

Clinical Trials in Duchenne Dystrophy

Standardization and Reliability of Evaluation Procedures

JULAIN M. FLORENCE,
SHREE PANDYA,
WENDY M. KING,
JENNY DIVELEY ROBISON,
LINDA C. SIGNORE,
MINDY WENTZELL,
and MICHAEL A. PROVINCE*

A multiclinic, collaborative study has been designed to assess the natural progression and efficacy of treatment of Duchenne muscular dystrophy. This article describes the protocol for the evaluation technique and the method used to establish within (intraobserver) and between (interobserver) reliability of the protocol evaluation procedures. Standardized patient evaluations were used, and consistency of evaluation was monitored by a computer. The reliability of the measures was analyzed 1) within observers by comparing the results of each of the first three tests done by each evaluator for all patients and 2) between observers by comparing, at multicenter group meetings, the results of each of the four evaluators' tests of the same patient. We have demonstrated reliability for an evaluation method that will provide an objective foundation on which to claim a drug or therapeutic procedure does or does not have an effect in treating Duchenne muscular dystrophy.

Key Words: *Clinical trials, Muscular dystrophy, Physical therapy.*

Throughout the history of therapeutic trials in Duchenne muscular dystrophy (DMD), various drugs and treatment procedures were thought to be beneficial. The medical community, patients and patients' families have hoped that the therapeutic procedures would improve the progress but have been disappointed after careful examination showed these treatments ineffective.^{1,2}

To establish the effect of any treatment on the natural course of a disease, standardized and reliable evaluation procedures are needed.^{1,2} Only through such documentation can the medical community ensure that claims about the effectiveness of a treatment are substantiated.

Ms. Florence is Coordinator Clinical Research, Irene Walter Johnson Rehabilitation Institute, and Research Assistant Professor, Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110 (USA).

Mrs. Pandya is Instructor, Department of Neurology, University of Rochester, Rochester, NY 14642.

Mrs. King is Therapy Supervisor, Muscular Dystrophy Clinic, Department of Neurology, Ohio State University, Columbus, OH 43210.

Mrs. Robison is Staff Physical Therapist and Chief Physical Therapist of Muscular Dystrophy Clinic, Vanderbilt University Medical Center, Nashville, TN 37212.

Mrs. Signore is Clinical Nurse Coordinator, Muscular Dystrophy Clinic, Department of Neurology, Ohio State University, Columbus, OH 43210.

Mrs. Wentzell is Chief of Physical Therapy, U.S. Naval Submarine Medical Center, New London, Groton, CT 06320.

Mr. Province is Statistician, Division of Biostatistics, Washington University Medical School, St. Louis, MO 63110.

* Other members of the Collaborative Investigation of Duchenne Dystrophy Group are Jerry R. Mendell, MD, Ohio State University; Robert C. Griggs, MD, and Richard Moxley, MD, University of Rochester; Gerald M. Fenichel, MD, Vanderbilt University; Michael H. Brooke, MD, Kenneth K. Kaiser, BS, Martha McCrate, BS, Philip Miller, AB, and Jack B. Schumate, MD, Washington University; and Phillip Bach, PhD, CLMG Clinical Laboratory.

This article was submitted September 24, 1982; was with the authors 15 weeks for revision; and was accepted July 11, 1983.

We have designed and established the reliability of a protocol to document the natural history and evaluate the results of treatment for DMD.³ The protocol has been in use for three years in a four-clinic collaborative study. This paper briefly describes the protocol³ of an ongoing study and, in further detail, describes the within (intraobserver) and between (interobserver) reliability of the protocol procedures during the two-year period of this report.

METHOD

Protocol

The protocol design is for a multiclinic, collaborative study. The team in each clinic consists of a principal investigator and clinical evaluator. The principal investigator is a physician who is responsible for the overall supervision of the protocol. The clinical evaluator is a physical therapist who is responsible for the assessment of the patient on each visit. All evaluators spend at least three days in a training workshop at the protocol coordinating center before they start the protocol.

