Outcome Measures for Pediatric Spinal Muscular Atrophy

Susan T. Iannaccone, MD; and the American Spinal Muscular Atrophy Randomized Trials (AmSMART) Group

Background: Spinal muscular atrophy (SMA) is a genetic disease of the anterior horn cell with a frequency of 8 per 100,000 live births and a high rate of mortality during infancy. The American Spinal Muscular Atrophy Randomized Trials (AmSMART) Group is an organization of 5 centers formed to perform clinical trials in children with SMA.

Objective: To devise reliable methods to measure strength, motor function, lung function, and quality of life for use as outcome measures in children with SMA.

Setting: Tertiary referral center, pediatric neurology department.

Patients and Methods: Twelve children with SMA aged 2 to 14 years were enrolled in a reliability study of 4 outcome measures: quantitative muscle testing (in children <5 years), gross motor function measure, pulmonary function tests, and quality of life. The Richmond Quantitative Measurement System was used to test grip, knee flexion and extension, and elbow flexion. Gross motor function measure was performed as described, and pulmonary function tests were measured using the KoKo system. Quality of life was assessed via the PedsQL and the PedsQL Neuromuscular Module for patients and parents.

Results: Ten children fulfilled the inclusion criteria and completed at least 3 visits with 3 evaluators in 6 months. Using a weighted κ, the gross motor function measure showed high interrater reliability. Quantitative muscle testing showed greater variability among the weakest children; the findings for pulmonary function tests and quality of life were inconclusive. The PedsQL Neuromuscular Module for parents had moderately high reliability.

Conclusion: A tool for motor function may be more useful in clinical trials of childhood SMA than one for quantitative muscle strength.

Arch Neurol. 2002;59:1445-1450

POTENTIAL CONFLICT OF INTEREST: None.

S P I N A L M U S C U L A R atrophy (SMA) (OMIM 253300) is a genetic disease of the anterior horn cell with a frequency of 8 per 100,000 live births.1-3 It has been classified into 3 types according to age at onset, severity of disease, and motor milestones achieved. Spinal muscular atrophy type I begins before age 6 months and has a high rate of mortality during infancy. The prevalence rate is greatest for SMA types II and III, which are associated with later onset and lower mortality rates during middle and late childhood.4,5 Death almost always is secondary to severe restrictive lung disease that is progressive, although muscle strength may be stable for decades.6 There is no known treatment for SMA. Until recently, no therapeutic trials have been attempted. The disease is caused by mutation of the SMN1 gene, the product of which is called SMN protein.7,9 New information regarding the nature and function of SMN protein in SMA and the availability of new pharmacologic agents now make it possible to consider clinical trials in this disease.10 Our goal is to devise reliable methods to measure lung function, strength, motor function, and quality of life in children with SMA.

From the Department of Neurology, Texas Scottish Rite Hospital for Children, Dallas. A complete list of the members of the AmSMART Group appears at the end of this article.
quantitative muscle testing (QMT) in children older than 5 years, (3) gross motor function measure (GMFM), and (4) quality of life (QOL).

Patients were recruited from the Pediatric Neuromuscular Clinic at Texas Scottish Rite Hospital for Children and were screened for inclusion and exclusion criteria. Inclusion criteria were as follows: 2 to 18 years old, weakness and a clinical diagnosis of SMA confirmed by mutation analysis of the SMN1 gene, forced vital capacity greater than 20% of predicted for age, less than 15% variance on test-retest using QMT after instruction, and informed consent. Exclusion criteria were as follows: no clinically significant evidence of renal dysfunction, central nervous system damage, or neurodegenerative or neuromuscular disease other than SMA and no mechanical ventilation of any type for more than 16 h/d.

