Reliability of 4 Outcome Measures in Pediatric Spinal Muscular Atrophy

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

Background: Spinal muscular atrophy is a common neurologic disorder of infants and children with a high mortality rate. Clinical trials have not been attempted in this population until recently.

Objective: To demonstrate that 4 outcome measures are reliable for use in clinical trials in patients with spinal muscular atrophy.

Design, Setting, Patients: Thirty-eight children with spinal muscular atrophy who fulfilled inclusion and exclusion criteria were enrolled at 5 pediatric centers for a reliability study. Paired samples statistics were performed comparing results of the qualifying variance visit with a fourth visit.

Main Outcome Measures: Quantitative muscle testing and the Gross Motor Function Measure.

Results: Thirty-four patients and 7 evaluators completed the study. Thirteen patients were aged 2 through 4 years and 21 were 5 through 17 years. The Gross Motor Function Measure was completed by 34 subjects. Six variables for pulmonary function tests were measured in 20 subjects. Quantitative muscle testing was performed on 21 subjects in 8 muscle groups. Thirty-three subjects completed the PedsQL Neuromuscular Module for Parents. The intraclass correlation coefficient and Bradley-Blackwood procedures indicated a very high level of agreement between measures.

Conclusion: The Gross Motor Function Measure, pulmonary function tests, quantitative muscle testing, and quality of life are reliable outcome measures for clinical trials in pediatric spinal muscular atrophy.

Arch Neurol. 2003;60:1130-1136

Spinal muscular atrophy (SMA) (OMIM 253300) is a genetic disease of the anterior horn cell with a frequency of 8 per 100,000 live births.1,2 Death almost always is secondary to severe restrictive lung disease that is progressive, although muscle weakness may be quite stable over decades.3,4 There is no known treatment for SMA. Until recently, no therapeutic trials have been attempted.5,6 Valid results of clinical trials depend on reliable outcome measures. For motor neuron disease, measures of muscle strength have been the most commonly used disease correlate. However, measuring muscle strength in children with SMA is very difficult because they are extremely weak. Even quantitative muscle testing, which has been the gold standard in adult motor neuron disease, has not been satisfactory in children with SMA because of large SDs.7,8 In studies of adults with SMA, inclusion criteria excluded the weakest subjects, who were unable to register strength during quantitative muscle testing.9 Such inclusion criteria would exclude most children with SMA from clinical trials. Thus, our goal was to find sensitive and reliable outcome measures that could be used for children with SMA beginning at age 2 years.

In a previous report,9 our group showed that a motor function tool, the Gross Motor Function Measure (a proprietary test available from the Neurodevelopmental Clinical Research Unit, Department of Clinical Epidemiology and Biostatistics, School of Rehabilitation Science, McMaster University, 1280 Main St W, Hamilton, Ontario, Canada L8S 4K1; attn: Diane Russell), appeared to be more reliable than quantitative muscle testing in children with SMA. That report described results in a small cohort of children (n=9) examined over 6 months by 6 evaluators. Interrater reliability was excellent for the Gross Motor Function Measure but not very good for quantitative muscle testing or pulmonary function tests. Data were insufficient to assess reliability for the PedsQL10 (a proprietary test avail-
able from Jim W. Varni, PhD, Center for Child Health Outcomes, Children's Hospital & Health Center, 3020 Children's Way, MC 5053, San Diego, CA 92123) quality-of-life tool, except for the PedsQL Neuromuscular Module for parents, which was found to have moderate intrarater reliability.

We now report results for intrarater reliability for the same outcome measures. Our group has shown in a long-term prospective study that patients with SMA are very stable with regard to strength. Thus, we made the assumption that individual patients would show identical results if tested by the same evaluator several times within a 6- to 8-week period. The results presented herein from this cohort show that the Gross Motor Function Measure, quantitative muscle testing, and pulmonary function tests have excellent reliability in children with SMA.

### METHODS

The American Spinal Muscular Atrophy Randomized Trials (AmSMART) group is an organization of 5 pediatric medical centers formed to perform clinical trials in children with SMA. The 3-year project was organized as follows: part 1, interrater reliability study; part 2, intrarater reliability study; and part 3, pilot drug trial.