Procedure

The patients admitted to the study fulfilled specific selection criteria to ensure they had DMD or Becker muscular dystrophy, the slowly progressive X-linked muscular dystrophy.³ The patients were evaluated on admission to the study; after 1 week, 4 weeks, and 8 weeks; and every 12 weeks thereafter. The evaluation included tests of muscle strength, range of motion (ROM), pulmonary function, and functional activities. Each assessment was performed without access to previous results to prevent any duplication or bias.

MUSCLE SCORE

Intraobserver ICC .95

Interobserver ICC .90

MUSCLE TESTING CONSISTENCY - INDIVIDUAL EVALUATORS

Evaluator	4	3	1
2	Individual muscles graded: 306 % Consistent: 93%	Individual muscles graded: 340 % Consistent: 83%	Individual muscles graded: 374 % Consistent: 93%
4		Individual muscles graded: 340 % Consistent: 91%	Individual muscles graded: 340 % Consistent: 88%
3			Individual muscles graded: 374 % Consistent: 88%

Fig. 1. Interobserver consistency (ICC) and reliability in grading muscle strength.

The following muscle groups were tested for strength: shoulder abductors and external rotators; elbow and wrist flexors and extensors; neck flexors and extensors; hip flexors, extensors, and abductors; knee flexors and extensors; ankle dorsiflexors, plantar flexors, invertors, and evertors. The position and procedure for testing each muscle group was defined in the protocol. The grading system used was a modification of the Medical Research Council (MRC) Scale (Appendix 1).⁴ The modification from the MRC scale was in Muscle Grade 4—the muscle group moves the joint against gravity and against some resistance. The rationale for our modifications and the procedure for selecting the reference muscle are described elsewhere.³ The individual muscle grades were allotted points (Appendix 1). These points were added to give the patient a “total muscle score.” Thirty-four individual muscle groups were tested. A person with normal strength would receive a total muscle score of 340.

Joint ROM was measured using a modified version of the system published by the American Academy of Orthopaedic Surgeons and British Orthopaedic Association.⁵ Measurements were taken with a goniometer at the shoulders, elbows, wrists, hips, knees, and ankles. The position and procedure for each joint measurement were strictly defined. The positions of measurement were described by Brooke et al.³ A “contracture index” was derived by comparing the measured ROM with the maximal “normal” ROM for that joint.⁵ A percentage contracture was derived for each joint and the average percentage of all 14 joints was the contracture index.

The pulmonary function testing included measuring forced vital capacity (FVC), maximum voluntary ventilation (MVV), and maximum expiratory pressure (MEP). The FVC and

MVV were measured by an electronic spirometer. The procedures for measurement are described elsewhere.⁶ The MEP was measured with a mercury manometer as described by Black and Hyatt.⁷

The functional testing consisted of graded functions of the upper and lower extremities and a series of timed tests. The upper and lower extremity functional grades indicated the severity of involvement in the extremity (Appendix 2). The hips and legs scale was based on the functional grading system developed by Vignos et al.⁸ The timed functions included standing up from a supine position, climbing four standard stairs, running or propelling a wheelchair 9 m (30 ft), rising from a chair, putting on a T-shirt, and cutting out a 8-cm (3-in) square. Each test was standardized using defined start and stop positions and using identical equipment in each clinic.

Intraobserver and Interobserver Reliability

The protocol was designed to monitor both the intraobserver and interobserver consistency to determine the reliability of our evaluation procedures. All test results were submitted to a computer that monitored for inconsistent or missing values and nonsensical errors. Intraobserver consistency was monitored by comparing the results of the first three tests done by each evaluator for all patients within the first month after admission to the study. Because the patient's condition was unlikely to change within this period, we attributed discrepancies in the ratings to clinical error.⁹ Interobserver reliability was evaluated at multicenter group meetings that were held on a regular basis. At these meetings, all four evaluators individually tested the same patients on the same day. Over a two-year period, seven meetings were held, and a total of 21 patients were tested. The patients ranged from age 5 years to 27 years, functional grades 1 to 9, and muscle scores 144 to 297.

