Advances in Physiotherapy 2002; 4:37–47

Taylor & Francis healthsciences

Reliability of the EK Scale, a Functional Test for Non-ambulatory Persons with Duchenne Dystrophy

BIRGIT F. STEFFENSEN, MSc, PT[1,2], SYLVIA A. HYDE, MBA, SRPT, FCSP[2], JØRN ATTERMANN, PhD[3] and EVA MATTSSON, PhD, PT[2]
[1]Institute for Neuromuscular Diseases, Århus, Denmark; [2]Department of Physical Therapy, Karolinska Institutet, Stockholm, Sweden; [3]Department of Biostatistics, University of Aarhus, Aarhus, Denmark

Abstract

The EK {Egen Klassifikation} scale was developed to assess overall functional ability in the non-ambulatory stage of Duchenne muscular dystrophy (DMD). The purpose of this study was to examine the reliability of the EK scale. Six subjects with DMD, selected as representative of the entire range of functional ability seen in the non-ambulatory phase of the disease, were video recorded and assessed using the EK scale. The assessment required both interview and performance of functional activities. The video records were shown to 17 healthcare professionals comprising seven physiotherapists, three physiotherapy students, four occupational therapists, two social workers and one physician. They viewed and assessed the video recordings. Seven of the professionals repeated the assessments after 6–8 weeks. Intra-class correlation coefficients determined for both inter- and intra-rater reliability were 0.98. The standard deviation of individual estimates of EK sum was 0.95 when an assessment was repeated by different evaluators on the same subject and 0.78 when repeated by the same evaluator. Weighted kappa values for individual categories ranged from 0.67 to 0.94. The EK scale was found to be highly reliable when used by healthcare professionals assessing the subjects from videotapes.

KEY WORDS: Assessment – functional ability – healthcare professionals – muscular dystrophy – reliability – videotapes.

INTRODUCTION

Duchenne muscular dystrophy (DMD) is a genetically determined neuromuscular disease affecting roughly one in 3500 newborn boys (1). The disease is characterized by a progressive loss of muscle strength and function throughout life leading to total dependency on technical devices and personal help in the advanced stage of the disease.

Traditionally, the course of the disease is described as having three stages: the stage of independent ambulation, the intermediate or transition stage where ambulation is gradually lost, and the advanced stage of wheelchair dependency. Increased life expectancy, largely resulting from more aggressive management of compromised ventilation (2), means that the person with DMD will spend a greater part of his life in the advanced stage of the disease.

A central part of quality of life is concerned with the domain of functional ability, which can be defined as the individual's ability to interact with his environment in a way that permits him to achieve tasks of daily living. A multi-disciplinary approach is needed to preserve functional ability and prevent secondary impairments (3,4). There is, however, a lack of instruments to specifically assess the functional abilities, which are present in the advanced stages of DMD.

The existing measures of function specifically used for subjects with DMD: the Vignos lower extremity classification (5), the Brooke upper extremity functional grade (6,7), and different kinds of timed tasks (6–8) are not useful in describing the overall function of the young person with DMD in the advanced stage of the disease. Furthermore, function in the sense of activities of daily living is not measured by these tests. Similarly, the existing generic measures of function, developed to assess daily activities in a wide range of disorders, do not focus on the specific

loss of function which is characteristic of DMD in the later stage of the disease. We therefore developed the EK {Egen Klassifikation} scale to satisfy this requirement (9) (Appendix 1). The title of this scale of functional ability is a direct translation from the Danish language and denotes the fact that the scale, in part, reflects the subject's self-perception of his functional ability, which is then further elucidated by the evaluator's assessment. This scale has been shown to be valid (10). The EK scale addresses the spectrum of functional ability and impairments characteristic for a young man with DMD from the time of loss of independent ambulation without knee-ankle-foot orthoses (KAFO) to total loss of function. Since this may require assessments over a period of more than 20 years it is essential that the instrument is reliable when used by different evaluators with different levels of experience and from different healthcare disciplines.

To explore the extent to which the EK scale is reliable in clinical practice it is important to determine the degree of confidence that can be ascribed to the units of the scale (11). It is also necessary to know the amount of training needed for administering the scale because DMD is a relatively rare disease and in normal clinical practice many clinicians might only see a few cases.