The outcome measures were as follows: (1) pulmonary function tests in children older than 3 years, (2) quantitative muscle testing in children older than 3 years, (3) the Gross Motor Function Measure, and (4) the PedsQL. All equipment and methods have been described previously.

Patients were recruited from the Pediatric Neuromuscular Clinics and examined at Texas Scottish Rite Hospital for Children, Dallas; Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; Shriner's Hospital for Children, Portland, Ore; Children's Hospital, Richmond, Va; and Gillette Children's Specialty Healthcare, St Paul, Minn. Inclusion criteria were as follows: age 2 through 17 years, weakness and clinical diagnosis of SMA confirmed by mutation analysis of the survival motor neuron (SMN) gene, forced vital capacity greater than 20% of predicted for age (if the child was older than 5 years), less than 15% variance on test-retest with the use of quantitative muscle testing after instruction, and informed consent. Exclusion criteria were as follows: evidence of renal dysfunction, central nervous system damage, neurodegenerative or neuromuscular disease other than SMA; and mechanical ventilation of any type for more than 16 hours per day.

All testing was done in the evaluation room located in the physical therapy department. It contained all equipment used for an evaluation session including the examining table, the Richmond Quantitative Measurement System (a proprietary system available from George Masiello, Technical Director of Motion Analysis Laboratory, Children's Hospital, 2924 Brook Rd, Richmond, VA 23220; telephone: [804] 228-5824; e-mail: gmasiello@chva.org), items for use in the Gross Motor Function Measure, 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 items of the evaluation session were performed in the same order for all subjects. Positioning for pulmonary function tests and quantitative muscle testing was constant for all subjects. Rest periods during the evaluation session were as follows: 30 to 60 seconds between attempts, 5 minutes between pulmonary function tests and quantitative muscle testing, and 15 minutes between quantitative muscle testing and the Gross Motor Function Measure.

Pulmonary function tests were measured according to American Thoracic Society standards and included maximum inspiratory pressure (centimeters of water), maximum expiratory pressure (centimeters of water), cough pressure (peak cough flow, liters per minute), forced vital capacity (liters), and forced expiratory volume in the first second (liters). Lung volumes were measured with a spirometry system (KoKo; Pulmonary Data Services, Inc, Louisville, Colo) that automatically calculated percentage of predicted on the basis of arm span.

The Richmond Quantitative Measurement System was used for the quantitative muscle testing on the following muscle groups: (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 subject had 3 attempts for each muscle group, while the computer recorded the best of 3 on prompting from the evaluator. Strength was recorded in pounds. Subjects sat for muscle groups 1, 2, and 3 and were supine for group 4.

The Gross Motor Function Measure tool contained 88 items in 5 dimensions as follows: (A) lying and rolling, (B) sitting, (C) crawling and kneeling, (D) standing, and (E) walking, running, and jumping. Each subject continued through all domains according to his or her abilities. Each item was scored individually on a 0-to-3 ordinal scale and weighted equally. Although the Gross Motor Function Measure tool was designed for children with cerebral palsy, it has been validated in able-bodied children. Studies in 61 healthy children showed that a score of 86.7% could be expected by age 3 years and 99.9% after the age of 5 years, as cited in the manual provided with the Gross Motor Function Measure.

Several quality-of-life forms were administered by the study coordinator. The generic PedsQL included a parent questionnaire for each of 4 age groups: 2 through 4 years, 5 through 7 years, 8 through 12 years, and 13 through 17 years. In addition, there was a child questionnaire for each of 3 age groups: 5 through 7 years, 8 through 12 years, and 13 through 17 years. AmSMART child psychologists developed a single questionnaire, the PedsQL Neuromuscular Module for Parents of children aged 2 through 17 years, and 2 questionnaires for children aged 5 through 7 years and 8 through 17 years.

