10-1-2018

Stability of the Gross Motor Function Classification System, Manual Ability Classification System, and Communication Function Classification System.

Robert J Palisano

Lisa Avery

Jan Willem Gorter

Barbara Galuppi

Sarah Westcott McCoy

DMCN Developmental Medicine & Child Neurology

Stability of the gross motor function, manual ability, and communication function classification systems

Journal:	*Developmental Medicine & Child Neurology*
Manuscript ID	DMCN-OA-17-10-0581.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Palisano, Robert; Drexel University, Department of Physical Therapy and Rehabilitation Sciences Avery, Lisa; Independent Statistician, Gorter, Jan Willem; McMaster University, Pediatrics; Galuppi, Barbara; McMaster University, CanChild Centre for Childhood Disability Research McCoy, Sarah; University of Washington, Rehabilitation Medicine
Keywords:	cerebral palsy, GMFCS, MACS, CFCS, Stability

Stability of the gross motor function, manual ability, and communication function classification systems

Robert J Palisano[1,2], Lisa Avery[2], Jan Willem Gorter[2,3], Barbara Galuppi[2], Sarah Westcott McCoy[4]

1 Department of Physical Therapy and Rehabilitation Science, Drexel University, Philadelphia, PA, USA.
2 Avery Information Services, Orillia, ON, Canada
3 Department of Pediatrics, McMaster University, Hamilton, ON, Canada.
4 Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA.

Correspondence to Robert J Palisano, Drexel University, College of Nursing and Health Professions, 1601 Cherry Street, MS 10501, Philadelphia, PA 19102, USA. E-mail: robert.j.palisano@drexel.edu

Word count: 3,182

Abstract

AIM To determine stability of the Gross Motor Function Classification System (GMFCS), Manual Ability Classification System (MACS), and Communication Function Classification System (CFCS) over 1-year and 2-year intervals using a process for consensus classification between parents and therapists.

METHOD Participants were 664 children with cerebral palsy (CP), 1.5 to 12.0 years of age, one of their parents, and 90 therapists. Consensus between parents and therapists on level of function was ≥ 92% for the GMFCS, MACS, and CFCS. A linearly weighted kappa coefficient of ≥ .75 was the criterion for stability.

RESULTS Kappa coefficients varied from .76-.88 for the GMFCS, .59-.73 for the MACS, and .57-.77 for the CFCS. For children < 4 years, level of function did not change for 58.2% on the GMFCS, 30.3% on the MACS, and 39.3% on the CFCS. For children ≥ 4 years, level of function did not change for 72.3% on the GMFCS, 49.1% on the MACS, and 55% on the CFCS.

INTERPRETATION The findings support repeated classification of children over time. The kappa coefficients for the GMFCS are attributed to descriptions of levels for each age band. Consensus classification facilitates discussion between parents and professionals that has implications for shared decision making.

What this paper adds

- The findings support repeated classification of children over time.
- Stability was higher for the GMFCS than the MACS and CFCS.
- The function of younger children was more likely to be reclassified.
- Percent agreement between parents and therapists using consensus classification varied from 92-97%
- The ICC overestimated stability compared with the weighted Kappa coefficient.

Running foot: Stability of classification systems for CP

The Gross Motor Function Classification System (GMFCS),[1, 2] Manual Ability Classification System (MACS),[3] and Communication Function Classification System(CFCS)[4] were developed to objectively classify children and adolescents with cerebral palsy (CP) for purposes of effective communication, setting goals, informing decisions on services and interventions, and applying research findings to practice. Each system has five levels that are intended to represent differences in function that are meaningful in daily life. The GMFCS includes separate descriptions for five age bands while the MACS and CFCS include a single description of each level of function that is applicable to all ages. For each system, a classification is made by determining which level best represents the child's current function throughout the day.