In the development of any new scale it is important to test reliability, and some sources of error that could render a test unreliable are (12):

- errors related to the instrument, such as unclear descriptions of the categories or the instructions for administration of the scale so that the categories are incomprehensible to the evaluator;
- errors related to the subject being evaluated, such as unwillingness or inability to explain and demonstrate how the tasks are performed;
- errors related to the evaluator, i.e. whether the inexperienced evaluator understands the assessment instrument correctly and the response of the subject evaluated, whether the experienced evaluator is open-minded to take in information that does not fit in with his experience, or whether the evaluator influences responses thus introducing bias.

A major concern in evaluating intra-rater reliability is the time interval between measurements. The time interval between successive tests should not be so long that the condition of the subject who is tested has changed, but long enough for the evaluator to forget the results of the first test. Videotaping has been used in several studies as a tool to make sure that the condition of the assessed subject is the same for the evaluator from test to test. It has also been used to eliminate subject variability in studies of inter-rater reliability (13–19).

The purpose of this study was to examine the intra-rater reliability of the EK scale when used by physiotherapists with different levels of experience and the inter-rater reliability when used by physiotherapists with different levels of experience and by other healthcare professionals.

MATERIALS AND METHODS

The EK scale

The EK scale (10, Appendix 1) is an ordinal scale ranging from 0 to 30 points where 0 represents the highest level of independent function and 30 the lowest level. The assessment comprises 10 parts referred to as categories, each concerning a major domain of functional ability and each category is scored in four items: 0–3. The EK scale permits a variety of categories and items to be assessed at the same time by summing of individual scores to attain an overall expression of function. The sum of scores of the individual categories is called the EK sum. The administration of the EK scale consists of an interview of the subject to learn how he performs the task in daily life (categories 1–9) or how he perceives his physical wellbeing (category 10) and a visual examination of the performance of those tasks which are observable (Appendix 2).

Evaluators

Seventeen evaluators participated in the study. They consisted of three physiotherapists experienced in treating subjects with DMD, three physiotherapy students with no experience in triton boys with DMD, four consultant physiotherapists and seven other healthcare professionals from different backgrounds.

The inclusion criteria for the experienced physiotherapists were that they were qualified, state registered and had personally been treating boys with DMD for at least 10 years. Three physiotherapists from a private clinic specializing in physiotherapy for boys with DMD, including those in the advanced stages of the disease, agreed to participate. They had been qualified for 15–22 years and had been treating boys with DMD for 10–18 years.

The inclusion criteria for the physiotherapy students were that they had not treated boys with DMD during their training. The students were recruited by placing a notice on the bulletin board inviting students entering the final term of the third year of training at the Physiotherapy School of Århus to participate. Three students who met the criteria accepted the invitation.

All professional staff of the Institute for Neuromuscular Diseases in Denmark who were involved with the care of individuals with DMD as advisors were also invited to participate and accepted. The professional staff comprised physiotherapists ($n = 4$), occupational therapists ($n = 4$), social workers ($n = 2$) and one physician. They had been acting as consultants to subjects with DMD, their families and their local healthcare professionals for periods of time ranging from 1 to 15 years. One of the physiotherapists in this group was one of the authors (BFS).

Each evaluator had been informed about the purpose of the study and was fully advised about the procedures to be used in the study.

For analysis of systematic differences in variance of scoring, the 17 evaluators were divided into seven sub-groups according to their experience and profession. The sub-groups were the experienced physiotherapists ($n = 3$), the students ($n = 3$), the physiotherapy consultants except BFS ($n = 3$), the occupational therapy consultants ($n = 4$), the social workers ($n = 2$), the physician ($n = 1$). BFS ($n = 1$) was assigned her own sub-group status to reflect her position as the only evaluator who actually met with the subjects for the purpose of administering the EK scale.

Subjects evaluated

Six male subjects with DMD (9, 12, 13, 17, 23 and 25 years of age), registered with the Institute for Neuromuscular Diseases in Denmark, participated in the study. Criteria for inclusion in the study were that the subject was:

- diagnosed in a neurological unit and met the established diagnostic criteria for DMD with respect to family history compatible with X-linked recessive inheritance, clinical course, muscle biopsy and serum creatinine kinase (SCK) activity (20);
- wheelchair dependent and had lost independent ambulation without KAFOs.

