Reliability and Responsiveness of the Gross Motor Function Measure-88 in Children With Cerebral Palsy

Jooyeon Ko, MinYoung Kim

Background. The Gross Motor Function Measure (GMFM-88) is commonly used in the evaluation of gross motor function in children with cerebral palsy (CP). The relative reliability of GMFM-88 has been assessed in children with CP. However, little information is available regarding the absolute reliability or responsiveness of GMFM-88.

Objective. The purpose of this study was to determine the absolute and relative reliability and the responsiveness of the GMFM-88 in evaluating gross motor function in children with CP.

Design. A clinical measurement design was used.

Methods. Ten raters scored the GMFM-88 in 84 children (mean age=3.7 years, SD=1.9, range=10 months to 9 years 9 months) from video records across all Gross Motor Function Classification System (GMFCS) levels to establish interrater reliability. Two raters participated to assess intrarater reliability. Responsiveness was determined from 3 additional assessments after the baseline assessment. The interrater and intrarater intraclass correlation coefficients (ICCs) with 95% confidence intervals, standard error of measurement (SEM), smallest real difference (SRD), effect size (ES), and standardized response mean (SRM) were calculated.

Results. The relative reliability of the GMFM was excellent (ICCs=.952–1.000). The SEM and SRD for total score of the GMFM were acceptable (1.60 and 3.14, respectively). Additionally, the ES and SRM of the dimension goal scores increased gradually in the 3 follow-up assessments (GMFCS levels I and II: ES=0.5, 0.6, and 0.8 and SRM=1.3, 1.8, and 2.0; GMFCS levels III–V: ES=0.4, 0.7, and 0.9 and SRM=1.5, 1.7, and 2.0).

Limitations. Children over 10 years of age with CP were not included in this study, so the results should not be generalized to all children with CP.

Conclusions. Both the reliability and the responsiveness of the GMFM-88 are reasonable for measuring gross motor function in children with CP.

J. Ko, PT, PhD, Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, Seongnam-si, Gyeonggi-do, Republic of Korea.

M. Kim, MD, PhD, Department of Rehabilitation Medicine, CHA Bundang Medical Center, 59 Yatap-ro, Bungdang-gu, CHA University, Seongnam-si, Gyeonggi-do, 463-712, Republic of Korea. Address all correspondence to Dr Kim at: kmin@cha.ac.kr.

[Ko J, Kim M. Reliability and responsiveness of the Gross Motor Function Measure-88 in children with cerebral palsy. *Phys Ther.* 2013;93:393–400.]

Published Ahead of Print: November 8, 2012
Accepted: November 2, 2012
Submitted: November 4, 2011

Cerebral palsy (CP) is the most common disorder resulting in physical disability in children. The extent of motor disability varies by the degree of brain damage, ranging from minor dysfunction to the most severely impaired cases.[1] As impairments in these children involve a variable range of functions, the children need comprehensive rehabilitation therapy.[2] Nonetheless, physical therapy is the most important modality because children's disabilities primarily entail motor dysfunction. It is important to determine the effects of therapeutic interventions on motor function with reliable and valid tests. Several evaluation measures are available to assess gross motor development in children with CP, such as the Pediatric Evaluation of Disability Inventory[3] and the Functional Independence Measure for Children.[4] Among them, to date, the Gross Motor Function Measure (GMFM)[5] is the best-known and most frequently used instrument around the world. The test requires a qualified therapist and a standardized environment.