The evaluation room was located in the physical therapy department and contained all equipment used for an evaluation session, including the examining table (Figure 2), the Richmond Quantitative Measurement System for QMT, items for use in the GMFM, and chairs for parents. Sessions for individual patients were scheduled at the same time of day (eg, morning or afternoon), with at least 48 hours between sessions. Individual parts of the evaluation session were performed in the same order for all patients. Positioning for PFTs and QMT was consistent for all patients. Rest periods during the evaluation session were as follows: 30 to 60 seconds between attempts, 15 minutes between PFTs and QMT, and 15 minutes between QMT and the GMFM.

Pulmonary function tests were performed according to American Thoracic Society standards and included maximum inspiratory pressure, maximum expiratory pressure, cough pressure (peak cough flow), forced vital capacity, and forced expiratory volume in the first second. Lung volumes were measured using the KoKo spirometer system (Pulmonary Data Services, Inc, Louisville, Colo), which calculates percentage predicted based on height and age.
The Richmond Quantitative Measurement System was used to test (1) right and left grip, (2) right and left knee extension, (3) right and left knee flexion, and (4) right and left elbow flexion. Each patient had 3 attempts for each muscle group, and the computer recorded the best of 3 test results on prompting from the evaluator. Strength was recorded in pounds. Patients sat during testing of muscle groups 1, 2, and 3 and were supine for testing of group 4.

The GMFM contained 88 items in 5 dimensions: lying and rolling (A), sitting (B), crawling and kneeling (C), standing (D), and walking, running, and jumping (E). Each patient continued through all dimensions according to his or her abilities. Each item was scored individually and weighted equally.

Ten QOL forms were administered by the study coordinator (K.R.) at the beginning of the first and last visits. The generic PedsQL included a parent questionnaire for each of 4 age groups: 2 to 4, 5 to 7, 8 to 12, and 13 to 18 years. In addition, there was a child questionnaire for each of 3 age groups: 5 to 7, 8 to 12, and 13 to 18 years. The AmSMART child psychologist developed a specific questionnaire, the PedsQL Neuromuscular Module, with one version for parents and 2 for children aged 5 to 7 and 8 to 18 years. All procedures were approved by the institutional review board of The University of Texas Southwestern Medical Center, Dallas.

TRAINING

All evaluators were licensed physical therapists with pediatric experience. Two evaluators in Dallas (D.C. and J.G.) were responsible for training evaluators from other centers. One consultant evaluator from Richmond, Va (J.M.), who had previous experience with QMT methods provided additional training. The respiratory therapist from Dallas (K.H.) provided training in the performance of PFTs. The Dallas group developed and distributed a training videotape and a manual for QMT and PFTs. From March 4 to 9, 2001, all evaluators met in Dallas for 1 week, during which they trained using the Richmond Quantitative Measurement System equipment and conducted evaluation sessions 3 and 4 for patients in phase 1.

DATA MANAGEMENT

All data were transcribed by evaluators onto case report forms (CRFs) developed in the Department of Academic Computing Services at The University of Texas Southwestern Medical Center. The CRFs were organized in a binder for each patient. The CRFs were checked carefully by the study coordinator (K.R.) and the principal investigator (S.T.I.), and a checklist was completed for each CRF before it was sent to the Department of Academic Computing Services for data entry and verification. Unique identifiers for a CRF were site number, patient number, patient initials, and visit number. The Department of Academic Computing Services at The University of Texas Southwestern Medical Center developed and distributed a variance calculator program to assist in determining whether the patient passed the inclusion criteria of less than 15% variance for right and left grip. The average interrater reliability (weighted $\kappa$ statistic) for the 37 items from dimensions A and B and a nonparametric Friedman test for repeated measurement for dimensions A and B and the total GMFM. Cronbach coefficient $\alpha$ was calculated for the parent version of the PedsQL Neuromuscular Module, and test-retest reliability was performed for the first and last visits.

RESULTS

Ten children (4 girls and 6 boys) fulfilled the inclusion criteria and completed at least 3 visits with 3 of the 6 evaluators in 6 months. The children were aged 2 to 14 years (mean, 7.4 years). There was 1 walker and 9 nonwalkers.

