Gross Motor Function Classification System and outcome tools for assessing ambulatory cerebral palsy: a multicenter study

D J Oeffinger*; C M Tylkowski, Shriners Hospitals for Children; M K Rayens, Biostatistics Consulting Unit, University of Kentucky Medical Center; R F Davis, Shriners Hospitals for Children, Lexington, KY; G E Gorton, III, Shriners Hospitals for Children, Springfield, MA; J D’Astous; D E Nicholson, Shriners Hospitals for Children, Intermountain, Salt Lake City, UT; D L Damiano, Washington University, St. Louis, MO; M F Abel, University of Virginia, Department of Orthopedics, Charlottesville, VA; A M Bagley, Shriners Hospitals for Children, Northern California, Sacramento, CA; J Luan, Biostatistics Consulting Unit, University of Kentucky Medical Center, Lexington, KY, USA.

*Correspondence to first author at 1900 Richmond Rd, Lexington, KY 40502, USA.
E-mail: doeffinger@shrinenet.org

The relationships between different levels of severity of ambulatory cerebral palsy, defined by the Gross Motor Function Classification System (GMFCS), and several pediatric outcome instruments were examined. Data from the Gross Motor Function Measure (GMFM), Pediatric Orthopaedic Data Collection Instrument (PODCI), temporal–spatial gait parameters, and oxygen cost were collected from six sites. The sample size for each assessment tool ranged from 226 to 1047 participants. There were significant differences among GMFCS levels I, II, and III for many of the outcome tools assessed in this study. Strong correlations were seen between GMFCS level and each of the GMFM sections D and E scores, the PODCI measures of Transfer and Mobility, and Sports and Physical Function, Gait Velocity, and Oxygen Cost. Correlations among tools demonstrated that the GMFM sections D and E scores correlated with the largest number of other tools. Logistic regression showed GMFM section E score to be a significant predictor of GMFCS level. GMFM section E score can be used to predict GMFCS level relatively accurately (76.6%). Study data indicate that the assessed outcome tools can distinguish between children with different GMFCS levels. This study establishes justification for using the GMFCS as a classification system in clinical studies.

See last page for list of abbreviations.
Gross Motor Function Measure (GMFM; Russell et al. 2002), the Pediatric Outcomes Data Collection Instrument (PODCI), temporal–spatial gait parameters, and energy utilization. The GMFM is a standardized observational criterion-referenced tool for assessing change in gross motor function over time in children with CP, with composite scoring in five dimensions: Lying and Rolling (section A), Sitting (section B), Crawling and Kneeling (section C), Standing (section D), and Walking, Running and Jumping (section E). The PODCI measures functional musculoskeletal health and assesses upper extremity function, transfer and mobility, sports and physical function, comfort, expectations for treatment, happiness, and satisfaction. A parent and adolescent self-reporting questionnaire format is used to obtain the information (Daltroy et al. 1998). Each of these outcome tools has been tested individually for content validity and test reliability (Nordmark et al. 1997, Palisano et al. 1997, Daltroy et al. 1998, Russell et al. 2000, Wood and Rosenbaum 2000).

Previous classifications of CP were based on subjective assessments of motor involvement (mild, moderate, or severe) and thus were never validated for reliability (Minear 1956). Consequently, within-group or intergroup comparisons for treatment could not be made objectively. Palisano and colleagues (Palisano et al. 1997) developed the Gross Motor Function Classification System (GMFCS) in 1997. The GMFCS provides a standardized system to classify the gross motor function of children with CP into five levels (level I the least severe to level V the most severe). The GMFCS has been found to be valid and reliable (Palisano et al. 1997, Wood and Rosenbaum 2000, Rosenbaum et al. 2002). Wood and Rosenbaum (2000) calculated intrarater reliability using a generalizability (G) coefficient and found high intrarater reliability (G = 0.93).

There have been several studies that have used outcome tools to evaluate the effectiveness of specific surgical procedures, the development of community function in children with disabilities, as well as the efficacy of physical therapy interventions (Boyce et al. 1995, Gowland et al. 1995; Ottenbacher et al. 1996, 1997; Daltroy et al. 1998; Damiano and Abel 1998; Novacheck et al. 2000; Haynes and Sullivan 2001). Some have found correlations between the PODCI and GMFM (Abel et al. 2003), the PODCI and gait parameters (Tervo et al. 2002, Novacheck et al. 2000), the GMFM and Pediatric Evaluation of Disability Index (PEDI; McCarthy et al. 2002), the GMFM and gait analyses (Damiano and Abel 1996), and the GMFM and temporal–spatial characteristics of gait (Damiano and Abel 1996, Drouin et al. 1996). Other studies have looked at the GMFCS and its ability to predict early classification of a child’s future motor function, as well as the relationship between the GMFM and the GMFCS (Palisano et al. 1997, 2000). However, no study has determined how children with CP of different levels of severity, as defined by the GMFCS classification, perform on commonly employed outcome tests.