The individual subjects were selected so that the group as a whole represented different functional stages from full independence to full dependence in the wheelchair.

The 9-year-old boy could still stand and walk a few steps indoors wearing his KAFOs, but had lost independent ambulation without orthoses. The 25-year-old man had been nocturnally ventilated via nasal mask since he was 20 years old and from the age of 24 he had been ventilated via tracheostomy 24 h a day. No subjects with untreated symptoms of impending respiratory insufficiency were found.

The medical scientific ethical committee of the county of Århus had approved the study and each person with DMD or in the case of those less than 18 years of age, their parents, had read and signed a document of informed consent for participation in the study.

Equipment and procedure

Administration of the EK scale required an interview, observation of movements and performance of tasks. All the assessments were undertaken in the subjects' homes, where one of the authors (BFS) interviewed and examined all the subjects (Appendix 2). Each category of the EK scale was addressed separately. The questions of the examiner, the answers of the subjects and the movements performed were all recorded using videotape.

Two VHS video recorders, each supported on a tripod, were used to obtain simultaneous recording of anterior and lateral views of the subjects. In order to achieve an image of the whole body, the cameras were placed at a distance of 2 m from the sternum and the point of the shoulder for the anterio-

posterior and lateral views respectively. The vertical height of the camera was adjusted so that the lens was focused at the subjects' shoulder level. To facilitate both recording and observation of the subjects' movements, they were dressed in T-shirts and shorts and were all sitting in their wheelchairs when interviewed and demonstrating movements in the sitting position. The tasks turning in bed and standing were demonstrated by those subjects who could perform them.

The subjects randomly chose an identification letter (A–F) which was used to label the films. The anterior and lateral films were later compiled as one film for each subject, showing first the anterior view of each category and then the lateral view. The compiling of the tapes was done by an external unit. Each new category of the assessment scale was introduced in advance by filmed text so that the evaluator was prepared to observe the task that followed.

Rating procedure

All the evaluators viewed the videotaped recordings of the subjects in alphabetic order at the Institute for Neuromuscular Diseases. One of the authors (BFS) was present throughout the evaluation procedures and no discussion of the results was allowed.

The EK scale and the method of scoring were explained to the evaluators who were then provided with a copy of the EK scale, which included written instructions on the method of scoring, and a scoring sheet. No training in assessing the videotaped functions was given before the assessments took place. Viewing and scoring of all six films took approximately 2.5 h; the scoring sheets were collected immediately after the assessment session.

The experienced physiotherapists ($n = 3$), the physiotherapy consultant BFS ($n = 1$) and the students ($n = 3$) assessed the video records of all the subjects in alphabetic order a second time after 6–8 weeks. The other healthcare professionals ($n = 10$) assessed the subjects only once.

Statistics

Inter-rater reliability of the individual categories. Kappa statistics was used for estimating agreement between evaluators in scoring the individual categories. Since disagreements between evaluators of only one item was considered less serious than disagreements of two or three items, the agreement between evaluators was assessed by weighted kappa statistics (21), where full agreement between a pair of evaluators was given a weight of 1, disagreement of one item a weight of 0.75, of two items 0.5, and of three items 0.25.

Weighted kappa was calculated for each pair of evaluators (a total of 136 pairs from 17 evaluators), and an overall kappa was calculated by averaging over pairs. Where there was more than one rating by an evaluator on the same subject, only the first rating was used for the calculations.

Inter- and intra-rater reliability of EK sum. A total of 144 ratings from 17 evaluators on six subjects were used in the statistical analyses. Graphical examination by probability plots showed that data were consistent with a normal distribution allowing EK sum to be analysed as a mixed effects model. The analysis was performed by means of the MIXED procedure in the statistical software package SAS (22). Significant components of variance were identified and estimated, and reliability coefficients {inter- and intra-class correlation, ICC (2,1)} (23) and 95% confidence intervals of individual estimates of EK sum were estimated from these (see Appendix 3).

RESULTS

The EK sum scores of the six subjects were spread over the scale ranging from 3 to 24 (see Table I).

Reliability of individual categories

Agreement between evaluators, estimated by means of overall weighted kappa statistics, is shown in Table II. The range is from 0.67 to 0.94 except for category 9 (speech) with estimated kappa equal to 0.24.

