

# Development and reliability of a system to classify gross motor function in children with cerebral palsy

Robert Palisano  
Peter Rosenbaum  
Stephen Walter  
Dianne Russell  
Ellen Wood  
Barbara Galuppi

To address the need for a standardized system to classify the gross motor function of children with cerebral palsy, the authors developed a five-level classification system analogous to the staging and grading systems used in medicine. Nominal group process and Delphi survey consensus methods were used to examine content validity and revise the classification system until consensus among 48 experts (physical therapists, occupational therapists, and developmental pediatricians with expertise in cerebral palsy) was achieved. Interrater reliability ( $\kappa$ ) was 0.55 for children less than 2 years of age and 0.75 for children 2 to 12 years of age. The classification system has application for clinical practice, research, teaching, and administration.

'Cerebral palsy' refers to a group of disorders in the development of posture and motor control, occurring as a result of a non-progressive lesion of the developing central nervous system (Bax 1964). This definition encompasses a wide variety of pathological and clinical entities that have in common a developmental motor disorder that can vary in etiology, manifestations, severity, prognosis, and comorbidities. Despite the best efforts of professionals from many disciplines, a great deal remains unknown about the natural history of cerebral palsy. It is recognized that motor outcome is roughly related to 'severity' but to date there has not been a generally accepted standardized system of classification of severity of motor disability for use in clinical practice and research.

Methods of classification that have been proposed are based on (a) pathophysiology or neuroanatomical location of the lesion (Fay 1950, Perlstein 1952, Minear 1956); (b) impairments in muscle tone, reflex activity, and voluntary control of movement (Fay 1950, Perlstein 1952); (c) parts of the body most involved (Balf and Ingram 1955, Minear 1956); (d) ambulatory status (Badell-Ribera 1985, Yokochi et al. 1993), or (e) degree of motor impairment (e.g. 'mild', 'moderate', 'severe') (Balf and Ingram 1955, Minear 1956, Yokochi et al. 1993). These methods of classification rely heavily on clinical judgment and are primarily of value for diagnosis. Their reliability and validity have not been investigated. Evans et al. (1989) developed a recording form to collect data on impairment and disability in children with neuromotor dysfunction for use in a register on childhood impairment. The form includes items to rate a child's head control, trunk control, gait, and upper limb function but does not include an overall classification of motor ability.

We have adopted an alternative approach to classification, based on the concepts of disability and functional limitation. The International Classification of Impairments, Disabilities, and Handicaps (ICIDH) developed by the World Health Organization (1980) defines disability as 'the restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being'. The concept of functional limitation is included in the models of disablement developed by Nagi (1965) and the National Center for Medical Rehabilitation Research in the United States (1993). Nagi defined functional limitation as a 'limitation in performance at the level of the whole person'. We believe that classification of children with cerebral palsy on the basis of abilities and limitations in gross motor function should enhance communication among professionals and families with respect to (1) determining a child's needs and making management decisions, (2) the creation of databases describing the development of children with cerebral palsy, and (3) comparing and generalizing the results of program evaluations and research into the outcome of treatment. Furthermore if the classification system is found to have predictive validity, early classification of a child (e.g. at age 2 years) would help parents to anticipate their child's later motor function.

The purposes of this study were (1) to construct a gross motor function classification system for children with cerebral palsy, analogous to the staging and grading systems used for tumours, (2) to examine content validity through nominal group process and Delphi survey consensus methods involving developmental therapists and pediatricians with expertise in cerebral palsy, and (3) to determine interrater reliability of the classification system. We wanted a descriptive classification

system that would be quick and easy to use, valid, and reliable.

### Method

The Gross Motor Disability Classification System was developed in four phases. First the authors drafted the system. During phases two and three, the validity of the content was examined using nominal group process and Delphi survey consensus methods, and the classification system was revised. Interrater reliability was examined in phase four.