The original version of the GMFM, the GMFM-88, consists of 88 items that have been categorized into 5 dimensions of gross motor function: lying and rolling; sitting; crawling and kneeling; standing; and walking, running, and jumping.[6] Because it allows quantitative evaluation of motor function, many studies have used the GMFM to assess the effectiveness of interventions in children with CP.[1,7–11] The psychometric properties of the GMFM-88 also have been evaluated through many studies since its development.[5,6] To our knowledge, however, previous studies provided only relative reliability,[1,5] and only a few studies have reported on the validity of the GMFM.[1,12]

Reliability refers to dependability, consistency, and stability of scores on an assessment tool. Reliability includes relative or absolute reliability[13] and interrater or intrarater reliability. The interrater reliability of a test is an estimate of how consistent the test is when used by different raters; the intrarater reliability is an estimate of the consistency of the score assigned to a single set of responses on 2 or more occasions by the same rater. Relative reliability examines the relationship between 2 or more measurements and the consistency of an individual's position within the group. The intraclass correlation coefficient (ICC) is used to quantify relative reliability.[14] Although the ICC is the most widely used method for reliability studies, it provides limited information on measurement error.[15] A high ICC does not necessarily indicate a low possibility of measurement error. Absolute reliability investigates the extent of measurement error, and the standard error of measurement (SEM) and smallest real difference (SRD) have usually been used to quantify it.[16] To date, there has been no study on absolute reliability for the GMFM-88. Thus, it is important to obtain SEM and SRD values for the GMFM-88 in children with CP.

Validity in such measures includes *responsiveness*, which is defined as the ability to differentiate clinically important differences.[17] For the longitudinal assessments, there should be appropriate power to detect significant changes. The effect size (ES)[18] and the standardized response mean (SRM)[19] have commonly been used to measure responsiveness. Although the more recently developed GMFM-66 is now available, Russell et al,[5] the developers of the GMFM, recommended the use of the GMFM-88 rather than the GMFM-66 when evaluating changes in gross motor function at the 2 extremes of impairment (ie, severely limited and mildly involved cases). In a study by Heyrman et al,[20] evaluating psychometric properties of the Trunk Control Measurement Scale for children with CP, the GMFM-88 was used instead of the GMFM-66 because the shorter version provided only a total score and not scores of the different dimensions or the goal total score. The goal total score is the mean score of selected dimensions on which the physical therapist decides to focus. Thus, the GMFM-88 needs to be further examined in terms of its responsiveness in evaluating children with CP, including those at the extremes of the range of functional ability and in terms of goal total scores.

The purpose of this study was to provide valuable information for clinical uses and research purposes by fully estimating the relative and absolute reliability and the responsiveness of the GMFM-88 for children with CP.

Method

Sample

We recruited 84 children with CP who had been admitted for 1 month of intensive rehabilitation therapy (2 sessions of physical therapy and 2 sessions of occupational therapy per day, 5 days per week) to CHA Bundang Medical Center, South Korea, between June and November 2010. The mean age at the time of admission (the baseline assessment) was 3.7 years (mean age=3.7 years, SD=1.9, range=10 months to 9 years 9 months; Tab. 1). The inclusion criteria were: (1) a confirmed diagnosis of CP from a pediatric rehabilitation medicine physician, (2) no surgical procedures within the previous 6 months, and (3) no injection of botulinum toxin type A within the previous 6 months. All caregivers of the children were informed of the procedure and the purposes of this study, and all signed informed consent forms.

Table 1.
General Characteristics of the Samples for the Intrarater and Interrater Reliability Analysis and Responsiveness Analysis[a]

Variable	Intrarater and Interrater Reliability Analysis			Responsiveness Analysis (n=60)
	Total (N=84)	<4 y (n=52)	≥4 y (n=32)	
Age, $\bar{X}$ (SD)	3.7 (1.9)	2.5 (0.9)	5.5 (1.5)	3.9 (2.0)
Sex				
Male	54 (64.3)	32 (61.5)	22 (68.8)	38 (63.3)
Female	30 (35.7)	20 (38.5)	10 (31.2)	22 (36.7)
GMFCS level				
I	14 (16.7)	9 (17.3)	5 (15.6)	7 (11.7)
II	9 (10.7)	6 (11.5)	3 (9.4)	9 (15.0)
III	22 (26.2)	19 (36.5)	3 (9.4)	17 (28.3)
IV	19 (22.6)	8 (15.4)	11 (34.4)	15 (25.0)
V	20 (23.8)	10 (19.2)	10 (31.3)	12 (20.0)

[a] Values are frequencies (%), except for age. GMFCS=Gross Motor Function Classification System.