Reliability of both the interobserver and intraobserver measurements were analyzed by a random-effects model analysis of variance, the intraclass correlation coefficient (ICC).¹⁰

Though muscle scores are not a classical interval or ratio scale, parametric (ICC) analysis was deemed appropriate because of its linear relationship to other variables.¹¹ Nonparametric analysis (kappa) yielded similar results to ICC, but for purposes of clarity, only the ICCs are reported here.

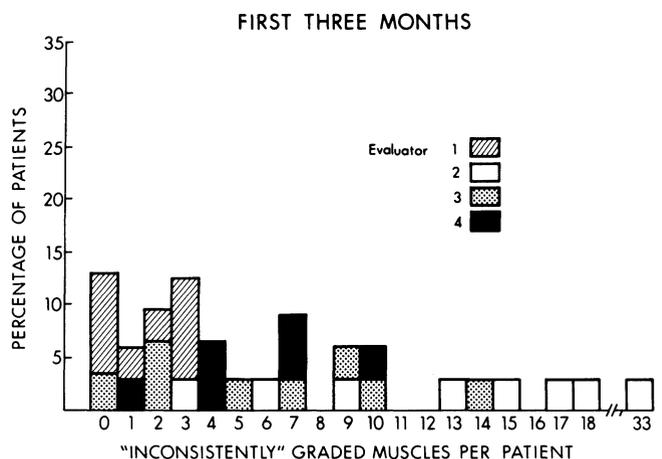


Fig. 2. Intraobserver inconsistency of muscle strength grading during the first three months of study (0 = no inconsistency).

Functional Testing

	<u>INTRA ICC</u>	<u>INTER ICC</u>
UE	.92	.87
LE	.99	.96

PULMONARY FUNCTION

FVC	.97	.98
MVV	.95	.97
MEP	.82	.69

TIMED FUNCTIONS

Stand from supine	.99	.91
Climb 4 standard stairs	.91	.04
Traverse 30 ft.	.97	.91
Push wc 30ft.	.81	.87
Arise chair	.99	.86
Put on T-shirt	.84	.90
Cut out 3" square	.88	.76

Fig. 6. Intraobserver and interobserver consistency (ICC) and reliability in pulmonary and functional testing.

and variance of each measurement can be determined. Hence, we know how much variance in the examination can be attributed to evaluation inconsistencies. This type of infor-

mation is essential in determining the validity of any therapeutic trial.

CONCLUSIONS

We designed a protocol for assessing the natural history and treatment of DMD. The reliability of all physical measurements has been analyzed, and we now have an objective foundation on which to claim a drug or procedure does or does not have an effect in the treatment of DMD.

REFERENCES

1. Thompson WHS, Smith I: X-linked recessive (Duchenne) muscular dystrophy (DMD) and purine metabolism: Effects of oral allopurinol and adenylate. *Metabolism* 27:151-163, 1978
2. Mendell JR, Weissberg DO: Lack of benefit of allopurinol in Duchenne dystrophy. *Muscle Nerve* 2:53-56, 1979
3. Brooke MH, Griggs RC, Mendell JR, et al: Clinical trial in Duchenne dystrophy. I. The design of the protocol. *Muscle Nerve* 4:186-197, 1981
4. Medical Research Council. Aids to the evaluation of peripheral nerve injuries. London, England, Her Majesty's Stationery Office, 1943, pp 11-46
5. Heck CV, Henderson IE, Rowe CR: Joint Motion, Method of Measuring and Recording. London, England, Churchill Livingstone Inc, 1966, pp 10-69
6. Florence JM, Brooke MH, Carroll JE: Evaluation of the child with muscular weakness. *Orthop Clin North Am* 9:409-430, 1978
7. Black LF, Hyatt RE: Maximal respiratory pressures: Normal values and relationships to age and sex. *Am Rev Respir Dis* 99:696-702, 1969
8. Vignos PJ, Spencer GE, Archibald KC: Management of progressive muscular dystrophy in childhood. *JAMA* 184:89-96, 1963
9. Brooke MH: A Clinician's View of Neuromuscular Diseases. Baltimore, MD, Williams & Wilkins, 1977, pp 95-109
10. Scheffe H: The Analysis of Variance. New York, NY, John Wiley & Sons Inc, 1959, pp 221-258
11. Brooke MH, Fenichel GM, Griggs RG, et al: Clinical investigation in Duchenne Dystrophy. 2. Determination of the "power" of therapeutic trials based on the natural history. *Muscle Nerve* 6:91-103, 1983