Stability of a classification system refers to the extent to which children remain in the same level of function over time. Evidence of stability of the GMFCS was provided in a study of 610 children with CP whose function was classified by physical therapists between 2-7 times (mean 4.3).[5] Mean time between first and last ratings was 33.5 months (SD 10.3). The weighted kappa coefficient for the first and last ratings was .84 for children < 6 years of age (percentage of agreement 75.7%) and .89 (percentage of agreement 82.9%) for children ≥6 years, indicating excellent chance-corrected agreement. Children were reclassified more than one level .08% of the time. Children whose function was initially classified at Levels I or V were least likely to be reclassified; children < 6 years were more likely to be reclassified to a lower level of function. In a study of 107 children at GMFCS levels II and III who had single event multi-level surgery, 95% remained in the same GMFCS level an average of five years post-surgery.[6] A physician who classified children Test-retest reliability of the Turkish version of the GMFCS was examined with a single physician, percent agreement was 75% and the ICC was .94.[7]

Evidence of stability of the MACS also has been reported. Ohrvall et al.[8] evaluated stability in children 4-17 years of age with CP. The intra-class correlation coefficient (ICC) between classifications (n=1,267) made 12 months apart by occupational therapists was .97(percentage of agreement 82%). The ICC between classifications (n=445) over 3-5 years was .96 (percentage of agreement 78%). Children were reclassified more than one level less than 1% of the time. The results did not differ between younger and older children. Imms et al.[9] evaluated stability of the MACS and GMFCS in 86 children with CP whose function was classified at a mean of 11 and 12 years of age by caregiver report. The ICC was .92 for both the GMFCS and MACS (percentage agreement 79% and 67% respectively). Test-retest reliability of the MACS has been reported for the Turkish[10] (ICC .91-97), Persian[11] (weighted kappa = .87), and Portuguese (Brazil)[12] (occupational therapy student rater, unweighted kappa=.83; occupational therapist, unweighted kappa =.95) language versions. To the best of our knowledge data on stability of the CFCS has not been reported.

Previously we reported that using a process for consensus classification, parents and therapists agreed on level of function 97.8%, 96.7%, and 94.5% of the time for the GMFCS, MACS, and CFCS respectively.[13] The aim of this study was to determine the stability of the GMFCS, MACS, and CFCS over a 1-year and 2-year interval using the same process for consensus classification. A linearly weighted kappa coefficient of ≥ .75 was the criterion used for stability.[14] We anticipated that stability would be higher for children ≥ 4 years compared with children < 4 years of age. For each classification system, we also examined whether children's function was more likely to be reclassified to a higher or lower level of function, whether change

in classification was related to distribution of limb involvement, country of residence (Canada, United States) or sex, and whether reclassification in one system was associated with reclassification in the other two systems. Because the GMFCS describes levels of function for age bands rather than a single description, we anticipated that stability would be higher for the GMFCS than the MACS and CFCS.

METHOD

Design

This study was part of a multi-site, prospective longitudinal observational cohort study of children with CP conducted in Canada and the United States referred to as the 'On Track Study'.

Participants

Participants were a convenience sample of 664 children with CP, 1.5 to 12.0 years of age at the start of the study (mean = 6.0, SD= 2.7), one of their parents, 88 physical therapists and two occupational therapists. Children had a diagnosis of CP reported by parents or were suspected to have a diagnosis of CP, i.e., they exhibited delayed motor development, muscle stiffness, and difficulties with balance and moving. Eligibility to participate was confirmed throughout the study so that the final sample represented children with CP. Therapist assessors provided detailed information for consideration of eligibility of 71 children either before or after recruitment. A physiatrist (JWG) reviewed the information and made recommendations regarding eligibility; 11 children were excluded from the final sample as a result of this review. The questionnaire and the three classification systems were available in English, French, and Spanish. Parents who could not read or communicate in one of these languages were not eligible to participate in this study. All parents completed the English language measures with the exception of two parents who completed the Spanish language measures.

The questionnaire completed by parents provided demographic information. Fifty percent of children were Canadian. Seventy-two percent of the children were White, 8% Black/African American, 6% Asian, 2% American Indian/Alaska Native, 11% identified in multiple race categories, and 1% did not respond. One hundred ninety-three (29.1%) children had unilateral limb involvement, 175 (26.4%) had diplegia, 295 (44.4%) had either tri- or quadriplegia, and limb distribution was missing for one child (0.1%). Eighty-eight percent of parents were mothers; 97% had some form of post-secondary education.