Functional ability was classified using the Gross Motor Function Classification System (GMFCS), from minimally involved (level I) to most severely involved (level V).[21] The enrolled participants were classified as GMFCS levels I (n=14), II (n=9), III (n=22), IV (n=19), and V (n=20) at the baseline assessment (Tab. 1). All participants were divided into 2 age subgroups: 52 were younger than 4 years of age (mean age=2.5 years, SD=0.9; 32 boys and 20 girls), and 32 were 4 years of age or older (mean age=5.5 years, SD=1.5; 22 boys and 10 girls). The results of a study by Rosenbaum et al[22] indicated most rapid changes in GMFM results occur during the first 4 years of life.

Instrument

A translated Korean version of the GMFM-88 was used.

Procedure

Tests for all items of the GMFM-88 were administered in a pediatric physical therapy room that was comfortable and familiar to the children. The tests were conducted in the order given in the GMFM manual, and all procedures were videotaped by 2 assistant therapists. The GMFM baseline assessment of reliability was performed at the time of admission. During the tests, the children were barefoot and used no assistive devices. It took 40 to 60 minutes to record each child's body movements. To determine interrater reliability, 10 pediatric physical therapists (raters A through J) served as the GMFM raters. The raters had a mean of 5 years and 3 months of experience (range=8 months to 10 years 3 months) in pediatric practice. All of the pediatric physical therapists attended a 30-hour GMFM workshop aimed at training the therapists to administer the instrument. After the workshop, the 10 raters independently viewed and scored the 84 GMFM records without dis-

The Bottom Line

What do we already know about this topic?

There is little information about the absolute reliability and responsiveness of the Gross Motor Function Measure (GMFM-88).

What new information does this study offer?

The GMFM-88 shows satisfactory reliability for both absolute and relative indexes. In addition, the GMFM-88 is responsive to functional improvement when administered to a child with cerebral palsy in a clinical setting.

If you're a caregiver, what might these findings mean for you?

The GMFM-88 is a good measure to better understand the gross motor ability and functional change of children with cerebral palsy.

Table 2.
Scores on the Gross Motor Function Measure (GMFM-88) and Values of Interrater Reliability Tests Assessed by 10 Raters (N=84)[a]

Dimension	$\bar{X}$ (SD)	Median	ICC (2,1)	95% CI	SEM	SRD
Lying and rolling	73.3 (29.0)	86.0	.975	.966–.982	4.59	8.98
Sitting	53.4 (34.4)	55.0	.986	.982–.990	4.07	7.98
Crawling and kneeling	41.5 (38.4)	39.0	.989	.985–.992	4.03	7.90
Standing	23.1 (32.4)	3.0	.987	.982–.991	3.69	7.23
Walking, running, and jumping	16.3 (27.4)	0.0	.994	.992–.996	2.12	4.16
Total	41.5 (29.2)	37.0	.997	.996–.998	1.60	3.14

[a] ICC (2,1)=intraclass correlation coefficient for interrater reliability, 95% CI=95% confidence interval, SEM=standard error of measurement, SRD=smallest real difference.

cussing the scores with one another. To determine intrarater reliability, 2 of the 10 raters rescored the same 84 video recordings 1 month after the first assessment. The assessment interval was used to minimize any influence of memory.[23] One of the 2 raters, rater A (the most experienced rater) had more than 10 years of experience in the pediatric field; the other rater, rater J (newly trained), had less than 1 year of experience. To evaluate the responsiveness of the GMFM-88 in the same children, 3 additional follow-up assessments were conducted at 1, 3, and 6 months after the baseline assessment by the therapists who participated in the reliability study, and the same therapist rated each child at the 3 time periods in which no video recording was used.

Among the 84 children, data from 60 children (mean age=3.9 years, SD=2.0, range=1 year 1 month to 9 years 5 months; Tab. 1) who continued rehabilitation therapy after discharge were available for analysis of responsiveness.