Visits were structured at follows: Visit 0 was a screening visit to determine that patients met diagnostic criteria and to obtain consent. Patients aged 2 through 4 years underwent evaluation at visits 1, 2, and 4. Patients aged 5 through 17 years performed pulmonary function tests, quantitative muscle testing, and the Gross Motor Function Measure at visits 1, 2, optional 3, and 4. Quality-of-life testing was administered at visits 0 and 4 for all subjects. To introduce the subject to the evaluation procedure, each child was seen twice within 2 weeks (visits 1 and 2). Thus, visit 1 served as a learning session for the subject and the evaluator. On the basis of data obtained from quantitative muscle testing during our intrarater reliability testing, we set the desirable variance at 15% or less. For patients aged 5 through 17 years, variance between visits 1 and 2 was required to be within 15% for at least 1 muscle group with quantitative muscle testing. If variance was greater than 15%, then the subject had a third chance to meet criteria (visit 3). Visits 1, 2, and optional 3 had to occur at least 48 hours apart but within 28 days. Visit 4 was carried out 4 weeks after visit 2 or 3. Protocol required that each subject be examined by the same evaluator for all visits. All subjects met criteria in 2 visits, except for one whose first evaluator was not available for visit 2. Therefore, this subject was seen by the same evaluator for visits 2, 3, and 4, while visit 1 was excluded from analysis. Of the possible 8 muscle groups, all patients qualified in at least 2 groups. With the use of 3 variance formulas (Table 1 and Figure), the muscle group most often qualified was the left elbow flexion (76%-86% of subjects), while the least often qualified was the left knee flexor (24%-29% of subjects). All supplemental data and information are posted on our Web site: http://acresearch.swmed.edu/AmSMART/.

©2003 American Medical Association. All rights reserved.
Table 1. Number of Patients for Each Muscle Group Meeting Inclusion Criteria for Variance by 3 Different Formulas*

<table>
<thead>
<tr>
<th>Muscle Group</th>
<th>Formula 1</th>
<th>Formula 2</th>
<th>Formula 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right grip</td>
<td>9 (43)</td>
<td>8 (38)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>Left grip</td>
<td>11 (52)</td>
<td>10 (48)</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Right knee flexor</td>
<td>13 (62)</td>
<td>11 (52)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Right knee extensor</td>
<td>11 (52)</td>
<td>9 (43)</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Left knee flexor</td>
<td>6 (29)</td>
<td>5 (24)</td>
<td>5 (24)</td>
</tr>
<tr>
<td>Left knee extensor</td>
<td>10 (48)</td>
<td>8 (38)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>Right elbow flexor</td>
<td>13 (62)</td>
<td>11 (52)</td>
<td>12 (57)</td>
</tr>
<tr>
<td>Left grip</td>
<td>18 (86)</td>
<td>16 (76)</td>
<td>17 (81)</td>
</tr>
<tr>
<td>Total (average)</td>
<td>91 (54)</td>
<td>78 (46)</td>
<td>85 (51)</td>
</tr>
</tbody>
</table>

*Formula 1 = (time 1 − time 2)/largest of 2. Formula 2 = (time 1 − time 2)/smallest of 2. Formula 3 = (time 1 − time 2)/average of 2.

Histogram showing the number of muscle groups meeting variance criteria. Means ± SDs are provided for each formula. Variance criteria were calculated by means of the following formulas: Formula 1 = (time 1 − time 2)/largest of 2, Formula 2 = (time 1 − time 2)/smallest of 2, Formula 3 = (time 1 − time 2)/average of 2.

Sessions for children aged 5 through 17 years took no more than 3 hours including standardized rest periods, while examination of the 2- through 4-year-olds took no more than 45 minutes. All procedures were approved by the University of Texas Southwestern Medical Center, Dallas, Institutional Review Board as well as the institutional review board of each participating center.

Paired samples statistics were performed for each of the 4 outcome measures. For pulmonary function tests, quantitative muscle testing, and Gross Motor Function Measure, we compared visits 2 and 4, except in one patient in whom we used visits 3 and 4. For quality of life, we compared visits 0 and 4. Reliability for each of the measures across 2 time points was determined by examining the results for both the intraclass correlation coefficient, a measure of reliability agreement, and the Bradley-Blackwood procedure, a test that simultaneously compares the means and variances of the 2 measurements.11 The pattern for good reliability with these 2 statistics is a significant intraclass correlation coefficient (indicating high agreement between measurements) and a nonsignificant Bradley-Blackwood procedure (the test-retest measurements have similar means and variances). If the Bradley-Blackwood procedure was significant, the paired t test and the Pitman test were used to evaluate where the bias between the 2 measurements occurred (equality of means and variances). Analyses were performed with SPSS version 11.0.1 (SPSS Inc, Chicago, Ill), and the significance level for reporting statistical results was set at $P<.05$.