### DEVELOPMENT OF THE CLASSIFICATION SYSTEM

The classification system was initially drafted using several methods of inquiry. These included a review of existing classification systems and research on development of children with cerebral palsy; examination of data from 275 children to whom the Gross Motor Function Measure (GMFM) (Russell et al. 1989) was administered twice within 6 months; a review of developmental records and videotapes of children with cerebral palsy identified by their current therapists as having mild, moderate, or severe involvement; and extensive discussions by the study team.

The initial impetus to create a gross motor function classification system arose in part from the examination of observations on gross motor function collected by the authors in

earlier studies of the development of the GMFM (Russell et al. 1989) and the Gross Motor Performance Measure (GMPM) (Boyce et al. 1995) (see also Scrutton and Rosenbaum 1996). Many children were studied in the course of developing these measures, including preschool-aged children without delays in motor development (N=60), and children with mild (N=53), moderate (N=99), or severe (N=63) cerebral palsy. Plots of GMFM total scores against age produced curves that appear, upon visual and statistical analysis, to differ importantly depending on the degree of disability as described by the severity level assigned to each child's condition (Fig. 1). The original designations of severity of cerebral palsy were made in an unstandardized way by each child's regular physical therapist. Based on the apparent utility of a classification of gross motor function for children with cerebral palsy, as illustrated by these 'motor growth curves', we believed that a standardized and reliable classification system could be created that would have wide applicability.

The next step in drafting the classification system was to examine individual scores for items that represent common motor milestones. These were examined in each of the five dimensions of the GMFM: lying and rolling; sitting; crawling and kneeling; standing; and walking, running, and jumping. There was wide variation and overlap between groups, partic-