PULMONARY FUNCTION TESTS

Six patients completed PFTs; 1 patient did not complete the peak cough flow measure. Analyses using the percentage change measure for all 5 measures of PFT (percent forced vital capacity, percent forced expiratory volume in the first second, maximum inspiratory pressure, maximum expiratory pressure, and peak cough flow) showed that 53% of measurements were within 10% and 73% were within 15% of each other across 3 evaluators (Table 1). Review of individual test results showed that there was consistency for individual patients in any given visit, but patients' results sometimes varied between visits secondary to reactive airway disease or illness.

QUANTITATIVE MUSCLE TESTING

Six patients completed QMT; 1 patient was missing data for right and left grip. The percentage change measure for 8 muscle groups showed that 18% of measurements were within 10% and 22% were within 15% of each other across 3 evaluators (Table 2).

GROSS MOTOR FUNCTION MEASURE

Preliminary analyses of the GMFM data showed that all patients could complete the items from dimensions A (lying and rolling) and B (sitting), but many were too weak to attempt dimensions C, D, and E. The average interrater reliability (weighted $\kappa$ statistic) for the 37 items from dimensions A and B using data from visits 2, 3, and 4 was excellent ($\kappa = 0.72$). The 88 items of the GMFM were ana-

---

Table 1. Summary Data for Pulmonary Function Tests in 6 Patients With Spinal Muscular Atrophy

<table>
<thead>
<tr>
<th>Measure</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forced vital capacity†</td>
<td>123</td>
<td>170</td>
<td>71.9</td>
<td>67</td>
</tr>
<tr>
<td>Forced expiratory volume in the first second†</td>
<td>123</td>
<td>199.3</td>
<td>68.0</td>
<td>66</td>
</tr>
<tr>
<td>Maximum inspiratory pressure, cm H2O</td>
<td>100</td>
<td>84</td>
<td>68.2</td>
<td>69</td>
</tr>
<tr>
<td>Maximum expiratory pressure, cm H2O</td>
<td>4</td>
<td>10</td>
<td>50.2</td>
<td>50</td>
</tr>
<tr>
<td>Peak cough flow, L/min†</td>
<td>335</td>
<td>100</td>
<td>199.3</td>
<td>170</td>
</tr>
</tbody>
</table>

*Percentage predicted for height.
†One patient did not complete the peak cough flow.

---
tor neurons from these mice may reveal compounds that are capable of up-regulating or modifying the SMN2 gene to produce more protein or more normal protein. Using the technique of “high throughput screening,” researchers estimate that they can test as many as 100000 compounds in a year. As soon as safety testing can be completed, candidate compounds could be ready for clinical trials. Thus, there may be improved rationale for drug treatment of SMA within 2 years.

Objective and reliable measures of disease state in childhood SMA have not yet been defined. All prospective and retrospective studies to date depend on measurement of strength, assumed to reflect the pathophysiologic process of motor neuron death. Standard methods of strength testing, such as the manual muscle test, are not useful in this population because patients are so weak that most muscle groups will not register more than 1 or 2 of a possible 5. Moreover, the scale is ordinal so that statistical analysis is problematic and small changes in strength cannot be measured. Thus, several groups have attempted to use QMT with devices that are sensitive to very weak force, including the handheld dynamometer and the fixed myometer, the Chatillon (AMETEK, Inc, Paoli, Pa), both of which showed varying degrees of reliability and sensitivity.

One such study, the DCN/SMA study, was conducted by 3 of the researchers involved in the present project. Between 1988 and 1994, 73 patients with SMA (41 girls and 32 boys; mean age, 17.3 years) were evaluated. Of the 14 muscle groups that underwent QMT, results from 6 (3 on each side) were considered unreliable. Thus, measures of shoulder abduction, shoulder adduction, and elbow extension were eliminated from the total muscle score.