RESULTs

Enrollment for part 2 began August 22, 2001, and was completed January 24, 2002. A total of 38 patients fulfilled initial inclusion criteria (in particular, homozygous deletion of SMN1), while 34 completed all visits. Of the 4 not included for data analysis, 1 was aged 2 years, 2 were aged 3 years, and 1 was aged 9 years. The 9-year-old had a forced vital capacity less than 20% of predicted (thereby not meeting all inclusion criteria), and the other 3 missed appointments, 2 because of illness.

Enrollment was fairly well distributed among the 5 centers and 7 evaluators, ranging from 5 to 9 patients each. Thirteen patients were aged 2 through 4 years and 21 were aged 5 through 17 years. For the 34 patients completing the study, 22 subjects were totally nonambulatory and 12 were ambulatory with or without orthoses.

The Gross Motor Function Measure was completed by 34 patients and showed excellent intrarater reliability both for raw scores and for percentage scores (Table 2). The intraclass correlation coefficients for each of the 5 domains ranged between 0.96 and 0.98 for both raw and percentage scores, with all 2-tailed $P$ values less than .001, indicating a very high reliability association between the measures. All Bradley-Blackwood procedures comparing the measurements at the 2 time points were nonsignificant. Nonsignificant Bradley-Blackwood results indicated that there was no bias between measures (the measures were equivalent) at 2 time points.

Results of the Gross Motor Function Measure were found to be associated with patients' motor ability. Thirty-four completed dimensions A (lying and rolling) and B (sitting). Twenty-two finished dimension C (crawling and kneeling); 15, dimension D (standing); and 13, dimension E (walking, running, and jumping). Hence, every subject was able to perform at least 2 dimensions and achieve a score with this tool.

Nine measures were included in pulmonary function tests (Table 3). Six were measured by means of both absolute values and percentages of predicted values for age as scores. Twenty patients completed all measures, with 1 patient being unable to complete maximum vol-
The quality-of-life tool contained several components, each specific for the age of the patient, except for the results of the Bradley-Blackwood procedure (P<.05 on F test).

To examine these measures further, total scores and subscale scores for muscle groups were added to create 5 new scores: grip plus elbow flexors for upper extremity, knee flexors plus knee extensors for lower extremity, right grip plus right elbow flexor plus right knee flexor plus right knee extensor for right side, left grip plus left elbow flexor plus left knee flexor plus left knee extensor for left side, and total or global muscle strength. All 5 of these combined muscle strength scores performed well, with all intraclass correlation coefficients being between 0.99 and 1.00 and all Bradley-Blackwood procedures being nonsignificant (P<.05) except for lower extremity, which was found to have biased means and variances (both higher at visit 2/3).

The quality-of-life tool contained several components, each specific for the age of the patient, except for
Spinal muscular atrophy is caused by homozygous deletion in the SMN1 gene on chromosome 5q. The disease severity and whether a child has type 1, 2, or 3 SMA correlate directly with the amount of SMN protein that can be measured in tissues or cells. New strategies for possible therapy for SMA focus on increasing the production of SMN protein by SMN2, a disease-modifying gene. Recent work in mouse models shows that administration of certain compounds, such as sodium butyrate, can increase protein levels in mouse lymphocytes and improve motor function in the SMA mouse. Similar studies with as yet to be determined compounds in human tissues may be forthcoming, making phase 3 clinical trials for SMA imminent.

However, clinical trials require reliable and sensitive outcome measures. Developing outcome measures that are relevant to the disease process has focused on measures of muscle strength, since direct measures of motor neuron function, which was used for all ages. A parent or guardian completed a questionnaire for each patient, while, in addition, each patient completed his or her own, beginning at age 5 years. On the parent PedsQL, 11 (of 13 in the age group) completed the questionnaire for ages 2 through 4 years, 5 (of 7) for ages 5 through 7 years, 8 (of 9) for ages 8 through 12 years, and 4 (of 5) for ages 13 through 17 years. On the child questionnaire, 5 (of 7) completed the questionnaire for ages 2 through 4 years, 5 (of 7) for ages 5 through 7 years, 8 (of 9) for ages 8 through 12 years, and 4 (of 5) for ages 13 through 17 years. Enough patients completed the PedsQL Neuromuscular Module for Parents, to allow calculation of reliability. For this module, data from 33 parents were available for analysis. All 3 dimensions of this module and the total score had high reliability, with all intraclass correlation coefficients being between 0.73 and 0.84, and all Bradley-Blackwood procedures were nonsignificant (smallest \( P = .29 \)). Since no more than 14 patients or parents completed any one of the other quality of life components, reliability analyses were not performed for these forms.