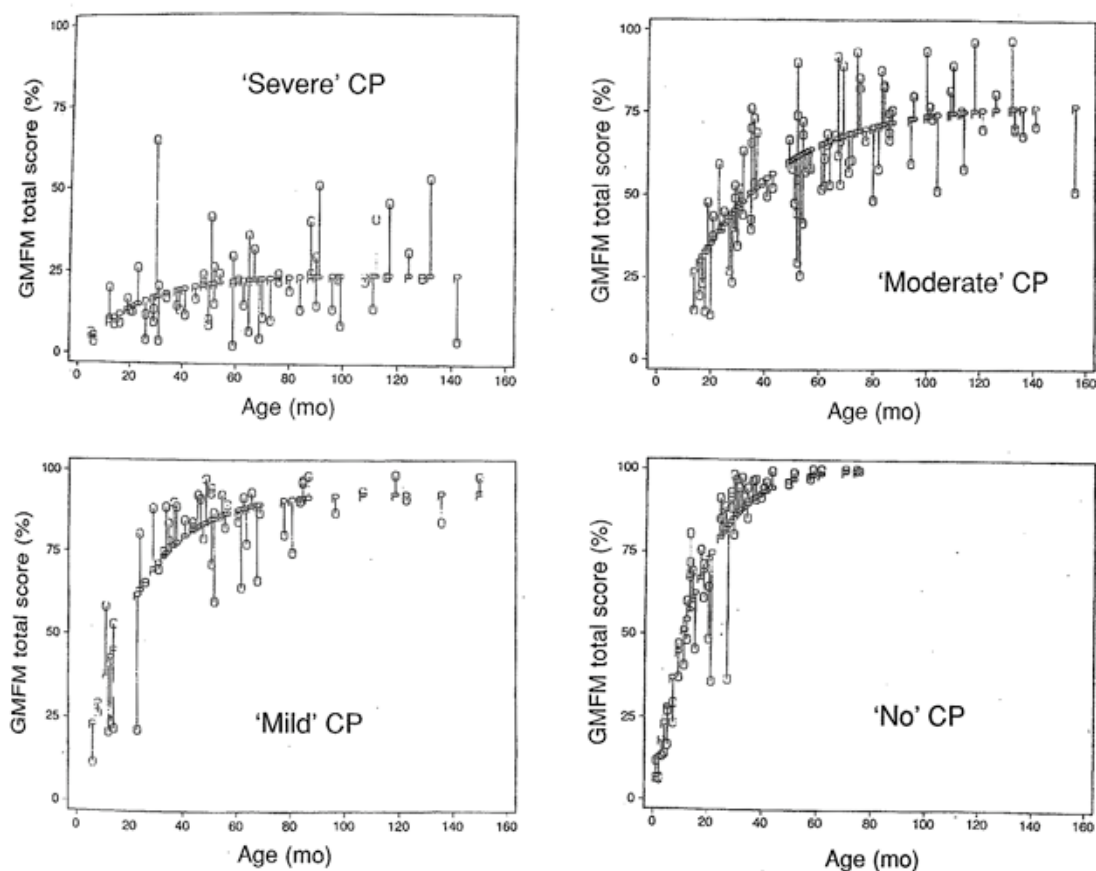


Figure 1: Plots of total scores on the Gross Motor Function Measure (GMFM) against age of children according to degree of motor involvement [children without motor delays (N=60) and children with mild (N=53), moderate (N=99) or severe (N=63) cerebral palsy].

ularly in the scores of children classified as having moderate or severe cerebral palsy. After examining the GMFM data and viewing the videotapes, we decided that a three-level classification system did not adequately represent the variation in gross motor development of children with cerebral palsy.

Eventually a five-level classification system was proposed which in the authors' opinion represented clinically meaningful distinctions in motor function. In keeping with the World Health Organization's construct of disability, we focused the classification system on self-initiated movement, with emphasis on function in sitting and walking. Distinctions between levels focused on functional limitations, the need for assistive technology including mobility devices (such as walkers, crutches, and canes), wheeled mobility, and to a lesser extent quality of movement. The system is ordinal, with no intent that the distance between levels be considered equal or that children with cerebral palsy are equally distributed among the five levels. The classification system was intended for use with children 18 months to 6 years of age. In keeping with the goal that the classification system be quick and easy to use based on reported or observed gross motor function rather than on standardized testing, brief descriptions were provided for each level as well as a summary of the distinctions between each pair of levels. The descriptions were broad and were not intended to assess in detail the development of individual children but rather to determine which of the five levels most closely resembles a child's current gross motor function.

#### NOMINAL GROUP PROCESS

In phase two, content validity was examined using a modified nominal group consensus method. Nominal group process consists of a structured meeting of a knowledgeable target group in which issues are discussed among participants in an attempt to reach a consensus (Fink et al. 1984). The procedure was modified, in that the first draft of the classification system was developed by the authors without input from the target group. The initial draft of the classification system and a questionnaire were sent to occupational therapists and physical therapists at three children's treatment centres in southern Ontario 3 weeks before the consensus meeting. Twenty-eight therapists provided feedback on the content and on the applicability of the classification system to children on their caseload. Of the 28 participants, 14 were occupational therapists and 14 were physical therapists. The therapists' clinical experience ranged from 5 to 30 years, with a mean of 9.6.

Representatives from each of the three centres brought the feedback from their centre and participated in the consensus meeting. In total, 13 therapists and two developmental pediatricians were voting participants in a half-day consensus exercise. The exercise started with an open discussion by participants of their general impressions of the classification system and the process used to accumulate the feedback from their centre. The group then proceeded to respond to 17 statements. One member of the study team chaired the session, two served as discussants, and a fourth member served as recorder/timer. The discussant introduced each preset statement and then facilitated discussion among the participants. The statements addressed the applicability of the classification system, how the system should be administered, the number of levels, the content of each level, distinctions between levels, and the ages of children who could be classified. When the time limit for discussion of each statement was reached,

the Chair took a formal vote. The participants were asked to approve or reject each statement. For each statement, the consensus target specified in advance was approval by at least 12 of the 15 participants.

#### DELPHI SURVEY

The Delphi consensus method is an attempt to obtain expert opinion in a systematic manner through questionnaires, with the ultimate goal of generating a group response (Fink et al. 1984). The survey is conducted over two or more rounds. After each round, the panel of experts is informed of the results, which are then used to construct the questionnaire for the next round. The process ends when the group reaches a preset level of agreement or when sufficient information has been exchanged to achieve the goals of the process.