Recruiting was done by regional coordinators and managed centrally by the project coordinator for each country. Using convenience sampling, participants were recruited from clinical settings in six provinces of Canada (British Columbia, Saskatchewan, Manitoba, Ontario, Nova Scotia, and Newfoundland) and four metropolitan regions in the United States (Seattle, WA; Atlanta, GA; Oklahoma City, OK; and Philadelphia, PA). Ethical approval was provided by the Health Sciences Research Ethics Board at Western University and ethics boards at McMaster University, Drexel University, the University of Washington, Mercer University, Oklahoma University of Health Sciences, and multiple agencies across participating sites. Signed informed consent/assent was obtained from parent/child participants. All parents consented to data being

used in publications. Therapist assessors for the On Track Study were identified by the regional coordinator at each site.

Classification Systems

The *Gross Motor Function Classification System (GMFCS)* was developed for children with CP 12 years of age and younger[1] and subsequently expanded to include a 12 to 18 year age band and revised to include environmental and personal considerations.[2] Classifications are made based on the child's self-initiated movements with emphasis on sitting and walking. Inter-rater reliability and validity has been reported.[1,2,5,15-17]

The *Manual Ability Classification System* (MACS) was developed for children with CP, 4 to 18 years of age.[3] Function is classified based on the child's self-initiated ability to handle objects during daily activities. Reliability and validity of the MACS have been demonstrated.[3, 18] After data collection began, the Mini-MACS was published for children with CP, 1-4 years of age and, therefore, was not used in this study.[19]

The *Communication Function Classification System (CFCS)* was developed for use with individuals with CP, 2 years of age and older.[4] Function is classified based on the child's everyday performance of all methods of communicating, including speech, gestures, eye gaze, facial expressions, augmentative, and alternative communication. Validity of the CFCS was reported for preschool age children with varied speech and language disorders.[20]

Procedure

Prior to data collection, therapists attended a one-day workshop for training on all measures used in the On Track Study including the GMFCS, MACS, and CFCS. Function was classified at the first assessment, 12 months (mean 12.5, SD 1.1) and 24 months (mean 23.5, SD 1.9) after the first classification. Among the 187 children < 4 years at the first assessment, 67 were ≥4 by the 12 month assessment and 96 were ≥4 by the 24 month assessment.

The process for consensus classification by Bartlett et al.[13] was used. At the beginning of each assessment, parents independently classified their children's level of function on the GMFCS, MACS, and CFCS. During the assessment, therapists independently classified the children's levels of function. Parents and therapists then discussed their classifications and the therapist documented: 1) the parent and therapist each classified the child as having the same level of function, 2) consensus on level of function was reached after discussion, or 3) consensus was not achieved. Guidelines were generated to reconcile disagreements. Fundamentally, we relied on parents' classifications. The level of function provided by the therapist was used only when the therapist provided compelling comments on the classification form. Our rationale is that parents know their children the best, see them in multiple settings, and are most able to describe usual performance.

Consensus on level of function between therapists and parents was 97% for the GMFCS, 96% for the MACS, and 94% for the CFCS at the initial assessment (664 children); 97% for the GMFCS, 93% for the MACS, and 92% for the CFCS at the 12 month assessment (645 children), and 97% for the GMFCS, 97% for the MACS, and 94% for the CFCS at the 24 month assessment (422 children). When consensus was not achieved, therapist and parent disagreement was most often within one level; 88-100% of the time for GMFCS, 81-93% of the time for MACS, and 71-92% of the time for CFCS. In these cases, the parent's classification level was

used with specific guidelines to determine if the assessor's classification level should be used instead.

Statistical Analysis

Statistical analysis were performed in R version 3.3.3.[21] Calculations of weighted kappa were performed with the psych package.[22] Four contingency tables were created for each classification system, two for children < 4 years and two for children ≥ 4 years of age. Within each age group, one contingency table compared the first and 12 month classifications and the second table compared the first and 24 month classifications. Chance-corrected agreement using the linearly weighted kappa statistic and simple percentage agreement were computed. Linear weighting accounts for the magnitude of disagreement between ratings; disagreement by one classification level is less severe than disagreement by two or more levels. For consistency with previous research on stability of the GMFCS[5] we used the criteria proposed by Fleiss[14] to interpret kappa; kappa <.40 poor agreement, .40-.75 fair to good agreement, and >.75 excellent chance-corrected agreement. To enable comparison with other studies, we computed the intra-class correlation coefficient (ICC), noting that the weighted kappa with squared weights is equivalent to the ICC.[23]

The proportion of children whose classification did not change and the proportion of children whose function was reclassified one or two times were computed to provide a better sense of the stability of the systems for individual children. Bowker's test of symmetry was used to determine if there was a propensity for function to be reclassified to a higher or lower functional level.[24] The alpha level for all analyses was p<.05.