In accordance with the original version of GMFM-88, the raters scored all 88 items for each assessment. The raw score of each dimension was converted into a percentage score.[24] The values obtained from the process described above were used for data analysis.

Data Analysis

Descriptive statistics for each dimension and total percentage scores were used. For relative reliability, ICC (2,1) and ICC (1,1) with 95% confidence intervals were used to evaluate the interrater and intrarater reliability of each GMFM dimensions and of the total scores. According to Polit and Hungler,[25] a reliability coefficient of at least .70 may be sufficient for group comparisons. However, when a measure is to be used for clinical decision making regarding an individual patient, a more stringent criterion is recommended. In such cases, an ICC of .90 for a total score is generally accepted as the minimum.[26]

Absolute reliability was determined by calculating both the SEM and the SRD. The SEM is the standard deviation of measurement errors,[27] and the SRD is the score that might be accepted as a smallest extent of real change.[28] The calculation formula for the parameters and their significance was: $SEM = SD\sqrt{1-r}$.[29] We used the ICC to determine the value of r in this study. The SRD value of less than $1.96 \times SEM$ is required in order to include 95% of the ratings when applied to interrater reliability.[14] In the case of intrarater reliability, SRD should be less than $\sqrt{2} \times 1.96 \times SEM = 2.77 \times SEM$ for 95% of the pairs of ratings to have meaning.[14] The interrater SEM and SRD were calculated using all ratings recorded during the 10-rater interrater trial, and the intrarater SEM and SRD were calculated from repeated scoring by raters A and J. Additionally, the interrater and intrarater reliability of GMFM-88 were assessed for each of the different age groups and GMFCS levels (<4 years and ≥4 years of age; GMFCS levels I and II and levels III–V).

The responsiveness was determined by use of the ES and the SRM for each dimension, total score, and goal total score of the GMFM-88. The goal dimensions in the GMFM-88 were individually selected by the raters. To identify the goal areas, the child's current functional status (ie, information from his or her GMFM assessment) and a number of other factors, including the child's age and his or her home and school environment, were considered.[5] For example, if a child was similar to a 6-year-old in functioning at GMFCS level IV and attended a kindergarten, the therapist might choose sitting and crawling and kneeling as goal areas. Thus, the mean goal total score would be the average of percentage scores for those 2 dimensions. Each goal total score was calculated as a mean score of the selected dimensions. The ES refers to the mean change between the baseline and the later score divided by the standard deviation of the baseline. The SRM was calcu-

Table 3.

Interrater and Intrarater Reliability of the Gross Motor Function Measure (GMFM-88) Total Score for Different Age and Functional Subgroups of Children With Cerebral Palsy (N=84)[a]

Subgroup	n	Interrater (Raters A–J)				Intrarater (Rater A)[b]		
		$\bar{X}$ (SD)	ICC (2,1)	SEM	SRD	ICC (1,1)	SEM	SRD
Age								
<4 y	52	39.9 (28.1)	.997	1.54	3.02	1.000	0.00	
≥4 y	32	44.1 (30.7)	.987	3.50	6.86	1.000	0.00	
GMFCS								
I	14	79.9 (23.7)	.952	5.19	10.2	1.000	0.00	
II	9	64.5 (20.6)	.994	1.60	3.13	1.000	0.00	
III	22	46.3 (17.1)	.990	1.71	3.35	.998	0.76	2.11
IV	19	30.6 (13.7)	.983	1.79	3.51	.999	0.43	1.19
V	20	9.5 (7.8)	.972	1.31	2.57	1.000	0.00	

[a] GMFCS=Gross Motor Function Classification System, ICC (2,1)=intraclass correlation coefficient for interrater reliability, ICC (1,1)=intraclass correlation coefficient for intrarater reliability, SE=standard error of measurement, SRD=smallest real difference.
[b] Rater A was the most experienced therapist.

Table 4.