In selecting a panel we considered the recommendations of Fink et al. (1984), who stated that consensus participants should have content expertise, be representative of their professions, and have the ability to implement findings. In addition, we selected participants who we thought would provide insightful and critical input and advice. Twenty-one physical therapists, occupational therapists, and developmental pediatricians from North America, Europe, and Australia were asked to participate and 20 agreed (Appendix A). The participants were recognized leaders in the field of developmental disability with expertise in cerebral palsy. None of the persons selected had participated in the nominal group consensus meeting. The years of professional experience of the panel members varied from 6 to 38 years, with a mean of 20 years.

The questionnaire for round one of the Delphi survey was divided into three parts and included 38 statements. Many statements were followed by an open-ended question to encourage the experts to elaborate upon their responses. Part A consisted of nine statements that addressed the need for and the construct of the classification system. Part B consisted of 23 statements that addressed the definition and description for each of the five levels and the distinctions between levels. The experts were asked to use the classification system to classify the motor function of children from their clinical practice. Those who did use it completed Part C, which included a statement on the ease of using the classification system and invited comments. Part D consisted of five statements addressing potential uses of the classification system.

Each statement was rated using a seven-point scale, with a rating of 1 indicating 'strong disagreement', 4 indicating 'indifference', and 7 indicating 'strong agreement' with the statement. Before mailing the questionnaires, consensus agreement for each item was defined as 16 or more of the 20 experts rating the statement as 5 or higher. Each participant was mailed the questionnaire for round one of the Delphi survey, the revised Gross Motor Function Classification System, an explanation of why the classification system was being developed, and the work to date on development of the classification system.

The questionnaire for round two included 28 statements. The 23 statements from Part B of the questionnaire used in round one were repeated. These statements addressed the definition and description for each of the five levels and the distinctions between the levels. The remaining five statements addressed how the classification system should be administered and the ages of children who could be accurately classified. These were areas where consensus was not achieved in

round one. Along with the questionnaire for round two of the Delphi survey, the experts were also mailed the revised Gross Motor Function Classification System, a summary of the responses from round one of the survey, and a copy of their own completed questionnaire from round one.

#### INTRATER RELIABILITY

Intrater reliability was examined after round two of the Delphi survey was completed, and the final revisions were made to the classification system. A convenience sample of five children's treatment centres in southern Ontario participated in the study. Physical therapists and occupational therapists at the five centres were asked to compile a list of children 12 years of age or younger with diagnosed or suspected cerebral palsy. The children included on the list were on active caseloads (seen within the past 6 months) and had been known to at least two therapists for at least 3 months. The lists were divided into two age strata: under 2 years of age and 2 years or older. Seventy-seven children (37 in the younger age group and 40 in the older age group) were randomly selected from the lists provided by the therapists, with the specification that no pair of therapists would classify more than five children.

Each child's level of gross motor function was classified independently by two therapists who were familiar with the child. Twenty-six physical therapists and 25 occupational therapists participated. The therapists' years of experience varied from 3 months to 31 years (mean 10.2 years). The therapists relied on their knowledge of the child's motor abilities to classify the child's level of gross motor function and were not required to observe or assess each child. Once the therapists had familiarized themselves with the classification system, the time required to classify a child should not have exceeded 10 minutes.

Intrater reliability was analyzed separately for each age group using the crude percentage of agreement and the kappa ( $\kappa$ ) statistic, a measure of chance-corrected agreement. In addition to the overall level of chance-corrected agreement, category-specific  $\kappa$  values were derived to determine chance-corrected agreement on each of the five levels of the classification system. Category associations were evaluated by computing the conditional probabilities that given that one randomly selected therapist chooses a particular level of function, the second therapist will choose each of the five levels of function.