There was an increment in the total muscle score for patients younger than 15 years that was statistically significant. Patients with SMA overall showed an increase of 2.2 kg in a mean of 4 years, or an increment of approximately 0.5 kg/y. A similar suggestion of improved motor function was seen in a cohort of patients with SMA younger than 6 years who were evaluated using a motor function scale. Such improvement during early childhood may be attributed to growth and development but remains way below what is expected for healthy children. For patients older than 15 years, there was no increment or decrement in the total muscle score.

Most important, no patient showed loss of strength during the observation. Based on our data, we suggested several options to obtain statistically reliable results in a clinical trial: (1) monitor motor function for at least 12 months using a sensitive tool yet to be developed; (2) monitor for an increase in QMT; (3) develop a strength measurement that is more sensitive and more reliable than QMT; (4) increase the sample size; or (5) increase the duration of the study. Because the number of patients and the length of the study affect the feasibility and cost of completing a study, it would be practical to develop sensitive methods of monitoring strength and motor function.

Until recently, there were few valid or reliable tools for measure of motor function in children. Those were of no use for nonambulatory children and offered no discriminatory increments for small changes in strength. The

| Table 2. Summary Data for Quantitative Muscle Testing by Muscle Group in Patients With Spinal Muscular Atrophy |
|-----------------|---------|---------|-------|------|
| Muscle Group    | Maximum | Minimum | Mean  | Median |
| Right grip (n = 5) | 6.71    | 0.32    | 2.84  | 2.71 |
| Left grip (n = 5)  | 6.60    | 0.79    | 3.62  | 2.79 |
| Right knee extensor (n = 6) | 7.95    | 0.28    | 1.83  | 1.23 |
| Left knee extensor (n = 6) | 9.02    | 0.20    | 2.10  | 1.15 |
| Right knee flexor (n = 6)  | 11.21   | 0.22    | 3.55  | 2.58 |
| Left knee flexor (n = 6)  | 13.45   | 0.44    | 3.15  | 1.62 |
| Right elbow flexor (n = 6) | 11.35   | 2.07    | 4.84  | 3.65 |
| Left elbow flexor (n = 6)  | 12.95   | 2.03    | 4.88  | 3.60 |

*To convert to kilograms, multiply by 0.45.
GMFM has been tested for validity and reliability but only in patients with cerebral palsy and not in those with neuromuscular disease. The GMFM contains 88 items, each of which has a graded score. The items are grouped according to body region, such as upper extremities, and some items reportedly may be omitted without affecting its validity. The GMFM seems to be excellent for evaluating very weak children, such as those with SMA. Our results showed good consistency for interrater and intrarater testing. However, the number of patients for part 1 was small, and we have no data yet to indicate how sensitive to change this measure might be.

Pulmonary function may better reflect disease state than muscle strength because most patients with SMA die of respiratory failure. However, PFTs in children are difficult to perform in a consistent way and rarely have been used as outcome measures for clinical trials. Furthermore, all lung volume variables must be compared with age-matched controls as percentage predicted for age, which is always calculated on the basis of height. Height measures in patients with SMA are notoriously inaccurate because of orthopedic deformities such as scoliosis and flexion contractures.

A new requirement for clinical trials is measurement of the patient's perception of disease state, the QOL. Defined as the degree of satisfaction or dissatisfaction felt by people with various aspects of their lives, QOL may be measured for individuals and for families. Health-related QOL depends on concurrent lifestyles, past experiences, hopes for the future, dreams, and ambitions, with objective and subjective components. The pediatric health-related QOL may include functional performance and patients' evaluations of their own experiences, with objective and subjective components. Testa and Simonson described 3 primary domains in pediatric health-related QOL: physical, psychological, and social.

In addition to more general surveys of information, QOL instruments have been designed around specific childhood diseases, such as cancer and arthritis. These measures are commonly used in clinical trials. The PedsQL incorporates associated modules for work in oncology, asthma, or diabetes mellitus. The PedsQL includes subdomains assessing physical, emotional, school, and social functioning. Reliability, validity, and specificity were established, and it allowed the addition of disease-specific questions. We do not yet have enough data to determine whether this tool will be useful for SMA.