Effect Size (ES) and Standardized Response Mean (SRM) for the Gross Motor Function Measure (GMFM-88) According to the GMFCS[a] Levels I and II[b,c] and III to V[d] (n=60)

Measure	Baseline–1 Month			Baseline–3 Months			Baseline–6 Months		
	Mean Change (SD)	ES	SRM	Mean Change (SD)	ES	SRM	Mean Change (SD)	ES	SRM
GMFCS I–II (n=16)									
Lying and rolling	1.0 (2.1)	0.4	0.5	1.3 (2.8)	0.5	0.5	1.6 (2.8)	0.6	0.6
Sitting	2.4 (3.2)	0.4	0.8	3.3 (3.9)	0.6	0.8	3.9 (3.9)	0.7	1.0
Crawling and kneeling	4.0 (4.6)	0.4	0.9	6.2 (5.7)	0.6	1.1	6.5 (6.1)	0.6	1.1
Standing	7.1 (6.0)	0.3	1.2	10.3 (8.0)	0.4	1.3	12.7 (11.4)	0.5	1.1
Walking, running, and jumping	6.3 (6.9)	0.2	1.0	7.6 (7.4)	0.3	1.0	10.8 (10.3)	0.4	1.0
Goal total	7.3 (5.6)	0.5	1.3	9.2 (5.0)	0.6	1.8	11.8 (6.0)	0.8	2.0
Total	4.1 (3.2)	0.3	1.3	5.7 (3.6)	0.5	1.6	7.1 (4.6)	0.6	1.6
GMFCS III–V (n=44)									
Lying and rolling	6.2 (5.9)	0.3	1.1	9.3 (8.3)	0.4	1.1	12.1 (9.3)	0.5	1.3
Sitting	6.9 (7.0)	0.3	1.0	11.8 (8.9)	0.4	1.3	16.4 (12.5)	0.6	1.3
Crawling and kneeling	4.0 (5.2)	0.1	0.8	10.2 (11.9)	0.4	0.9	12.6 (14.1)	0.4	0.9
Standing	4.8 (8.3)	0.5	0.6	6.8 (11.6)	0.7	0.6	7.1 (10.9)	0.7	0.7
Walking, running, and jumping	0.7 (1.7)	0.1	0.4	1.7 (3.0)	0.3	0.6	2.1 (3.7)	0.4	0.6
Goal total	7.4 (5.1)	0.4	1.5	12.7 (7.4)	0.7	1.7	16.4 (8.4)	0.9	2.0
Total	4.5 (3.2)	0.3	1.4	8.3 (6.3)	0.5	1.3	10.1 (6.8)	0.6	1.5

[a] GMFCS=Gross Motor Function Classification System.
[b] One child was not tested at 3 and 6 months after admission due to health problems of the child.
[c] One child was not tested at 6 months after admission due to parental reasons.
[d] One child was not tested at 6 months after admission due to health problems of the child.

lated by dividing the mean change by the standard deviation of the changes. There is no absolute standard for interpreting ES and SRM values, so we used the interpretation of Cohen.[30] According to those criteria, an ES greater than 0.8 is large, 0.5 to 0.8 is moderate, and 0.2 to 0.5 is small. It could be interpreted that the larger ES becomes, the greater the change. To analyze responsiveness, we classified participants by age and functional impairment.

All statistical analyses were performed using SPSS version 19.0 software (SPSS Inc, Chicago, Illinois).

Results

Relative and absolute reliability for intrarater reliability were satisfactory (ICC [1,1]=.994–1.000, SEM=0.00–2.23, and SRD=2.28–6.18, respectively). Indexes of relative interrater reliability, ICC (2,1) for each dimension, and total score on the GMFM-88 ranged from .975 to .997 (Tab. 2). The interrater and intrarater ICCs when children were classified by the 2 age subgroups and 5 GMFCS levels were >.952 for all examinations (Tab. 3). All of the ICC values were excellent.

For absolute reliability, intrarater SEMs and SRDs ranged from 0.00 to 2.23 and from 2.28 to 6.18, respectively. The indexes of interrater reliability SEM and SRD were in the ranges of 1.60 to 4.59 and 3.14 to 8.98, respectively, for the 5 dimensions and total scores (Tab. 2). The SEM and SRD values according to age and GMFCS level subgroups are shown in Table 3. For responsiveness, the ES and the SRM of each dimension, goal total, and total score of the GMFM-88 for GMFCS levels I and II and levels III to V were found to be responsive, showing moderate to large ES and SRM value at 6 months after the baseline assessment (Tab. 4).