#### Results

##### NOMINAL GROUP PROCESS

Consensus or approval of a statement by at least 12 of the 15 voting participants was reached for 12 of the 17 statements. All of the participants rejected two of the statements: 'The classification system is sufficient to represent the variation in motor disability among children with cerebral palsy' and 'The system should focus on best ever performance.' Only 11 of the 15 voting participants approved the statement 'A child can be classified accurately from a parent/caregiver report using a structured interview.' Only 10 approved the statement that 18 months should be the lower age limit for the classification system, and only five approved the statement that 12 months should be the lower age limit for the classification system. Based on discussion among the participants of the statements where consensus was not achieved, the descriptions for each level of gross motor function and the distinctions between lev-

els of the classification system were modified. The consensus that the classification system should reflect a child's typical performance in home, school, and community settings, and that age limits below 18 months and above 6 years should be explored, were also implemented. The age range of children with cerebral palsy who could be classified was expanded to include 12 months to 12 years.

#### DELPHI SURVEY

For round one of the survey, consensus among the 20 experts was achieved for 26 of the 38 statements. All of the experts agreed that there is a need for a classification system for children with cerebral palsy that is based on the construct of disability and functional limitation. The statements for which consensus agreement of 80% or greater was not achieved are listed in Table I. Of 18 respondents, only 12 to 14 of the experts indicated that the description for each of the five levels adequately distinguished among children of varying ages. Of 19 respondents, only 14 agreed with the definition for level I, and 15 indicated that the description for level I was clinically meaningful. Fifteen of 19 respondents agreed that the distinction between levels II and III was well defined and helpful in making the classification. Only 14 of 20 of the experts indicated that a child's level of motor function could be accurately classified based on a structured interview with a parent or caregiver, while 15 of the 20 indicated that it could be accu-

**Table I: Statements for which the 20 experts did not reach consensus after round one of the Delphi survey**

Statement	Responses $\geq 5$ (total respondents)
1. A child's level of motor function can be accurately classified based on a professional's familiarity with the child over a period of at least 3 months, without necessitating any additional assessments.	15 (20)
2. A child's level of motor function can be accurately classified using a structured interview with a parent/caregiver	14 (20)
3. This classification system can be accurately used for infants as young as 12 months of age.	12 (20)
4. The Level I definition is appropriate.	14 (19)
5. The Level I description is clinically meaningful.	15 (19)
6. The Level I description provides adequate distinctions for children at varying ages.	12 (18)
7. The Level II description provides adequate distinctions for children at varying ages.	12 (18)
8. The Level III description provides adequate distinctions for children at varying ages.	12 (18)
9. The Level IV description provides adequate distinctions for children at varying ages.	13 (18)
10. The Level V description provides adequate distinctions for children at varying ages.	14 (18)
11. The distinction between Level II and Level III is well defined.	15 (19)
12. The distinction between Level II and Level III is helpful in making the classification.	15 (19)

Each statement was rated using a seven-point scale with a rating of 1 indicating 'strong disagreement', a rating of 4 indicating 'indifference', and a rating of 7 indicating 'strong agreement'. The table includes only the statements for which the preset level of consensus (80% of responses at or above a rating of 5) was not reached.

rately classified based on a professional's familiarity with the child over a period of at least 3 months. Only 12 of 20 of the experts agreed that the classification system could be used for infants as young as 12 months of age.

All of the experts indicated that a classification system for children with cerebral palsy has applications for research and teaching. Nineteen of the 20 experts indicated that it has application for clinical practice, 17 indicated that it has application for administration, and 15 indicated that it has application for sports participation.

Eleven of the 20 experts used the system to classify the motor function of between 2 and 16 children with cerebral palsy. All 11 experts indicated that they found the classification system easy to use. The system was revised a second time based on the results of the Delphi survey and the experts' written comments.