### Table 4. Descriptive Statistics and Reliability Results for Quantitative Muscle Testing

<table>
<thead>
<tr>
<th>Strength Measurement, lb</th>
<th>No. of Patients</th>
<th>Visit 2, Mean (SD)</th>
<th>Visit 4, Mean (SD)</th>
<th>Paired Differences, Mean (SD)</th>
<th>ICC†</th>
<th>Paired t Test</th>
<th>Bradley-Blackwood Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F dfn/dfd</td>
<td>Value P dfn/dfd Value P</td>
</tr>
<tr>
<td>Right grip</td>
<td>21</td>
<td>8.28 (9.76)</td>
<td>8.98 (10.86)</td>
<td>−0.70 (1.71)</td>
<td>0.99</td>
<td>9.65</td>
<td>2/19 .001 .19 .07 .36 .002</td>
</tr>
<tr>
<td>Left grip</td>
<td>21</td>
<td>7.47 (3.34)</td>
<td>7.88 (9.84)</td>
<td>−0.41 (1.38)</td>
<td>0.99</td>
<td>2.45</td>
<td>2/19 .11 −1.38 .18 −1.68 .11</td>
</tr>
<tr>
<td>Right knee flexion</td>
<td>21</td>
<td>4.73 (4.94)</td>
<td>4.55 (4.42)</td>
<td>0.18 (1.12)</td>
<td>0.97</td>
<td>3.07</td>
<td>2/19 .07 0.75 .46 2.34 .03</td>
</tr>
<tr>
<td>Right knee extension</td>
<td>21</td>
<td>3.81 (4.04)</td>
<td>3.34 (3.67)</td>
<td>0.47 (1.40)</td>
<td>0.93</td>
<td>1.95</td>
<td>2/19 .17 1.55 .14 1.20 .24</td>
</tr>
<tr>
<td>Left knee flexion</td>
<td>21</td>
<td>4.06 (4.19)</td>
<td>4.11 (4.25)</td>
<td>−0.05 (0.97)</td>
<td>0.97</td>
<td>0.06</td>
<td>2/19 .94 −0.26 .80 −0.24 .81</td>
</tr>
<tr>
<td>Left knee extension</td>
<td>21</td>
<td>3.95 (5.28)</td>
<td>3.36 (4.48)</td>
<td>0.59 (1.35)</td>
<td>0.96</td>
<td>8.32</td>
<td>2/19 .003 2.02 .06 3.27 .004</td>
</tr>
<tr>
<td>Right elbow flexion</td>
<td>21</td>
<td>8.67 (8.91)</td>
<td>8.78 (9.45)</td>
<td>−0.11 (2.28)</td>
<td>0.97</td>
<td>0.60</td>
<td>2/19 .56 −0.22 .83 −1.07 .30</td>
</tr>
<tr>
<td>Left elbow flexion</td>
<td>21</td>
<td>8.97 (9.86)</td>
<td>8.44 (9.28)</td>
<td>0.53 (1.15)</td>
<td>0.99</td>
<td>6.22</td>
<td>2/19 .008 2.10 .06 2.60 .02</td>
</tr>
<tr>
<td>Combined scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower extremities</td>
<td>21</td>
<td>16.55 (16.70)</td>
<td>15.35 (15.50)</td>
<td>1.20 (2.63)</td>
<td>0.99</td>
<td>5.10</td>
<td>2/19 .02 2.09 .05 2.23 .04</td>
</tr>
<tr>
<td>Upper extremities</td>
<td>21</td>
<td>33.39 (37.32)</td>
<td>34.09 (38.70)</td>
<td>−0.70 (4.47)</td>
<td>0.99</td>
<td>1.27</td>
<td>2/19 .30 −0.72 .48 −1.42 .17</td>
</tr>
<tr>
<td>Right side</td>
<td>21</td>
<td>25.49 (25.36)</td>
<td>25.65 (26.40)</td>
<td>−0.16 (3.58)</td>
<td>0.99</td>
<td>0.90</td>
<td>2/19 .42 −0.20 .84 −1.33 .20</td>
</tr>
<tr>
<td>Left side</td>
<td>21</td>
<td>24.45 (26.42)</td>
<td>23.79 (26.00)</td>
<td>0.65 (2.03)</td>
<td>1.00</td>
<td>1.50</td>
<td>2/19 .25 1.47 .16 0.92 .37</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>49.94 (51.60)</td>
<td>49.44 (52.32)</td>
<td>0.50 (5.03)</td>
<td>1.00</td>
<td>0.30</td>
<td>2/19 .75 0.45 .66 −0.63 .54</td>
</tr>
</tbody>
</table>