Discussion

Our results, based on ICC, SEM, SRD, ES, and SRM indexes, showed satisfactory levels of responsiveness and of relative and absolute reliability for the GMFM-88 for children with CP, across all GMFCS functional levels.

In terms of relative reliability, the ICCs for interrater and intrarater reliability for the 5 dimensions and total scores were >.97 and >.94, respectively. In a previous study by Russell et al,[5] the original GMFM developers, the values of interrater ICCs ranged from .87 to .99, and those of intrarater ICCs ranged from .92 to .99 for all dimensions and total scores. In the present study, the years of experience of the pediatric therapists ranged from less than 1 year to more than 10 years. Despite 30 hours of workshop training, the results for the least experienced therapist were poorer than those for the more experienced therapists. Thus, an adequate amount of training is required before conducting GMFM-88 assessments, and retraining may be necessary to minimize differences among therapists.

Nordmark et al[31] examined the interrater and intrarater reliability of the GMFM-88 for the dimensions of lying and rolling, sitting, and crawling and kneeling by repeated administration using a video recording of 3 children with CP by 15 physical therapists using the GMFM manual without previous experience or training in using the measure. The ICC (2,1) was .77 and .88 at the first and second assessments, respectively, and the ICC (1,1) was .68 at the second assessment. When compared with previous studies, our results were similar or showed even higher reliability (Tabs. 1 and 2). Because demographic data, including age and severity of motor involvement, might show different clinical manifestations, we investigated interrater and intrarater reliability according to age and functional level by classifying participants into subgroups. All results from the subgroups demonstrated high ICC values (Tab. 3). However, it became clear that the use of ICC alone would be insufficient for a reliability analysis because the ICC alone does not enable measuring actual differences.[14,16]

The SEM is an estimate of error associated with measurement. Currently, there is no criterion for the interpretation of SEM levels for the GMFM-88. Liaw et al[32] suggested that SEM scores less than 10% of the total mean score were acceptable. Applying this criterion to our results for interrater reliability, the SEMs of the lying and rolling, sitting, and crawling and kneeling dimensions were acceptable; however, the SEM values for standing and for walking, running, and jumping were slightly higher than 10% of the mean score (3.69 >10% of 23.1 and 2.12 >10% of 16.3, respectively). The SEMs of the total gross motor score for GMFCS levels III and IV for the intrarater reliability were reasonable (0.76 and 0.43, respectively), and for the other levels, the SEM values were zero. The SRD, derived from the SEM, indicates whether the intervention effect exceeds the measurement error and enables clinical judgment about the occurrence of real change as a result of a specific interventions.[28] To our knowledge, the SRD has not been documented previously for the GMFM-88.

In the present study, we calculated the SRD for the GMFM-88 in each dimension and the total score, providing a threshold for interpreting changes in GMFM-88 scores over time. For example, the total gross motor ability improvement of less than 3.14 could be due to natural variability in measurement error, even if no real change was obtained through the therapy. With this threshold in mind, the results of this

study indicated slight improvement at 1 month after the intervention, showing increases of 4.1 points in GMFCS levels I and II and 4.5 points in GMFCS levels III to V. However, the increments were certainly greater than the SRM at 3 months and 6 months after the intervention in both GMFCS groups (Tab. 4). The interrater SRD was large for GMFCS level I (Tab. 3). Children at GMFCS level I have good function in walking, stair walking, and jumping, which are dynamic activities. Although the raters had pretest training, such dynamic activities might be difficult for therapists with less experience to measure. Again, the tendency of lowered reliability in more difficult items suggests that training is needed prior to performing the GMFM.