For round two of the survey, completed questionnaires were received from 18 of the 20 experts. Consensus among the 18 experts was achieved for all 29 statements. All 18 of the experts agreed with 15 of the statements and 17 agreed with 11 of the statements. Sixteen of the experts agreed with the statement that the classification system could be accurately used for children 1 to 2 years of age. Fifteen agreed that an accurate classification could be made from a structured interview with a parent or caregiver. Nine of the 20 experts used the revised system to classify the motor function of between 3 and 16 children with cerebral palsy. All nine experts indicated that they found the system easy to use. Based on the experts' written comments, revisions were made to clarify some of the definitions and descriptions. The current version of the Gross Motor Function Classification System is in Appendix B.

#### INTER-RATER RELIABILITY

Table II lists the distribution of agreements and disagreements among the five levels of the classification system by age group. Overall, the ratings were fairly evenly distributed among the five levels. For the 37 children less than 2 years of age, paired therapists agreed on the level of gross motor function for 24, disagreed by one level for 11, and disagreed by two levels for 2. Of these 13 disagreements, 4 were between Levels IV and V, and 1 was between Levels III and V. For the 40 children 2 to 12 years of age, paired therapists disagreed by one level for only 8; of these disagreements, 5 were between Levels I and II.

The  $\kappa$  coefficient is a standard agreement statistic used to correct for chance agreement. For children under 2 years of

age,  $\kappa$  was 0.55 while for children 2 years of age or older,  $\kappa$  was 0.75. The results provide moderate support for the overall reliability of the classification system. It has been suggested that values of  $\kappa$  greater than 0.75 represent excellent agreement beyond chance, and values in the range 0.40 to 0.75 represent fair to good agreement beyond chance (Fleiss 1981).

Association and agreement statistics that analyze inter-rater reliability for each of the five levels of the classification system by age group are presented in Tables III and IV respectively. In both age groups, the proportion of the total number of classifications made was reasonably evenly distributed over the five levels. For children under 2 years of age, the  $\kappa$  coefficients (Table III) indicate better chance-corrected agreement for Levels I and II than for the other three levels. For instance the  $\kappa$  for Level I is 0.841, indicating excellent distinguishability of Level I from the others, whereas that for Level V is only 0.368, indicating that Level V is more often confused with the other levels.

The category association statistics (Table IV) indicate that for children under age 2 the association between the classifications of the paired therapists was particularly strong for Levels I and II. That is, if the first observer selected Level I, there was an 88% chance that the second observer would agree. In contrast, disagreement between paired therapists was higher for the other levels, especially for Levels IV and V, which are more difficult to distinguish from each other in the younger age group. For example, if Level V was chosen by the first observer, there was only a 44% chance that the second observer would agree, with the discordant ratings being primarily at Level IV (44% chance) or occasionally at Level III (11% chance).

For children above the age of 2 years, the patterns were somewhat reversed (Table III). The  $\kappa$  coefficients were highest (all above 0.8) for Levels III to V, while agreement on Levels I and II was poorer. The association statistics (Table IV) indicate some difficulty in distinguishing between Levels I and II. When Level I or II was chosen by the first observer, the second observer agreed only 67% and 55% of the time respectively. There was, however, clear distinction of Levels I and II from the other levels: disagreement above and below the Level-II to Level-III boundary was never observed. There was only minor disagreement in distinguishing among Levels III to V. The second observer agreed 92, 87, and 89% of the time for categories III, IV, and V respectively; the disagreements always concerned adjacent categories.

Table II: Percentages of agreement and disagreement between paired therapists for each level of the classification system, by age group

Level	Younger age group (<2 y)						Older age group (2-12 y)					
	N=37			N=40			N=40			N=40		
	Agreements		Disagreements		Agreements		Disagreements		Agreements		Disagreements	
	N	%	Levels	N	%	Level	N	%	Level	N	%	
I	7	18.9	I/II	2	5.4	I	5	12.5	I/II	5	12.5	
II	7	18.9	II/III	3	8.1	II	3	7.5	III/IV	1	2.5	
III	4	10.8	III/IV	2	5.4	III	6	15.0	IV/V	2	5.0	
IV	4	10.8	IV/V	4	10.8	IV	10	25.0				
V	2	5.4				V	8	20.0				
			II/IV	1	2.7							
			III/V	1	2.7							