To determine factors associated with stability of each classification system, the 411 children classified three times were dichotomized as 'stable' if their level of function did not change or 'not stable' if their level of function changed the second or third time. Logistic regression was used to determine likelihood of reclassification based on initial classification level and age. Finally, Spearman correlations were computed to determine whether reclassification in one system was associated with reclassification in one or both of the other two systems.

RESULTS

Cross-tabulations, kappa coefficients, percentage of agreement, ICCs, and tests of symmetry are presented in Tables I-III for the GMFCS, MACS, and CFCS respectively. For the GMFCS, linearly weighted kappa varied from .76-.88 (percentage of agreement 64.5% - 80.3%) and the ICC varied from .89-.95 (Table I). For the MACS, linearly weighted Kappa varied from .59-.73 (percentage of agreement 49.2% - 66.7%) and the ICC varied from .77-.87 (Table II). For the CFCS, linearly weighted Kappa varied from .57-.77 (percentage of agreement 51.6% - 69.7%) and the ICC varied from .71-.89 (Table III).

Children < 4 years of age whose function was reclassified at the 24 month assessment were more likely to be classified to a higher level of function on the MACS (p=.04) and CFCS (p<.001). Children in both age groups whose function was reclassified at the 12 month assessment, were more often re-classified to a higher level of function on the CFCS (p<.05).

For children < 4 years of age, level of function did not change for only 58.2% of children on the GMFCS, 30.3% on the MACS, and 39.3% on the CFCS. The proportion of children whose

function was reclassified twice was 9% for the GMFCS, 24.6% for the MACS, and 22.1% for the CFCS. For children ≥ 4 years of age, level of function did not change for 72.3% of children on the GMFCS, 49.1% on the MACS, and 55% on the CFCS. The proportion of children whose function was reclassified twice was 8.3% for the GMFCS, 18.7% for the MACS, and 15.2% for the CFCS.

Results of the logistic regression are given in Table IV. Likelihood of at least one reclassification was related to initial classification level and age for the GMFCS and MACS (Table IV). Younger children were more likely to be reclassified. Children were more likely to change classification level if their initial GMFCS level was II-IV (OR 2.29-2.56); initial MACS level was III (OR 5.65) or IV (OR 2.81), or initial CFCS level was II-V (OR 5.55-16.36). Children whose function was reclassified on one system were not more likely to have their function reclassified on either of the other two systems (Spearman correlations varied from -0.06 to 0.15).

DISCUSSION

The kappa coefficients, the primary measure of stability in our study, provide evidence of stability of the GMFCS, MACS, and CFCS for children with CP 12 years of age and younger. For the GMFCS, chance-corrected agreement for classifications made at 12 month and 24 month intervals was excellent (kappa coefficients ≥ .75). For the MACS and CFCS, chance-corrected agreement was good (kappa coefficients .57-.73) and there was excellent chance-corrected agreement on the CFCS for children ≥ 4 years of age for the 12 month interval. With one exception, kappa coefficients were higher for children ≥ 4 years of age and for classifications made 12 months apart, however, differences were not analyzed statistically and many were small. As hypothesized, we attribute the higher chance-corrected agreement for the GMFCS to the descriptions of levels for each age band rather than the single description across ages for the MACS and CFCS.

The number of children whose function was reclassified, especially children ≤ 4 years of age on the MACS and CFCS, indicates that children with CP do not always remain at the same level of function over time. The percentage agreement between classifications 12 and 24 months apart on the GMFCS for children 4-12 years of age are comparable to the percentage agreement previously reported for children whose function was classified at a mean of 11and 12 years of age[8] and children and youth 4-17 years over a 3-5 year period.[7] Our findings for the MACS are similar to the percentage agreement reported by Imms et al.[8] but lower than the percentage agreement reported by Ohrvall et al.[7] The MACS was developed for children 4-18 years of age, therefore, the low percentage agreement for children ≤ 4 years of age in our study is not entirely unexpected, also given the moderate interobserver reliability of the MACS for young children.[18] The Mini-MACS[19] has recently been published with an emphasis on age-appropriate descriptions of manual abilities and should be used to classify children with CP < 4 years of age.