Agreement within evaluators, as rated by experienced physiotherapists ($n = 3$) and physiotherapy students ($n = 3$), twice on each subject is shown in Table III. Out of 360 pairs of ratings no difference was found in 316, and 357 were within ± 1.

Inter- and intra-rater reliability of EK sum

There was no significant effect of group of evaluators on the mean of EK sum (F test, $p = 0.21$), nor on the

TABLE I: EK summary scores for each of the subjects (evaluators' first ratings only)

Evaluators		Subjects					
Group	Evaluator number	A	B	C	D	E	F
Experienced physiotherapists	1	20	15	12	4	12	24
	2	20	14	12	4	12	22
	3	19	16	11	3	11	20
Physiotherapy students	1	20	16	13	3	10	21
	2	21	13	13	4	11	22
	3	19	14	14	4	11	21
Physiotherapy consultants	1	20	16	12	4	12	22
	2	18	15	13	4	15	21
	3	20	14	11	4	12	21
Occupational therapy consultants	1	21	14	12	5	11	20
	2	19	16	14	4	14	22
	3	20	16	13	4	12	21
	4	19	18	12	5	13	22
Social workers	1	19	15	13	4	12	22
	2	19	16	14	4	13	22
Physician	1	21	16	12	4	12	20
BFS	1	19	16	12	4	11	20

magnitude of variance (chi-squared test, $p = 0.53$). Hence, there were no statistically significant differences in reliability between different groups of evaluators and the EK sum could be adequately described by the model given in the Materials and Methods section.

The estimates of the overall mean and the variances and corresponding p-values are shown in Table IV.

The estimates of the inter- and intra-rater reliability coefficients were 0.977 and 0.984, respectively, with lower 95% confidence limits 0.954 and 0.930, respectively. For a fixed subject, the standard deviation in EK sum rated by different evaluators was 0.95 and the 95% credibility interval of the true EK sum was $X \pm 1.86$, where X is the rating performed by a randomly chosen evaluator. For a fixed subject within a fixed evaluator, the standard deviation in EK sum was 0.78 and the 95% credibility interval of the true EK sum was $X \pm 1.52$, where X is the rating performed by the fixed evaluator.

DISCUSSION

The results from this study show an ICC (2,1) of 0.98 of the EK scale when estimated from videotaped subjects with DMD, who were representing most of the scale. This reliability must be considered as high. Some of the explanation of the high ICC could be the wide spread of measurements of EK sum, ranging from 3 to 24 (Table I), which prevents range effects from producing spuriously low reliability coefficients, since the reliability coefficient is the ratio of variability between subjects to the total variability. There was no significant effect of different healthcare

TABLE II: Inter-rater agreement of the assessment of individual categories by 17 evaluators as estimated by overall weighted kappa statistics

Categories of EK	Overall weighted kappa
1. Ability to use wheelchair	0.69
2. Ability to transfer from wheelchair	0.76
3. Ability to stand	0.94
4. Ability to balance in the wheelchair	0.72
5. Ability to move the arms	0.78
6. Ability to use hands and arms when eating	0.77
7. Ability to turn in bed	0.86
8. Ability to cough	0.80
9. Ability to speak	0.24
10. Physical wellbeing	0.67

TABLE III: Intra-rater agreement of the assessment of individual categories illustrated as differences between first and second assessment of the categories of EK

Difference	−3	−2	−1	0	1	2	3
Categories of EK							
1. Ability to use wheelchair			6	30			
2. Ability to transfer from wheelchair			3	33			
3. Ability to stand				36			
4. Ability to balance in the wheelchair			1	29	6		
5. Ability to move arms				34	2		
6. Ability to use hands and arms when eating			2	29	5		
7. Ability to turn in bed		1	5	30			
8. Ability to cough			5	30	1		
9. Ability to speak				34	2		
10. Physical wellbeing				31	3	1	1
Sum of differences		1	22	316	19	1	1

The evaluators were experienced physiotherapists ($n = 3$) and physiotherapy students ($n = 3$) who scored six subjects twice. "0" indicates no difference between first and second assessment. Negative numbers indicate that the second score was higher than the first. Positive numbers indicate that the first score was higher than the second.

professionals or physiotherapists with or without experience on the mean of EK sum or on precision, i.e. the closeness of agreement between independent test results. However, the lack of differences between the different groups of evaluators might be explained by the small number of evaluators in each group (1–4).