It is important that a function-measuring tool have responsiveness (ie, the ability to detect changes over time). Some researchers have studied the responsiveness of the GMFM-88 in monitoring the effectiveness of interventions.[1,12] Nordmark et al[31] reported the need to classify individuals according to age and severity of functional impairment to assess responsiveness in children with CP. In the present study, we also classified the children between the ages of 13 months to 9 years 5 months into 2 subgroups according to their severity as less impaired (GMFCS levels I and II) and more severely impaired (GMFCS levels III–V) to analyze responsiveness of the GMFM-88, respectively. The results showed large responsiveness of the GMFM-88 goal total score as reflected in changes after 6 months, regardless of functional severity (GMFCS levels I and II versus levels III–V: ES=0.8, 0.9 and SRM=2.0, 2.0, respectively), according to Cohen's criteria.[30] This finding is similar to that in a previous study, which demonstrated responsiveness of GMFM, but included only GMFCS levels I and II.[12]

In the present study, the values of SRM were higher than those of ES across all scores on the GMFM-88, indicating that the standard deviation of the changes in score was more homogenous than that of the baseline scores. The SRM seems to be more representative than the ES when applied to a group with varying functional ability at baseline. Because SRM uses the between-subject variability of the individual change scores over time as a denominator, it provides more appropriate standardization than ES, which uses between-subject variability of the baseline score instead. A higher SRM indicates better responsiveness.

In summary, we investigated the reliability and responsiveness of the GMFM-88 for a relatively large sample of children with CP. The results indicated reliability and responsiveness.

Limitations

Because the present study was conducted at 1 center and the age range of the patient population was limited, the results should not be generalized to all children with CP. Thus, a multicenter study with a broader range of ages of the study sample should be conducted.

Conclusion

Each dimension and the total score of the GMFM-88 showed acceptable levels of relative reliability and absolute reliability. Also, responsiveness of the goal total score of the GMFM-88 was high in children with CP regardless of their functional severity.

Both authors provided concept/idea/research design, writing, data collection, project management, and facilities/equipment. Dr Ko provided data analysis and clerical support. Dr Kim provided study participants, institutional liaisons, and consultation (including review of rapid proof before submission).

The authors express their gratitude to the children and families who participated in the study. Special thanks go to the pediatric physical therapists in the Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, who administered the assessments for the study.

Approval for this study was granted by the Institutional Review Board of CHA Bundang Medical Center, CHA University.

DOI: 10.2522/ptj.20110374

References

1 Lundkvist Josenby A, Jarnlo GB, Gummesson C, Nordmark E. Longitudinal construct validity of the GMFM-88 total score and goal total score and the GMFM-66 score in a 5-year follow-up study. *Phys Ther.* 2009; 89:342–350.

2 Aisen ML, Kerkovich D, Mast J, et al. Cerebral palsy: clinical care and neurological rehabilitation. *Lancet Neurol.* 2011;10: 844–852.

3 Duncan B, Shen K, Zou LP, et al. Evaluating intense rehabilitative therapies with and without acupuncture for children with cerebral palsy: a randomized controlled trial. *Arch Phys Med Rehabil.* 2012; 93:808–815.

4 Davis MF. Measuring impairment and functional limitations in children with cerebral palsy. *Disabil Rehabil.* 2011;33:2416–2424.

5 Russell DJ, Rosenbaum PL, Avery LM, Lane M. *Gross Motor Function Measure (GMFM-66 and GMFM-88) User's Manual.* London, United Kingdom: Mac Keith Press; 2002.

6 Russell DJ, Rosenbaum PL, Cadman DT, et al. The Gross Motor Function Measure: a means to evaluate the effects of physical therapy. *Dev Med Child Neurol.* 1989;31: 341–352.

7 Rosenbaum P, Paneth N, Leviton A, et al. A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl.* 2007;109:8–14.

8 Zarrinkalam R, Russo RN, Gibson CS, et al. CP or not CP? A review of diagnosis in a cerebral palsy register. *Pediatr Neurol.* 2010;42:177–180.

9 Harbourne RT, Willett S, Kyvelidou A, et al. A comparison of interventions for children with cerebral palsy to improve sitting postural control: a clinical trial. *Phys Ther.* 2010;90:1881–1898.