To our knowledge, this is the first report of stability of the CFCS. Our impression is that distinguishing between levels of communication function (sending and receiving information) including differences in communicating with familiar and unfamiliar partners is more challenging for parents and therapists than distinguishing between levels of gross motor function. Over a 12 and 24 month period, changes in environmental and personal factors that impact function in daily life may have contributed to the number of children whose function was

reclassified on the CFCS. Additionally, the MACS and CFCS require judgement of expectations for manual ability and communication function at different ages, especially for younger children.

Our results suggest that the ICC overestimates stability. There is a discrepancy between high ICCs and the number of children whose function was reclassified in our study and previous research. As we stated earlier, the ICC is equivalent to a weighted kappa with quadratic weights, which differs from a linearly weighted kappa in the amount that discordant ratings are penalized. Because the weighted kappa is constrained to -1 to +1 the higher the penalty imposed upon ratings further apart (as happens in the ICC) the lower the influence of ratings that only differ by a single level. Because ratings that differ by only a single level comprise almost all of the discordant ratings in these classification systems, the linearly weighted kappa will always be lower than the ICC. This, in our opinion leads to a situation where the ICC amplifies the true stability. This is best illustrated by an example from Table 1; the agreement between classifications 12 months apart of children >4 years on the GMFCS. The percentage agreement is high, at 79.4%, but the corollary is that over 20% of children changed levels. However, because only two children had ratings that differed by more than a single level, the ICC is 0.95, implying very high stability. In contrast, the linearly weighted kappa is 0.87, which we think reflects both the stability of the measure and the fact that the initial classification is not immutable.

Our perspective is that the GMFCS,[2] MACS,[3] and CFCS[5] are complementary and collectively provide valuable information for shared decisions on goals, services, and interventions for children and youth with CP. Our finding that children whose function was reclassified on one system were not more likely to be reclassified on either of the other two systems supports this perspective. For research, we asked parents and therapists to independently make classifications prior to discussion. In practice, we envision collaboration among parents, children, and service providers, especially since classifications are based on usual performance in daily life. Although our findings provide evidence of stability, the percentage agreement between classifications made 12 and 24 months apart indicates that function will be reclassified for some children, hence the need to gross motor (GMFCS), manual ability (MACS), and communication(CFCS) function repeatedly over time, especially for children under age 4. The value of consensus classification is that the process facilitates discussion between parents and professionals that has implications for shared decision making on goals, services and interventions.

REFERENCES

1. Palisano RJ, Rosenbaum P, Bartlett D, et al. Content validity of the expanded and revised gross motor function classification system. *Dev Med Child Neurol* 2008; **50**:744-50.

2. Palisano RJ, Rosenbaum P, Walter S, et al. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997; **39**:214-23.

3. Eliasson AC, Krumlinde Sundholm L, et al. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* 2006; **48**: 549-54.

4. Hidecker MJC, Paneth N, Rosenbaum PL, et al. Developing and validating the Communication Function Classification System (CFCS) for individuals with cerebral palsy. *Dev Med Child Neurol* 2011; **53**:704-10.

5. Palisano R, Cameron D, Rosenbaum PL, et al. Stability of the Gross Motor Function Classification System. *Dev Med Child Neurol* 2006; **48**:424-28.

6. Rutz E, Tirosh O, Thomason P, et al. Stability of the Gross Motor Function Classification System after single-event multilevel surgery in children with cerebral palsy. *Dev Med Child Neurol* 2012; 54:1109-13.

7. Ozlem El, Baydar M, Berk H, Peker O, Kosay C, Demiral Y. Interobserver reliability of the Turkish version of the expanded and revised gross motor function classification system. *Disabil Rehabil* 2012; 34:1030-3.

8. Öhrvall AM, Krumlinde-Sundholm L, Eliasson AC. The stability of the Manual Ability Classification System over time. *Dev Med Child Neurol* 2014; **56**:185-89.

9. Imms C, Carlin J, Eliasson AC. Stability of caregiver-reported manual ability and gross motor function classifications of cerebral palsy. *Dev Med Child Neurol* 2010; 52:153-59.