Abbreviations: dfn, degrees of freedom for numerator; dfd, degrees of freedom for denominator; ICC, intraclass correlation coefficient.

*All ICCs had \( P < .001 \).

### Table 5. Descriptive Statistics and Reliability Results for PedsQL Neuromuscular Module for Parents

<table>
<thead>
<tr>
<th>Dimension</th>
<th>No. of Patients</th>
<th>Screening Visit, Mean (SD), Score</th>
<th>Visit 4, Mean (SD), Score</th>
<th>Paired Differences,* Mean (SD)</th>
<th>ICC†</th>
<th>Paired t Test</th>
<th>Bradley-Blackwood Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F dfn/dfd</td>
<td>Value P dfn/dfd Value P</td>
</tr>
<tr>
<td>Disease</td>
<td>33</td>
<td>61.06 (19.77)</td>
<td>61.70 (15.22)</td>
<td>−0.64 (9.15)</td>
<td>0.84</td>
<td>0.58</td>
<td>2/31 .57</td>
</tr>
<tr>
<td>Communication</td>
<td>32</td>
<td>63.22 (32.62)</td>
<td>61.47 (28.94)</td>
<td>0.75 (18.90)</td>
<td>0.81</td>
<td>0.78</td>
<td>2/30 .47</td>
</tr>
<tr>
<td>Family resources</td>
<td>33</td>
<td>60.19 (23.09)</td>
<td>55.45 (22.41)</td>
<td>4.73 (16.85)</td>
<td>0.73</td>
<td>1.29</td>
<td>2/31 .29</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>61.42 (19.76)</td>
<td>59.91 (18.01)</td>
<td>1.52 (11.45)</td>
<td>0.82</td>
<td>0.69</td>
<td>2/31 .51</td>
</tr>
</tbody>
</table>

Abbreviations: dfn, degrees of freedom for numerator; dfd, degrees of freedom for denominator; ICC, intraclass correlation coefficient.

*All t tests had \( P > .05 \).

†All ICCs had \( P < .001 \).
function such as electrodiagnostic studies require considerable skill on the part of the evaluator and discomfort on the part of the subject. Such difficulties are compounded by the fact that the population of interest is very young. This is why we have tried to use outcome measures that are age appropriate while being disease relevant.

Pulmonary function tests seem to be particularly relevant to SMA, since nearly all morbidity and mortality are caused by respiratory infection or failure. Pulmonary function tests have been used rarely as outcome measures in pediatric cohorts, but have been successfully used in multicenter clinical trials in asthma. Although the first data set for pulmonary function tests was disappointing, we have shown very good reliability in the current study. This is largely attributed to a special training program for the evaluators. This 2-day program was modeled and carried out by pulmonologists previously involved in asthma clinical trials. All evaluators indicated their confidence in performing pulmonary function tests after that session. Moreover, each visit thereafter was monitored by one of the respiratory therapists before acceptance for data entry. Such intensive and specialized training is a requirement for reliable and valid multicenter trials.

The Gross Motor Function Measure was developed as an outcome measure for children with cerebral palsy and has been used with good reliability in several clinical trials in that population. This tool has not previously been used in any neuromuscular disease of childhood, to our knowledge. Our preliminary study showed that the Gross Motor Function Measure was reliable, but the present cohort provides more evidence of its reliability in SMA. Previous attempts to measure motor function in young patients with SMA, such as those by the Dallas-Cincinnati-Newington group, were disappointing, showing great variability. Another example of such an attempt is the Hammersmith motor ability scale, which was validated in patients with Duchenne muscular dystrophy. Only the Dallas-Cincinnati-Newington group has used a functional motor scale to observe patients with SMA prospectively. It may be very interesting to determine whether the Gross Motor Function Measure changes over time in a cohort of patients whose mutation analysis and SMN2 copy number were known, thereby providing some phenotype-genotype correlation with motor function. This could be the focus of another study.