The primary purpose of this study was to assess the relationships between the condition severity level of children with ambulatory CP, as defined by the GMFCS functional levels I–III, and data obtained from the GMFM, PODCI, temporal–spatial gait parameters, and oxygen cost. This study also investigated: (1) the descriptive characteristics of the outcome tools, including differences in average scores among levels I to III of the GMFCS; (2) the relationships among the outcome tools; (3) whether any of these outcome tools predict GMFCS level; and (4) whether the predictors of GMFCS level can be used to cluster patients in a way that is consistent with the GMFCS.

Method

Children with the diagnosis of CP and a GMFCS classification of level I, II, or III, were identified for inclusion in this retrospective study. GMFCS levels I to III incorporate the majority of ambulatory children with CP (i.e. those most likely to be seen in motion analysis laboratories). Participants were males and females between the ages of 4 and 21 years old, mean age was 11 years 2 months (SD 4 years).

The specific measurement parameters of GMFM sections D (Standing) and E (Walking, Running, and Jumping), parent report PODCI questionnaire, temporal–spatial gait data, and oxygen cost were chosen as appropriate measures of function in this population. Only sections D and E of the GMFCS were included as they are most applicable to the ambulatory population and are more routinely collected in motion analysis laboratories. Oxygen cost (ml/Kg/m) was chosen over oxygen consumption (ml/kg/min) measures to allow for comparisons across participants walking at different speeds. While not the only available measures, the investigators felt that the chosen outcome tools include a representation of technical measures (temporal–spatial data and oxygen cost), clinician observed rating (GMFM), and parent report (PODCI) of a child’s functional level.

A multicenter approach was used both to obtain a large sample size and to obtain findings that would be applicable to a wide array of pediatric patients with CP. Six pediatric orthopaedic facilities that routinely treat children with CP and administer the GMFEM and PODCI, perform standard gait analyses, and measure oxygen cost were identified and asked to participate in the study (Shriners Hospitals for Children, Lexington, Springfield, Intermountain, and Northern California, Kluge Children’s Rehabilitation Center at the University of Virginia, and Washington University in St. Louis, MO). All six sites received the required Institutional Review Board (IRB) approval for retrospective review of data and contributed the relevant data from all eligible participants. The retrospective nature of the data collection precluded an a priori power analysis and sound statistical practice made estimating power after the sample sizes unwise (Hoening and Heisey 2001); however, the number of cases included is relatively large compared with similar published studies of this population of patients.

Data were compiled for those patients who met the inclusion criteria. The data were coded with only local site identifiers to ensure patient confidentiality. For each participant, GMFM section D and E scores, parent-report PODCI scores, temporal–spatial data obtained from a standard gait study, and oxygen cost data were recorded in a standardized spreadsheet provided to each site by the lead investigators.

Temporal–spatial data are reported as a percentage of an age-matched normally developing population. Each site sent their raw data to the lead facility where the raw values were converted to a percentage of normal according to the database of normative values from the Motion Lab at the Shriners Hospital for Children, Lexington. This normalization allowed for comparisons across ages and groups.

All data were compiled at the primary site for analyses. All available data for each participant were included. However, due to the retrospective nature of the study, we could not
control participant fulfillment of all assessments. Each participant varied in the number of assessments completed. There were fewer data for some of the assessments due to factors such as participant age (PODCI) and cooperation (oxygen cost). The sample size for each assessment tool (including GMFM, PODCI, temporal–spatial variables, and oxygen cost) ranged from 226 to 1047 participants.

**STATISTICAL ANALYSIS**

Data from all participating facilities were combined and summarized, as appropriate to the level of measurement, using descriptive methods. The descriptive analysis included means, standard deviations, and ranges for continuous variables and frequency distributions for categorical variables.

To investigate if there were significant differences among the three GMFCS levels for each parameter, the data were analyzed using a series of one-way analysis of variance (ANOVA) models, with GMFCS as the factor (independent variable) in the model. Post-hoc analysis was accomplished using Scheffe’s post-hoc procedure for pairwise comparisons with significance set at \( p \leq 0.05 \). The 95% confidence intervals were developed for each of the outcome tools by GMFCS level in order to provide a range within which the average value for the assessment tools is expected to occur for the given level of the GMFCS.