The EK scale was designed to assist in clinical decision making and to be used in a multi-disciplinary team. For this purpose the 95% credibility interval of an individual score of EK sum is a more useful expression of reliability than the ICC (11). Furthermore, it is expressed in the units of the scale whereas the ICC is a unitless quantity. In this case it means that the true EK sum can vary ± 1.86 EK sum units when a subject is assessed by different evaluators and ± 1.52 EK sum units when the subject is assessed by the same evaluator.

In the individual categories of EK, the kappa values varied between 0.67 and 0.94 for nine of the 10 categories (Table II). The category 9, "Ability to speak", had a very low kappa value (0.24). Close scrutiny of the raw data of the category 9 revealed high prevalence of scores in the item 0 for all six subjects in this category. This is expected to reflect a true agreement between the evaluators, since all six subjects in this sample had well-preserved voice and speech. Even the ventilator-dependent subject had normal speech as his reduced respiratory capacity was compensated for by the ventilator, but it is also possible that this reflects a weakness in kappa statistics, a paradox discussed by several authors (24). When the prevalence of a score approaches 100% or 0% the kappa value falls.

The categories 3 and 7, "Ability to stand" and "Ability to turn in bed", were those with the highest kappa values (Table II). These categories were probably more easily evaluated because they are clear and precisely defined and more easily understood than broad and complex activities (25–27).

The categories 1 and 10, "Ability to use wheelchair" and "Physical wellbeing", showed the largest disagreement (Table II). The scoring of item 3 of category 1 and all items of category 10 were mainly based on interview. The scoring of these

TABLE IV: Estimates of the overall mean and the variances and corresponding p-values

Parameter	Estimate	Standard error	DF	p
μ	14.1115	2.5128	5	0.0025
τ^2	37.8401		1	<0.0001
ν^2	0.2959		1	0.043
σ^2	0.6010			

μ = overall mean of EK sum, τ^2 = variance of a_i, ν^2 = variance of b_{ij}, σ^2 = variance of e_{ijk}, see Appendix 3.

categories was therefore more dependent upon interpretation of the responses of the subjects by the evaluators. The interview on the category "Physical wellbeing" is dealing with symptoms of hypoventilation. Some of the boys complained of problems with sleeping, and it was difficult to differentiate whether this was basically a respiratory problem or a musculo-skeletal problem. However, this would also be difficult to determine in real life when based on an interview only. Furthermore, the items of category 10 are not mutually exclusive and therefore more difficult to score.

Video recording was used in this study to standardize the test situation and to eliminate the error due to changes in the subjects over time. By using video recording as the medium we could include subjects whom it would have been difficult and costly to see for repeated measurements, as they were severely restricted in their mobility. However, the natural diurnal variation of the subjects with DMD could not be determined by this method. This was not considered to detract from the results in this study because the items of the scale are so broad that they are not expected to change during a day. The potential variation in the physical and verbal responses of the subjects to different examiners with different personality or professional background could not be determined either. An element of bias may have been introduced to the study by using only one physiotherapist, who was experienced, for examining all subjects. Theoretically this might have improved reliability and contributed to the lack of difference between the professionals. However, since most of the categories are simple everyday tasks with which the subjects are familiar and they are only asked to demonstrate and explain what they do in daily life, the information obtained for scoring cannot be expected to vary much in real life. Furthermore, the subjects with DMD at this age (>9 years) are usually co-operative and accompanied by attendants who would be able to give supplementary information if necessary.

Video recording has been used in the construction of the reliability study, which is in common with other studies in which assessment instruments for evaluating movement performance (13–19) and interviews (28,29) have been tested for reliability. However, in some of the reported studies, training of the evaluators in observing the patients on the video was a major task (14). This was not necessary in our study due to the simplicity of the scale. The majority of the items of the scale are mutually exclusive, which makes it easy to score.

The relatively small number of subjects with DMD, the profound impact of the disease on function and the involvement of specialists from many different professional disciplines make it important to have a tool for evaluating function that is both easy to administer, does not require special training, and is reliable.

CONCLUSION

In this study the EK scale was shown to be highly reliable when tested for intra- and inter-rater reliability among subjects with DMD when used by physiotherapists with different levels of experience, and by other healthcare professionals.