10. Akpinar P, Tezel CG, Eliasson AC, Icagasioglu A. Reliability and cross-cultural validation of the Turkish version of Manual Ability Classification System (MACS) for children with cerebral palsy. *Disabil Rehabil* 2010; 32:1910-6.

11. Riyahi A, Rassafiani M, AkbarFahimi N, Sahaf R, Yazdani F. Cross-cultural validation of the Persian version of the Manual Ability Classification System for children with cerebral palsy. *Int J Ther Rehabil* 2013; 20:19-24.

12. Silva DB, Funayama CA, Pfeifer LI. Manual Ability Classification System (MACS): reliability between therapists and parents in Brazil. *Braz J Phys Ther* 2015; 19:26-33.

13. Bartlett DJ, Galuppi B, Palisano RJ, et al. Consensus classifications of gross motor, manual ability, and communication function classification systems between therapists and parents of children with cerebral palsy. *Dev Med Child Neurol* 2016; 58:98-9.

14. Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York: John Wiley and Sons, 1981, pp 38-46.

15. Wood E, Rosenbaum P. The Gross Motor Function Classification System for Cerebral Palsy: a study of reliability and stability over time. *Dev Med Child Neurol* 2000; **42**: 292-96.

16. Rosenbaum, PL, Walter SD, Hanna SE, et al. Prognosis for Gross Motor Function in Cerebral Palsy: Creation of Motor Development Curves. *JAMA* 2002; 288:1357-63.

17. Gorter JW, Ketelaar M, Rosenbaum P, et al. Use of the GMFCS in infants with CP: the need for reclassification at age 2 years or older. *Dev Med Child Neurol* 2009; **51**:46-52.

18. Plasschaert VF, Ketelaar M, Nijnuis MG, et al. Classification of manual abilities in children with cerebral palsy under 5 years of age: how reliable is the Manual Ability Classification System? *Clin Rehabil* 2009; **23**:164-70.

19. Eliasson AC, Ullenhag A, Wahlström U, et al. Mini-MACS: development of the Manual Ability Classification System for children younger than 4 years of age with signs of cerebral palsy. *Dev Med Child Neurol* 2017; **59**:72-8.

20. Hidecker MJ, Cunningham BJ, Thomas-Stonell N, et al. Validity of the Communication Function Classification System for use with preschool children with communication disorders. *Dev Med Child Neurol* 2017; 59:526-30.

21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2016.

22. Revelle W. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-roject.org/package=psych Version = 1.5.8. 2015.

23. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and psychological measurement, Vol. 33, 1973, pp 613–619.

24. Agresti A. Categorical Data Analysis, 2nd ed. Hoboken: John Wiley and Sons, 2002, p 437.

Table I: GMFCS: cross-tabulations, Kappa coefficients, percent agreement, and tests of symmetry

Children <4 years (n=187)

First Visit	12 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (70)	**66**	3	1	0	0
Level II (36)	8	**17**	10	1	0
Level III (22)	3	6	**11**	2	0
Level IV (32)	0	0	4	**23**	5
Level V (27)	0	0	0	5	**22**

Percent Agreement = 74.3%

Linearly Weighted Kappa = 0.83, 95% CI (0.78, 0.87)

Squared Weighted Kappa (ICC) = 0.92, 95% CI (0.09, 0.95)

Bowker's Test of Symmetry = 8.61, p = 0.57

26 reclassified as more functional; 22 as less functional

Children <4 years (n=124)

First Visit	24 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (47)	**39**	7	0	1	0
Level II (28)	4	**12**	11	1	0
Level III (14)	1	2	**7**	4	0
Level IV (18)	0	0	3	**10**	5
Level V (17)	0	0	0	5	**12**

Percent Agreement = 64.5%

Linearly Weighted Kappa = 0.76, 95% CI (0.69, 0.83

Squared Weighted Kappa (ICC)= 0.89 95% CI (0.84, 0.94)

Bowker's Test of Symmetry = 17.57, p = 0.06

15 reclassified as more functional; 29 as less functional

Children >4 years (n=465)