10 Unlu E, Cevikol A, Bal B, et al. Multilevel botulinum toxin type a as a treatment for spasticity in children with cerebral palsy: a retrospective study. *Clinics (Sao Paulo).* 2010;65:613–619.

11 Sterba JA. Adaptive downhill skiing in children with cerebral palsy: effect on gross motor function. *Pediatr Phys Ther.* 2006; 18:289–296.

12 Vos-Vromans DC, Ketelaar M, Gorter JW. Responsiveness of evaluative measures for children with cerebral palsy: the Gross Motor Function Measure and the Pediatric Evaluation of Disability Inventory. *Disabil Rehabil*. 2005;27:1245–1252.

13 Lu YM, Lin JH, Hsiao SF, et al. The relative and absolute reliability of leg muscle strength testing by a handheld dynamometer. *J Strength Cond Res*. 2011;25:1065–1071.

14 Bland JM, Altman DG. Measurement error and correlation coefficients. *BMJ*. 1996; 313:41–42.

15 Harris KD, Heer DM, Roy TC, et al. Reliability of a measurement of neck flexor muscle endurance. *Phys Ther*. 2005;85: 1349–1355.

16 Flansbjer UB, Holmbäck AM, Downham D, Lexell J. What change in isokinetic knee muscle strength can be detected in men and women with hemiparesis after stroke? *Clin Rehabil*. 2005;19:514–522.

17 Giraudeau B, Ravaud P, Chastang C. Importance of reproducibility in responsiveness issues. *Biometrical J*. 1998;40: 685–701.

18 Kazis LE, Anderson JJ, Meenan RF. Effect size for interpreting changes in health status. *Med Care*. 1989;27:S178–S189.

19 Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopedic evaluation. *Med Care*. 1990; 28:632–642.

20 Heyrman L, Molenaers G, Desloovere K, et al. A clinical tool to measure trunk control in children with cerebral palsy: the Trunk Control Measurement Scale. *Res Dev Disabil*. 2011;32:2624–2635.

21 Palisano RJ, Rosenbaum PL, Walter SD, et al. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 1997;39:214–223.

22 Rosenbaum PL, Walter SD, Hanna SE, et al. Prognosis for gross motor function in cerebral palsy: creation of motor development curves. *JAMA*. 2002:288; 1357–1363.

23 Jeng SF, Yau KI, Chen LC, Hsiao SF. Alberta Infant Motor Scale: reliability and validity when used on preterm infants in Taiwan. *Phys Ther*. 2000;80:168–178.

24 Ketelaar M, Verneer A, Helders P. Functional motor abilities of children with cerebral palsy: a systemic literature review of assessment measures. *Clin Rehabil*. 1998; 12:369–380.

25 Polit DF, Hungler BP. *Essential of Nursing Research: Methods, Appraisal, and Utilization*. Philadelphia, PA: Lippincott Williams & Wilkins; 1989.

26 Nunnally JC. *Psychometric Theory*. 2nd ed. New York, NY: McGraw-Hill; 1978.

27 McKenna L, Cunningham J, Straker L. Inter-tester reliability of scapular position in junior elite swimmer. *Phys Ther Sport*. 2004;5:146–155.

28 Beckerman H, Roebroeck ME, Lankhorst GJ, et al. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001;10:571–578.

29 Damstra J, Huddleston Slater JJ, Fourie Z, Ren Y. Reliability and the smallest detectable differences of lateral cephalometric measurements. *Am J Orthod Dentofacial Orthop*. 2010;138:546.e1–546.e8.

30 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

31 Nordmark E, Hägglund G, Jarnlo GB. Reliability of the Gross Motor Function Measure in cerebral palsy. *Scand J Rehabil Med*. 1997;29:25–28.

32 Liaw LJ, Hsieh CL, Lo SK, et al. The relative and absolute reliability of two balance performance measures in chronic stroke patients. *Disabil Rehabil*. 2008;30: 656–661.