Two-sample \( t \)-tests were used to compare the average PODCI scores obtained in this study with two sets of published normative data (Haynes and Sullivan 2001, Hunsaker et al. 2002). Comparisons were performed between both sets of normative values for the complete group of study participants as well as by GMFCS level.

Spearman’s rank correlation was used to assess the relationships between the GMFCS and each of the continuous outcome tools used in the study. This form of the test for association was chosen because the GMFCS is ordinal rather than continuous. However, as the level of measurement of these tools is continuous, relationships among the different outcome tools were assessed using Pearson’s product–moment correlation.

A logistic regression model was developed to investigate which performance measures were predictive of GMFCS levels. The response variable (dependent variable) in the prediction model was GMFCS level. A logistic regression, using a cumulative logit model appropriate for the case of a categorical dependent variable with more than two levels, was chosen to evaluate the impact of the outcome tools (continuous independent variables) on the GMFCS (three-level ordinal variable). The selection of outcome tool parameters for inclusion in the model was based on the Spearman’s rank correlation analysis (i.e. variables significantly related to GMFCS were considered as candidates). The list of potential predictors of GMFCS level included: GMFM section D and E scores; PODCI scores of Transfer and Mobility, Sports and Physical Function, and Global Function; Gait Velocity, Stride Length, and Cadence; and Oxygen Cost. A stepwise procedure was used to determine the final model. The stepwise regression process begins without any predictors in the model. Assuming there is at least one potential outcome tool parameter significantly associated with the GMFCS level, the first predictor to enter the model is the one most strongly related to the dependent variable. The variable that has the second-strongest association with the dependent variable is added as a second predictor and then the significance of both variables in the model is evaluated. If either predictor is no longer significant (as evidenced by a \( p \) value \( > 0.05 \)), it is removed from the model and the predictor with the third most significant relationship with the dependent variable is added to the model. This process continues until each potential predictor has been tested for inclusion in the model and all variables remaining in the model are significant.

Using the results of the logistic model, a cluster analysis based on the average linkage method was performed to determine whether any variables that were identified as predictors of GMFCS level could be used to form groups of participants, such that the participants within a group were similar while the participants from two groups were distinct. The average linkage method is one of the forms of cluster analysis that is hierarchical. The process begins with each observation forming a cluster of size one. In the first step (i.e. when each observation is its own cluster), the two observations that are the most similar to each other in the dataset (i.e. are the closest to each other or have the smallest distance between them) are joined and that forms a cluster of size two. The next step is to find the two clusters that are most similar (this could be two other observations or an observation and the cluster of two) and join them to form a new cluster. The average linkage method determines which clusters to join by not only minimizing the

<table>
<thead>
<tr>
<th>NCMRR Dimension</th>
<th>Description (Palisano et al. 1994)</th>
<th>Addressed using</th>
<th>Addressed using</th>
<th>Addressed using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathophysiology</td>
<td>Interruption or interference of normal physiology and developmental processes or structures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impairment</td>
<td>Loss or abnormality of body structure or body function</td>
<td>PODCI comfort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional Limitation</td>
<td>Restriction of ability to perform activities</td>
<td>GMFM sections D and E</td>
<td>Gait velocity self-selected</td>
<td>Energy expenditure</td>
</tr>
<tr>
<td>Disability</td>
<td>Inability to participate in typical societal role functions</td>
<td>GMFCS</td>
<td>PODCI Physical disability sections</td>
<td></td>
</tr>
<tr>
<td>Societal Limitation</td>
<td>Barriers to full participation in society that result from attitudes, architectural barriers, and social policies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NCMRR, National Center for Medical Rehabilitation Research; GMFM, Gross Motor Function Measure; PODCI, Pediatric Orthopaedic Data Collection Instrument; GMFCS, Gross Motor Function Classification System.
distance between clusters to be joined in a given step but also by maximizing distance between clusters that remain unjoined in the same step. As the goal of this cluster analysis was to determine if any of the assessment tools could divide the participants into groups that were similar to GMFCS groupings, this process continued until the number of clusters retained was three.