APPENDIX 1 Strength Grading Scale

Muscle Grade	Definition	Points Awarded
5	Normal strength.	10
5-	Barely detectable weakness.	9
4S	Same as 4 but stronger than reference muscle.	7
4	Muscle is weak but moves the joint against a combination of gravity and some resistance.	7
4W	Same as 4 but weaker than reference muscle.	7
3+	The muscle is capable of transient resistance but collapses abruptly. This degree of weakness is difficult to put into words, but it is a muscle that is able to move the joint against gravity and an additional small amount of resistance. It is not to be used for muscles capable of sustained resistance throughout the whole range of movement.	5
3	Muscle cannot move against resistance but moves the joint fully against gravity. With the exception of knee extensors, the joint must be moved through the full mechanical range against gravity. If a patient has contractures that limit movement of the joint, the mechanical range will obviously be to the point at which the contractures cause a significant resistance to the movement.	4
3-	Muscle moves the joint against gravity but not through the full extent of the mechanical range of the joint.	3
2	Muscle moves the joint when gravity is eliminated.	2
1	A flicker of movement is seen or felt in the muscle.	1
0	No movement.	0

APPENDIX 2 Functional Grades

Grade**Lower Extremities**

- | | |
|----|---|
| 1 | Walks and climbs stairs without assistance. |
| 2 | Walks and climbs stairs with aid of railing. |
| 3 | Walks and climbs stairs slowly with aid of railing (over 12 seconds for four standard steps). |
| 4 | Walks unassisted and rises from a chair but cannot climb stairs. |
| 5 | Walks unassisted but cannot rise from a chair or climb stairs. |
| 6 | Walks only with assistance or walks independently with long leg braces. |
| 7 | Walks in long leg braces but requires assistance for balance. |
| 8 | Stands in long leg braces but unable to walk even with assistance. |
| 9 | Is in wheelchair. |
| 10 | Confined to bed. |
-

Upper Extremities

- | | |
|---|---|
| 1 | Starting with arms at the sides, the patient can abduct the arms in a full circle until they touch above the head. Can place a weight of 0.5 kg or more on a shelf above eye level. |
| 2 | Can raise arms above head as previously but cannot place a 0.5 kg weight on a shelf. |
| 3 | Can raise arms above head only by flexing the elbow (ie, shortening the circumferences of the movement) or using accessory muscles. |
| 4 | Cannot raise hands above head but can raise 237 mL (8 oz) glass of water to mouth. |
| 5 | Cannot raise hands to mouth but can use hands to hold pen or pick up pennies from table. |
| 6 | Cannot raise hands to mouth and has no useful function of hands. |
-

APPENDIX 3 Consistency Criteria

Strength grading

Consistent—Strength evaluation of a given muscle did not differ by more than one consecutive step (ie, 3– to 3, 4S to 5–) if the highest rating was 2 or more.

Possible—Strength evaluation of a given muscle differed by two consecutive steps (ie, 3– to 3+) if the highest rating was strength 3 or more.

Inconsistent—Any differences in strength other than those described above.

Joint range measurement

Consistent—Measurement difference by 10 degrees or less.

Possible—Difference between 10 degrees and 15 degrees.

Inconsistent—Greater than 15 degrees difference.