First Visit	12 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (143)	**123**	20	0	0	0
Level II (112)	24	**85**	2	0	1
Level III (50)	0	8	**34**	8	0
Level IV (86)	0	1	9	**65**	11
Level V (74)	0	0	0	12	**62**

Percent Agreement = 79.4%

Linearly Weighted Kappa = 0.87, 95% CI (0.85, 0.89)

Squared Weighted Kappa (ICC) = 0.95, 95% CI (0.93, 0.96)

Bowker's Test of Symmetry = 10.72, p = 0.38

54 reclassified as more functional; 42 as less functional

Children >4 years (n=299)

First Visit	24 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (88)	**69**	19	0	0	0
Level II (70)	13	**54**	3	0	0
Level III (32)	0	7	**23**	2	0
Level IV (57)	0	0	2	**44**	11
Level V (52)	0	0	0	2	**50**

Percent Agreement = 80.3%

Linearly Weighted Kappa = 0.88, 95% CI (0.85, 0.91)

Squared Weighted Kappa (ICC) = 0.96, 95% CI (0.95, 0.97)

Bowker's Test of Symmetry = 10.56, p = 0.39

24 reclassified as more functional; 35 as less functional

Table II: MACS: cross-tabulations, Kappa coefficients, percent agreement, and tests of symmetry

Children <4 years (n=187)

	12 Month Visit				
First Visit	**Level I**	**Level II**	**Level III**	**Level IV**	**Level V**
Level I (39)	**27**	12	0	0	0
Level II (73)	18	**39**	11	5	0
Level III (25)	0	10	**12**	3	0
Level IV (40)	1	4	13	**18**	4
Level V (10)	0	0	1	2	**7**

Percent Agreement = 55.1%

Linearly Weighted Kappa = 0.61, 95% CI (0.53, 0.68)

Squared Weighted Kappa (ICC) = 0.77, 95%CI (0.70, 0.83)

Bowker's Test of Symmetry = 16.68, p = 0.08

49 reclassified as more functional; 35 as less functional

Children <4 years (n=124)

	24 Month Visit				
First Visit	**Level I**	**Level II**	**Level III**	**Level IV**	**Level V**
Level I (22)	**16**	6	0	0	0
Level II (55)	19	**30**	6	0	0
Level III (16)	0	7	**5**	4	0
Level IV (26)	0	3	7	**7**	9
Level V (5)	0	0	0	2	**3**

Percent Agreement = 49.2%

Linearly Weighted Kappa = 0.59, 95% CI (0.51, 0.67)

Squared Weighted Kappa (ICC) = 0.80, 95%CI (0.74, 0.85)

Bowker's Test of Symmetry = 19, p = 0.04

38 reclassified as more functional; 25 as less functional

Children >4 years (n=465)

	12 Month Visit				
First Visit	**Level I**	**Level II**	**Level III**	**Level IV**	**Level V**
Level I (92)	**68**	20	4	0	0
Level II (186)	28	**137**	20	1	0
Level III (69)	5	29	**27**	8	0
Level IV (71)	0	3	17	**44**	7
Level V (47)	0	0	2	11	**34**

Percent Agreement = 66.7%

Linearly Weighted Kappa = 0.73, 95% CI (0.69, 0.77)

Squared Weighted Kappa (ICC) = 0.86, 95%CI (0.83, 0.89)

Bowker's Test of Symmetry = 16.12, p = 0.1

95 reclassified as more functional; 60 as less functional

Children >4 years (n=299)

	24 Month Visit				
First Visit	**Level I**	**Level II**	**Level III**	**Level IV**	**Level V**
Level I (50)	**35**	12	3	0	0
Level II (125)	29	**83**	12	1	0
Level III (43)	1	14	**20**	6	2
Level IV (47)	0	1	9	**25**	12
Level V (34)	0	0	0	7	**27**

Percent Agreement = 63.5%

Linearly Weighted Kappa = 0.72, 95% CI (0.68, 0.77)

Squared Weighted Kappa (ICC) = 0.87, 95%CI (0.84, 0.90)

Bowker's Test of Symmetry = 12.87, p = 0.23

61 reclassified as more functional; 48 as less functional

Table III: CFCS: cross-tabulations, Kappa coefficients, percent agreement, and tests of symmetry

Children <4 years (n=187)

