L, Jean S, Mathieu J. Fatigue and daytime sleepiness in patients with myotonic dystrophy type 1: to lump or split? Neuromuscul Disord

2009; 19: 397–402.

13. Gagnon C, Chouinard MC, Laberge L, Brisson D, Gaudet D, Lavoie M, et al. Prevalence of lifestyle risk factors in myotonic dystrophy type 1. Can J Neurol Sci 2013; 40: 42–47.

14. Laberge L, Mathieu J, Auclair J, Gagnon E, Noreau L, Gagnon C. Clinical, psychosocial, and central correlates of quality of life in myotonic dystrophy type 1 patients. Eur Neurol 2013; 70: 308–315.

15. 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 J Rare Dis 2014; 9: 186.

16. Bertrand JA, Jean S, Laberge L, Gagnon C, Mathieu J, Gagnon JF, et al. Psychological characteristics of patients with myotonic dystrophy type 1. Acta Neurol Scand 2015; 132: 49–58.

17. Raymond K, Auger LP, Cormier MF, Vachon C, St-Onge S, Mathieu J, et al. Assessing upper extremity capacity as a potential indicator of needs related to household activities for rehabilitation services in people with myotonic dystrophy type 1. Neuromuscul Disord 2015; 25: 522–529.

18. Petitclerc E, Hebert LJ, Mathieu J, Desrosiers J, Gagnon C. Lower limb muscle strength impairment in late-onset and adult DM1 phenotypes. Muscle Nerve 2017; 56: 57–63.

19. Gallais B, Gagnon C, Mathieu J, Richer L. Cognitive decline over time in adults with myotonic dystrophy type 1: a 9-year longitudinal study. Neuromuscul Disord 2017; 27: 61–72.

20. 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. Neuromuscul Disord 2017; 27: 673–682.

21. Mathieu J, Boivin H, Meunier D, Gaudreault M, Bégin P. Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. Neurology 2001; 56: 336–340.

22. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991; 39: 142–148.

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

24. Berg K, Wood-Dauphinée S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiother Can 1989; 41: 304–311.

25. Hammaren E, Kjellby-Wendt G, Kowalski J, Lindberg C. Factors of importance for dynamic balance impairment and frequency of falls in individuals with myotonic dystrophy type 1 – a cross-sectional study – including reference values of Timed Up & Go, 10m walk and step test. Neuromuscul Disord 2014; 24: 207–215.

26. American Society of Hand Therapists. Clinical assessment recommendations 2nd edn. Chicago: American Society of Hand Therapists; 1992.

27. Aldehag AS, Jonsson H, Littorin S, Ansved T. Reliability of hand function testing instruments in patients with muscular dystrophies. Int J Ther Rehabil 2008; 15: 211–218.

28. Tiffin J, Asher EJ. The Purdue pegboard; norms and studies of reliability and validity. J Appl Psychol 1948; 32: 234–247.

29. Hammaren E, Kjellby-Wendt G, Lindberg C. Muscle force, balance and falls in muscular impaired individuals with myotonic dystrophy type 1: a five-year prospective cohort study. Neuromuscul Disord 2015; 25: 141–148.

30. Wiles CM, Busse ME, Sampson CM, Rogers MT, Fenton-May J, van Deursen R. Falls and stumbles in myotonic dystrophy. J Neurol Neurosurg Psychiatry 2006; 77: 393–396.

31. Lindeman E, Leffers P, Reulen J, Spaans F, Drukker J. Quadriceps strength and timed motor performances in myotonic dystrophy, Charcot-Marie-Tooth disease, and healthy subjects. Clin Rehabil 1998; 12: 127–135.

32. Mahler E, Cuperus N, Bijlsma J, Vliet Vlieland T, van den Hoogen F, den Broeder AA, et al. Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis. Scand J Rheumatol 2016; 45: 518–527.

33. Gorman SL, Harro CC, Platko C, Greenwald C. Examining the function in sitting test for validity, responsiveness, and minimal clinically important difference in inpatient rehabilitation. Arch Phys Med Rehabil 2014; 95: 2304–2311.

34. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther 2013; 93: 158–167.

35. Beauchamp MK, Harrison SL, Goldstein RS, Brooks D. Interpretability of change scores in measures of balance in people with COPD. Chest 2016; 149: 696–703.

36. Ailliet L, Rubinstein SM, de Vet HC, van Tulder MW, Terwee CB. Reliability, responsiveness and interpretability of the neck disability index-Dutch version in primary care. Eur Spine J 2015; 24: 88–93.

37. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009; 17: 163–170.

38. Winblad S, Samuelsson L, Lindberg C, Meola G. Cognition in myotonic dystrophy type 1: a 5-year follow-up study. Eur J Neurol 2016; 23: 1471–1476.

39. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999; 48: 1507–1515.

40. Schwartz CE. Applications of response shift theory and methods to participation measurement: a brief history of a young field. Arch Phys Med Rehabil 2010; 91: S38–S43.