Results
The total study population included data from 1047 participants combined from the six facilities. Of these, 457 (44%) were classified as being GMFCS level I, 286 (27%) were classified as being GMFCS level II, and 304 (28%) were classified as GMFCS level III. There was no difference in mean age at evaluation among the three GMFCS levels \((F=0.8; p=0.4)\). GMFCS level I

Table III: Comparisons of Pediatric Orthopaedic Data Collection Instrument (PODCI) data for participants with CP, GMFCS levels I, II, and III with published normative values

<table>
<thead>
<tr>
<th>PODCI subscale</th>
<th>Normative dataset 1 (AAOS)</th>
<th>Normative dataset 2 (Houston)</th>
<th>Current study all participants</th>
<th>Current study GMFCS I</th>
<th>Current study GMFCS II</th>
<th>Current study GMFCS III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
<td>n</td>
<td>Mean</td>
</tr>
<tr>
<td>Upper Extremity and Physical Function</td>
<td>98.8</td>
<td>99.4</td>
<td>76.3^a</td>
<td>83.0^a</td>
<td>75.9^a</td>
<td>68.0^a</td>
</tr>
<tr>
<td>Transfer and Mobility</td>
<td>99.1</td>
<td>99.7</td>
<td>75.5^a</td>
<td>87.7^a</td>
<td>77.8^a</td>
<td>57.6^a</td>
</tr>
<tr>
<td>Sports and Physical Function</td>
<td>93.7</td>
<td>95.5</td>
<td>50.6^a</td>
<td>66.6^a</td>
<td>47.7^a</td>
<td>33.7^a</td>
</tr>
<tr>
<td>Pain and Comfort</td>
<td>89.0</td>
<td>91.1</td>
<td>72.0^a</td>
<td>69.7^a</td>
<td>71.8^a</td>
<td>75.4^a</td>
</tr>
<tr>
<td>Happiness</td>
<td>81.5</td>
<td>89.1</td>
<td>79.3^a</td>
<td>83.5^f</td>
<td>74.0^a</td>
<td>78.8^d</td>
</tr>
<tr>
<td>Global Function</td>
<td>95.2</td>
<td>96.4</td>
<td>70.9^a</td>
<td>79.0^a</td>
<td>69.0^a</td>
<td>62.8^a</td>
</tr>
</tbody>
</table>

^aANOVA F-test for this GMFCS group comparison was not significant so post-hoc analyses were not considered. Results of ANOVA post-hoc tests for pairwise differences (Scheffe’s least significant difference procedure) among GMFCS level for each parameter are represented by ^b(significantly different for all comparisons with \(p<0.05\)) and ^c(significant difference between the GMFCS levels at \(p<0.05\)).
participants had a mean age 10 years 9 months (standard deviation [SD] 4 years) with a range of 3 years to 20 years 4 months, level II participants had a mean age of 11 years 2 months (SD 4 years 3 months) with a range of 3 years to 20 years 5 months, and level III participants had a mean age of 10 years 7 months (SD 4 years 2 months) with a range of 3 years to 21 years 9 months.

**DESCRIPTIVE STATISTICS, ANALYSIS OF VARIANCE, AND CONFIDENCE INTERVALS**

Descriptive statistics, summarized by GMFCS level, for GMFM sections D and E, PODCI scores, temporal–spatial gait parameters, and oxygen cost are presented in Table II. Table II also contains ANOVA comparisons among GMFCS levels and the 95% confidence intervals for the means at each level of GMFCS. There were significant differences among the levels of GMFCS for many of the outcome tools assessed in this study. In particular, GMFM sections D and E scores, Gait Velocity, Stride Length, Cadence, Oxygen Cost, and PODCI measures of Upper Extremity and Function, Transfer and Mobility, Sports and Physical Function, and Global Function all demonstrated a significant GMFCS level effect, with the exception of the sub-scale of Happiness. These findings are true whether the study participants were assessed as a whole group or per GMFCS level. These data are reported in Table III.

**COMPARISON OF STUDY FINDINGS WITH NORMATIVE DATA**

Results of comparisons between each of the two sets of published normative values for control individuals without disabilities and mean PODCI scores found in this study showed that the study participants scored significantly lower on all PODCI measures of function, with the exception of the sub-scale of Happiness. These findings are true whether the study participants were assessed as a whole group or per GMFCS level. These data are reported in Table III.

**CORRELATIONS BETWEEN GMFCS AND OUTCOME TOOLS**

Spearman’s rank correlations (rho values) were assessed between the GMFCS and outcome tools. Those with statistical significance at a level of $p \leq 0.05$ are reported in Table IV. The strongest correlations were seen between GMFCS level and each of the GMFM sections D and E scores; the PODCI measures of Transfer and Mobility, Sports and Physical Function, Gait Velocity, Stride Length, and Oxygen Cost. There was no significant correlation between the GMFCS level and each of the PODCI measures for Pain and Comfort ($rho = 0.12$, $p = 0.06$), Comorbidity ($rho = 0.07$, $p = 0.2$), Expectations ($rho = -0.09$, $p = 0.2$), and Happiness ($rho = -0.11$, $p = 0.08$). Correlations of $rho \geq 0.5$ or higher are generally considered to indicate moderate to strong relationships.