First Visit	12 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (52)	**47**	4	1	0	0
Level II (31)	10	**13**	5	3	0
Level III (46)	7	12	**21**	6	0
Level IV (45)	2	3	11	**24**	5
Level V (13)	0	0	2	5	**6**

Percent Agreement = 59.4%

Linearly Weighted Kappa = 0.65, 95% CI (0.58, 0.72)

Squared Weighted Kappa (ICC) = 0.78, 95%CI (0.72, 0.85)

Bowker's Test of Symmetry = 19.78, p = 0.03

52 reclassified as more functional; 24 as less functional

Children <4 years (n=124)

First Visit	24 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (31)	**30**	1	0	0	0
Level II (23)	10	**10**	2	1	0
Level III (29)	6	9	**9**	3	2
Level IV (33)	3	4	9	**12**	5
Level V (8)	0	0	2	3	**3**

Percent Agreement = 51.6%

Linearly Weighted Kappa = 0.57, 95% CI (0.47, 0.66)

Squared Weighted Kappa (ICC) = 0.71, 95%CI (0.62, 0.80)

Bowker's Test of Symmetry = 35.37, p < 0.001

46 reclassified as more functional; 14 as less functional

Children >4 years (n=465)

First Visit	12 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (190)	**164**	23	3	0	0
Level II (84)	36	**38**	10	0	0
Level III (78)	4	13	**49**	9	3
Level IV (77)	1	6	17	**44**	9
Level V (36)	0	0	0	7	**29**

Percent Agreement = 69.7%

Linearly Weighted Kappa = 0.77, 95% CI (0.73, 0.80)

Squared Weighted Kappa (ICC) = 0.89, 95%CI (0.86, 0.91)

Bowker's Test of Symmetry = 24.96, p = 0.01

84 reclassified as more functional; 57 as less functional

Children >4 years (n=299)

First Visit	24 Month Visit Level I	Level II	Level III	Level IV	Level V
Level I (120)	**105**	14	1	0	0
Level II (46)	16	**21**	7	2	0
Level III (58)	3	16	**25**	12	2
Level IV (48)	1	2	16	**23**	6
Level V (27)	0	0	0	8	**19**

Percent Agreement = 64.5%

Linearly Weighted Kappa = 0.74, 95% CI (0.69, 0.78)

Squared Weighted Kappa (ICC) = 0.87, 95%CI (0.84, 0.91)

Bowker's Test of Symmetry = 12.61, p = 0.25

62 reclassified as more functional; 44 as less functional

Table IV: Logistic regression coefficients examining the likelihood of change in classification at least once across three assessments, subsample of N=411. Negative coefficients (B) indicate a lower likelihood of re-classification relative to reference group. Reference group is classification level I.

	B	SE	OR	z	p
DV is GMFCS Classification constant across 3 assessments yes=0 no=1*					
Intercept	-0.41	0.30	0.67	-1.3	0.182
GMFCS Level II	0.91	0.30	2.48	3.0	**0.002**
GMFCS Level III	0.83	0.37	2.29	2.2	**0.024**
GMFCS Level IV	0.94	0.32	2.56	3.0	**0.003**
GMFCS Level V	-0.26	0.38	0.77	-0.7	0.499
Age in months	-0.01	0.004	0.99	-3.3	**0.001**
DV is MACS Classification constant across 3 assessments yes=0 no=1					
Intercept	0.66	0.34	1.94	2	0.048
MACS Level II	0.32	0.29	1.38	1.1	0.267
MACS Level III	1.73	0.42	5.65	4.1	**<0.001**
MACS Level IV	1.03	0.36	2.81	2.9	**0.004**
MACS Level V	-0.17	0.42	0.84	-0.4	0.684
Age in months	-0.01	0.003	0.99	-3.8	**<0.001**
DV is CFCS Classification constant across 3 assessments yes=0 no=1					
Intercept	-1.12	0.34	0.33	-3.3	<0.001
CFCS Level II	2.79	0.37	16.36	7.6	**<0.001**
CFCS Level III	2.35	0.32	10.48	7.3	**<0.001**
CFCS Level IV	2.25	0.33	9.50	6.9	**<0.001**
CFCS Level V	1.71	0.42	5.55	4.1	**<0.001**
Age in months	-0.01	0.004	0.99	-1.8	0.076

DV: dependent variable