ACKNOWLEDGEMENTS

We gratefully acknowledge the co-operation of the young men who were videotaped, the physiotherapy students and the colleagues who evaluated the videotaped boys. We also wish to thank Stina Wallman, MSc, PT, for her help with filming the boys. This study was supported by Muskelsvindfonden, Danske Fysioterapeuters Forskningsfond and Lily Benthine Lunds Fond af 01.06.78.

APPENDIX 1: THE EK SCALE

Each of 10 categories consists of four items (0–3) and the EK sum is the sum of scores over categories.

1. Ability to use wheelchair

0 Able to use a manual wheelchair on flat ground, 10 m on less than 1 min.
1 Able to use a manual wheelchair on flat ground, 10 m in more than 1 min.
2 Unable to use manual wheelchair, requires electric wheelchair.
3 Uses electric wheelchair, but occasionally has difficulty in steering.

2. Ability to transfer from wheelchair

0 Able to transfer from wheelchair without help
1 Able to transfer independently from wheelchair with use of aid.
2 Needs assistance to transfer with or without additional aids (lift, easy glide).
3 Needs to be lifted with support of head when transferring from wheelchair.

3. Ability to stand

0 Able to stand with knees supported, as when using braces.
1 Able to stand with knees and hips supported, as when using standing aids.
2 Able to stand with full body support.
3 Unable to be stood, marked contractures.

4. Ability to balance in the wheelchair

0 Able to push himself upright from complete forward flexion by pushing up with hands.
1 Able to move the upper part of the body more than 30° from the upright position in all directions, but cannot push himself upright from the total forward flexed position.
2 Able to move the upper part of the body less than 30° from one side to the other.
3 Unable to change position of the upper part of the body, cannot sit without total support of trunk and head.

5. Ability to move the arms

0 Able to raise the arms above the head with or without compensatory movements.
1 Unable to lift the arms above the head, but able to raise the forearms against gravity, i.e. hand to mouth with or without elbow support.
2 Unable to lift the forearms against gravity, but able to use the hands against gravity when the forearm is supported.
3 Unable to move the hands against gravity but able to use the fingers.

6. Ability to use the hands and arms when eating

0 Able to cut meat into pieces and eat with spoon and fork. Can lift a filled cup (approx. 250 ml) to the mouth without support at elbow.
1 Eats and drinks with support at elbow.
2 Eats and drinks with elbow support and with reinforcement of the opposite hand ± feeding aids.
3 Has to be fed.

7. Ability to turn in bed

0 Able to turn himself in bed with bedclothes.
1 Able to turn himself on a couch, but not in bed.
2 Unable to turn himself in bed. Has to be turned three times or less during the night.
3 Unable to turn himself in bed. Has to be turned four times or more during the night.

8. Ability to cough

0 Able to cough effectively.
1 Has difficulty to cough and sometimes needs manual reinforcement. Able to clear the throat.
2 Always needs help for coughing. Only possible to cough in certain positions.
3 Unable to cough. Needs suction or positive pressure breathing techniques in order to keep the airways clear.

9. Ability to speak

0 Powerful speech. Able to sing and speak loudly.
1 Speaks normally, but cannot raise his voice.
2 Speaks with quiet voice and needs a breath after 3–5 words.
3 Speech is difficult to understand except to close relatives.

10. Physical wellbeing

0 No complaints, feels good.
1 Easily tires. Has difficulty resting in a chair or in bed.
2 Has loss of weight, loss of appetite. Scared of falling asleep at night, sleeps badly.
3 Experience additional symptoms such as: change of mood, stomach ache, palpitations, perspiring.

APPENDIX 2: ADMINISTRATION OF THE EK SCALE

The administration of the EK scale consists of a question to the individual and his helper on how the task is performed in daily life. The items are scored according to the explanation given and observation of the performance.