The AmSMART Group has made a first attempt to develop valid and reliable outcome measures for children with SMA aged 2 to 18 years, whether ambulatory or nonambulatory. Reliability results depended on the patients' and the evaluators' familiarity with the procedures. As in previous studies, we found the patients' learning curve to be 2 to 3 visits. Evaluators were especially uncomfortable with measuring PFTs. They lacked the judgment for recognizing less than full effort and did not recognize abnormal flow loops indicating that the patient had reactive airways. Therefore, we incorporated a follow-up training session of 2 days that included extensive discussion of normal lung physiology following American Thoracic Society guidelines. We concluded, moreover, that training sessions should be conducted at a time when data were not being collected so that evaluators could concentrate on learning technique without feeling pressure to collect data.

In conclusion, after reliability testing with a small number of patients, the GMFM seems to be the most reliable outcome measure used. A tool that measures motor function may be more useful in clinical trials of childhood SMA than one that measures quantitative muscle strength. Performance of PFTs in children requires special training. Evaluator training sessions should be separated in time and space from sessions used for collection of reliability data. Inconsistency of outcome measures may be secondary to patient, evaluator, or tool factors.

Accepted for publication April 10, 2002.

Author contributions: Study concept and design (Drs Iannaccone, Morton, Reisch, Schochet, Luckett, Samaha, Russman, and Lesher); acquisition of data (Drs Iannaccone, Morton, Schochet, and Wong; Mss Rabb, Carman, Gordon, Harris, J. Smith, Fritch, Zilke, Mayhew, and Stout; and Mr Webster); analysis and interpretation of data (Drs Iannaccone, Hynan, Schochet, Samaha, and S. Smith); drafting of the manuscript (Drs Iannaccone and Morton and Mss Rabb, Carman, Gordon, Harris, Fritch, Zilke, and Stout); critical revision of the manuscript for important intellectual content (Drs Iannaccone, Hynan, Reisch, Schochet, Luckett, Wong, Samaha, Russman, Lesher, and S. Smith and Ms J. Smith); statistical expertise (Dr Hynan); obtained funding (Dr Iannaccone); administrative, technical, and material support (Drs Iannaccone, Morton, Schochet, Luckett, Lesher, and S. Smith; Mss Rabb, Carman, Gordon, Harris, J. Smith, Zilke, and Mayhew; and Mr Webster); study supervision (Drs Iannaccone, Reisch, Schochet, Luckett, and Samaha and Mss Rabb and J. Smith).

This study was supported by grant 1-RO1-NS 39327-01A1 from the National Institutes of Health, Bethesda, Md, and by the Muscular Dystrophy Association, Tucson, Ariz.

Members of the American Spinal Muscular Atrophy Randomized Trials Group: Texas Scottish Rite Hospital for Children: Susan T. Iannaccone, MD; Karen Rabb, RN; Deanna Carman, BS, PT; Jennifer Gordon, MS, PT; Kathalene Harris, BS, RRT; and Anne Morton, PhD. The University of Texas Southwestern Medical Center: Linda Hynan, PhD; Joan Reisch, PhD; Janet Smith, BA; Joe C. Webster, BA; Peter Schochet, MD; Peter Luckett, MD; and Laura Herbelin. Children's Hospital Medical Center, Cincinnati, Ohio: Brenda Wong, MD; Fred Samaha, MD; and Ann Fritch, MS, PT. Shriner's Hospital for Children, Portland, Ore: Barry Russman, MD, and Kirsten Zilke, BS, PT. Children's Hospital, Richmond, Va: Robert Lesher, MD, and Jill Mayhew, BS, PT. Gillette Children's Specialty Health Care, St Paul, Minn: Stephen Smith, MD, and Jean Louis Stout, MS, PT.

Corresponding author and reprints: Susan T. Iannaccone, MD, Texas Scottish Rite Hospital for Children, 2222 Welborn St, Dallas, TX 75219 (e-mail: susan.iannaccone@tsrh.org).
REFERENCES