Quantitative muscle testing has been used extensively in adult clinical trials of patients with neuromuscular disease, particularly amyotrophic lateral sclerosis and recently Duchenne muscular dystrophy. Most such trials required that the subjects have strength of at least 3+ in one or more muscle groups by means of the manual muscle test. Such inclusion criteria would eliminate as many as 90% of children with SMA from clinical trials. Thus, we sought a quantitative muscle testing system developed specifically for small children with limited strength. The Richmond Quantitative Measurement System was designed with pediatric patients in mind, using a sensitive transducer that can measure as little as 1 lb of force, meaning that even a muscle group that measures 2 on the manual muscle test could register force with the Richmond Quan-

titative Measurement System. However, this system still shows greater variability than we desire in the weakest muscle groups, such as the lower extremities. It is possible that a more sensitive transducer than currently used could improve variability.

Our quality-of-life questionnaire is a work in progress. The PedsQL has established reliability and validity, although it is a generic tool. We added questions that were specific to the lifestyle of a child with SMA, and our results indicated that this tool (Neuromuscular Module for Parents) is valid for the SMA population.

In summary, we have shown reliability and validity for the Gross Motor Function Measure, pulmonary function tests, and quantitative muscle testing in SMA. We still lack outcome measures for children younger than 2 years, a group in which only mortality has been available as an outcome measure. Recently, Bromberg and Swoada used motor unit number estimate in infants and toddlers with SMA and showed that the motor neuron count correlated with the severity of disease. Such a tool could be incorporated in a clinical trial if treatment were to be given very early in the disease process, before motor unit counts reached their lowest level. On the other hand, we know from the work of this group that motor unit counts seem to remain stable up to 18 months. Therefore, a clinical trial in patients with stable SMA with the use of motor unit number estimate might depend on showing an increase in motor units or increased compound muscle action potential with treatment.

Accepted for publication March 7, 2003.

Author contributions: Study concept and design (Drs Iannaccone, Hynan, Morton, Reisch, Schochet, Luckett, Wong, Samaha, Russman, and Leshner; Mss Carman, Gordon, and Stout; Mr Webster); acquisition of data (Drs Iannaccone, Wong, Russman, Leshner, and Smith and Mss Raab, Carman, Gordon, Harris, Smith, Goldsmith, Walters, Fritch, Morehart, Zilke, Sienko-Thomas, Mayhew, Grillo, and McCarty); analysis and interpretation of data (Drs Iannaccone, Hynan, Schochet, and Wong); drafting of the manuscript (Drs Iannaccone, Hynan, Samaha, and Smith and Mss Harris, Smith, Walters, Zilke, May-
hew, Grillo, Stout, and McCarty); critical revision of the manuscript for important intellectual content (Drs Iannaccone, Hynan, Morton, Reisch, Schochet, Luckett, Wong, Russman, and Leschner; Ms Raab, Carman, Gordon, Smith, Goldsmith, Fritch, Morehart, and Sienko-Thomas; and Mr Webster); statistical expertise (Dr Hynan); obtained funding (Drs Iannaccone and Samaha); administrative, technical, and material support (Drs Iannaccone, Hynan, Morton, Reisch, Schochet, Samaha, Leschner, and Smith; Ms Carman, Gordon, Harris, Smith, Goldsmith, Walters, Morehart, Zilke, Sienko-Thomas, Mayhew, and Grillo; and Mr Webster); study supervision (Drs Iannaccone, Morton, Reisch, and Russman and Ms Raab and Smith).

This study was supported by grant 1-RO1-NS 39327-02 from the National Institutes of Health, Bethesda, Md, and the Muscular Dystrophy Association, Tucson, Ariz.

This study was presented in part at the 7th International Congress of the World Muscle Society; October 2-5, 2002; Rotterdam, the Netherlands.

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

REFERENCES