1. How do you drive your wheelchair (items 0–3)? Please show me how you do it (items 0–2).
2. How do you transfer from the wheelchair (items 0–3)? Please show me how you do it (items 0–1).
3. Do you stand up (items 0–3)? How do you stand? Please show me or explain to me how you do it (items 0–2).
4. Do you change position in the wheelchair (items 0–3)? Please show me how much you can lean forwards and to the sides and get back to the upright position (items 0–2).
5. Can you move your fingers, hands and arms against gravity (items 0–3)? Please show me how you do it (items 0–3).
6. How do you feed yourself (0–3)? Please show me or explain to me how you do it (items 0–2).
7. How do you turn in bed during the night (items 0–3)? Please explain to me how you do it (items 0–1) and how often (items 2–3).
8. What do you do to produce the most effective cough (items 0–3)? Please show and explain to me how you do it (items 0–3).
9. Do you speak loudly and clearly enough to make people understand you at the other end of the classroom (items 0–3)?
10. How is your physical wellbeing (items 0–3)? Category 10 focuses on symptoms of respiratory insufficiency and the descriptions of the items from 0–3 are used for questioning and scoring for instance: do you need to rest during the day? Do you sleep well during night? How is your appetite?

APPENDIX 3: STATISTICAL TREATMENT FOR DETERMINING RELIABILITY COEFFICIENTS

First a large model was fitted, including sources of variance from subjects, evaluators, interactions between subjects and evaluators, as well as residual error. In this model, different groups of observers were allowed to have different magnitudes of variation. Moreover, the mean of EK sum was allowed to depend on the group of evaluators.

This large model was successively reduced by removing components in a stepwise reduction analysis where the statistical significance of each component was tested. Where two successive models both fit the data, the deviance (i.e. twice the log likelihood ratio) between the models has an approximate chi-squared distribution with degrees of freedom equal to the difference in the number of free parameters in the models. This well-known statistical property provides justification for the removal of components on the basis of their p-value.

Maximum likelihood was used for inference about fixed effects (dependence of mean EK sum on group of evaluators), whereas restricted maximum likelihood was used for inference about variance components.

In every case the basic assumptions of approximate normal distribution of residuals and best linear unbiased predictors (BLUPs) were examined by graphical methods (22).

If data are well behaved, they can be described by the model

$$X_{ijk} = \mu + a_i + b_{ij} + e_{ijk}$$

where: X_{ijk} is the EK sum for ith subject rated by the jth evaluator at the kth repetition; μ is the overall mean; a_i is the deviation from mean due to ith subject; b_{ij} is the deviation from the level of the ith subject due to interaction between ith subject and jth evaluator; e_{ijk} is the residual error; and a_i, b_{ij} and e_{ijk} are normally distributed with mean 0 and variances τ^2, ν^2, and σ^2, respectively. In this case the inter-rater reliability coefficient, defined as the correlation between ratings on the same subject performed by different evaluators, is $\tau^2/(\tau^2 + \nu^2 + \sigma^2)$. The intra-rater reliability coefficient, defined as the correlation between ratings on the same subject performed by the same evaluator, is $(\tau^2 + \nu^2)/(\tau^2 + \nu^2 + \sigma^2)$.

The 95% credibility intervals of the individual estimates of EK sum can be calculated in the following way. For a fixed subject i, the standard deviation in EK sum rated by different evaluators is $\sqrt{\nu^2 + \sigma^2}$, and hence, if the inter-rater reliability is high, the corresponding 95% credibility interval of the true EK sum is $X_{ijk} \pm 1.96 \times \sqrt{\nu^2 + \sigma^2}$, where X_{ijk} is the rating performed by a randomly chosen evaluator. The standard deviation of EK sum for a fixed subject within a fixed evaluator j is $\sigma = \sqrt{\sigma^2}$, and the corresponding 95% credibility interval of the true EK sum is $X_{ijk} \pm 1.96 \times \sigma$.

REFERENCES

1. Emery AEH. Population frequencies of inherited neuromuscular diseases-a world survey. Neuromusc Disord 1991;1:19–29.

2. European Consortium on Chronic Respiratory Insufficiency. Workshop report. Respiratory insufficiency and ventilatory support. 39th ENMC international workshop, 26–28 January 1996. Neuromusc Disord 1996; 6: 431–5.

3. Fowler Jr WM. Rehabilitation management of muscular dystrophy and related disorders: II. Comprehensive care. Arch Phys Med Rehab 1982;63:322–38.

4. Vignos Jr PJ. Physical models of rehabilitation in neuromuscular disease. Muscle Nerve 1983;6:323–38.

5. Vignos Jr PJ, Spencer GE, Archibald KC. Management of progressive muscular dystrophy of childhood. JAMA 1963;184:89–110.

6. Brooke MH, Griggs RC, Mendell JR, Fenichel M, Shumate JB, Pellegrino J. Clinical trial in Duchenne dystrophy. I. The design of the protocol. Muscle Nerve 1981;4:186–97.

7. Hiller LB, Wade CK. Upper extremity functional assessment scales in children with Duchenne muscular dystrophy: a comparison. Arch Phys Med Rehab 1992;73:527–34.

8. Wagner NB, Vignos PJ, Carlozzi C, Hull AL. Assessment of hand function in Duchenne muscular dystrophy. Arch Phys Med Rehab 1993;74:801–4.

9. Wade DT. Measurement in neurological rehabilitation. Oxford: Oxford University Press, 1992:35.

10. Steffensen BF, Hyde SA, Lyager S, Mattsson E. Validity of the EK scale – a functional assessment of individuals with Duchenne muscular dystrophy or spinal muscular atrophy, who are non ambulatory. Phys Research Int 2001;16:119–34.

11. Keating J, Matyas T. Unreliable inferences from reliable measurements. Aust J Physiother 1998;44:5–10.

12. Rothstein JM. Measurement and clinical practise: theory and application. In: Rothstein JM, editor. Measurement in physical therapy. New York: Churchill Livingstone, 1985:5–15.

13. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped gait-analysis assessments. Phys Ther 1991;71:465–72.

14. Gowland C, Boyce WF, Wright V, Russell DJ, Goldsmith CH, Rosenbaum PL. Reliability of the gross motor performance measure. Phys Ther 1995;75:597–602.

15. Hooper J, Tayler R, Pentlon B, Whittle IR. Rater reliability of Fahn's tremor rating scale in patients with multiple sclerosis. Arch Phys Med Rehabil 1998;79:1076–9.

16. Hughes KA, Bell F. Visual assessment of hemiplegic gait following stroke: pilot study. Arch Phys Med Rehabil 1994;75:1100–7.

17. Jelles F, Van Bennekom CAM, Lankhorst GJ, Sibbel CJP, Bouter LM. Inter- and intra-rater agreement of the rehabilitation activities profile. J Clin Epidem 1995;48:407–16.

18. Myhr U, von Wendt L, Sandberg KW. Assessment of sitting in children with cerebral palsy from videofilm: a reliability study. Phys Occup Ther Ped 1993;12:21–35.

19. Shields RK, Enloe LJ, Evans RE, Smith KB, Steckel SD. Reliability, validity, and responsiveness of functional tests in patients with total joint replacement. Phys Ther 1995;75:169–79.

20. Emery AEH, ed. Diagnostic criteria for neuromuscular disorders, 2nd edn. London: Royal Society of Medicine Press, 1997:1–4.

21. Altman DG. Practical statistics for medical research, 1st edn. London: Chapman & Hall, 1991:406–9.

22. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. SAS System for mixed models. Cary, NC: SAS Institute Inc.; 1996:135–70,229–51.

23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull 1979;2:420–8.

24. Feinstein GG, Cichetti DV. High agreement but low kappa. I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–9.

25. Jette AM, Deniston OL. Inter-observer reliability of a functional status assessment instrument. J Chronic Dis 1978;31:573–80.

26. Jette AM. Functional status index: reliability of a chronic disease evaluation instrument. Arch Phys Med Rehabil 1980;61:395–401.

27. Avlund K, Thudium D, Davidsen M, Fuglsang-Sørensen B. Are self-ratings of functional ability reliable? Scand J Occup Ther 1995;2:10–6.

28. Perris C, Ericsson U, Jacobsson L, Lindstrom H, Perris H. Interprofessional communicability and reliability of the Comprehensive Psychopathological rating Scale (CPRS) as assessed by videotaped interviews. Acta Psychiatr Scand 1979;60:144–8.

29. Wiseman MR, Vizard E, Bentovim A, Leventhal J. Reliability of video taped interviews with children suspected of being sexually abused. BMJ 1992;304:1089–91.

ADDRESS FOR CORRESPONDENCE:
Birgit F. Steffensen
Institute for Neuromuscular Diseases
Kongsvang Allé 23
DK-8000 Århus C
Denmark
Tel: +45-89482222
Fax: +45-89482238
E-mail: bist@muskelsvindfonden.dk

Submitted 23 May 2001;
accepted for publication 20 November 2001