**CORRELATIONS AMONG OUTCOME TOOLS**

Linear relationships among different parameter scores across tools were also examined using Pearson’s product–moment correlations ($r$ values). The tools with the strongest correlations ($r \geq 0.5$) are reported in Table V. The GMFM sections D and E scores correlated with the largest number of other tools. In addition to the relationships summarized in Table V, gait velocity was significantly related to the PODCI scores of Transfer and Mobility ($r = 0.52$, $p < 0.0001$) and Sports and Function ($r = 0.54$, $p < 0.0001$). Additionally, oxygen cost was related to the PODCI scores of Transfer and Mobility ($r = -0.54$, $p < 0.0001$) and Sports and Function ($r = -0.50$, $p < 0.0001$).

**LOGISTIC REGRESSION ANALYSIS**

The number of participants with complete data on a given measure varied considerably, as discussed previously. To be included in the logistic regression analysis, a participant had to be complete on all 10 variables included in the initial model. A total of 113 participants had complete information on all of the included variables (GMFM sections D and E scores; PODCI scores of Transfer and Mobility, Sports and Physical Function, and Global Function; Gait Velocity, Stride Length, and Cadence; and Oxygen Cost) as well as the GMFCS, and they

---

**Table IV: Significant Spearman’s rank correlations ($p \leq 0.05$) between GMFCS and each outcome tool are reported**

<table>
<thead>
<tr>
<th>Measure</th>
<th>rho</th>
<th>$p$</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GMFM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section D (%)</td>
<td>-0.72</td>
<td>&lt;0.0001</td>
<td>762</td>
</tr>
<tr>
<td>Section E (%)</td>
<td>-0.77</td>
<td>&lt;0.0001</td>
<td>761</td>
</tr>
<tr>
<td><strong>PODCI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Extremity and Function</td>
<td>-0.28</td>
<td>&lt;0.0001</td>
<td>721</td>
</tr>
<tr>
<td>Physical Function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer and Mobility</td>
<td>-0.61</td>
<td>&lt;0.0001</td>
<td>271</td>
</tr>
<tr>
<td>Sports and Physical Function</td>
<td>-0.58</td>
<td>&lt;0.0001</td>
<td>237</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>-0.17</td>
<td>0.0047</td>
<td>263</td>
</tr>
<tr>
<td>Global Function</td>
<td>-0.44</td>
<td>&lt;0.0001</td>
<td>226</td>
</tr>
<tr>
<td><strong>Temporal–spatial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity (%norm)</td>
<td>-0.68</td>
<td>&lt;0.0001</td>
<td>1004</td>
</tr>
<tr>
<td>Stride Length</td>
<td>-0.27</td>
<td>&lt;0.0001</td>
<td>1004</td>
</tr>
<tr>
<td>Cadence (%norm)</td>
<td>-0.52</td>
<td>&lt;0.0001</td>
<td>1004</td>
</tr>
<tr>
<td>Oxygen cost (ml/kg/m)</td>
<td>0.61</td>
<td>&lt;0.0001</td>
<td>419</td>
</tr>
</tbody>
</table>

Relationships between GMFCS and each GMFM, temporal–spatial data, and PODCI are negative as there is an inverse relationship between each of these performance measures and GMFCS level. In the case of oxygen cost, there is a positive relationship between this measure and GMFCS level.

**Table V: Relationships between different outcome tools. Summary of significant Pearson’s product–moment correlations (denoted $r$; all associations significant with $p < 0.0001$) among selected study variables**

<table>
<thead>
<tr>
<th>Measure</th>
<th>GMFM Section D</th>
<th>GMFM Section E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PODCI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer and Mobility</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>Sports and Physical Function</td>
<td>0.65</td>
<td>0.70</td>
</tr>
<tr>
<td>Global Function</td>
<td>0.55</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>Temporal–spatial</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velocity (%norm)</td>
<td>0.64</td>
<td>0.69</td>
</tr>
<tr>
<td>Cadence (%norm)</td>
<td>0.52</td>
<td>0.51</td>
</tr>
<tr>
<td>Oxygen cost (ml/kg/m)</td>
<td>-0.70</td>
<td>-0.68</td>
</tr>
</tbody>
</table>

Ambulatory CP: GMFCS and Outcome Tools D J Oeffinger et al. 315
formed the sample for this analysis. Based on stepwise selection of independent variables in the logistic procedure, only the GMFM section E score entered and remained in the model as a significant predictor of GMFCS level (the $p$ value for the predictor is $<0.0001$). This logistic regression indicated that there was a significant positive relationship between the GMFM section E score and the GMFCS level, such that higher scores on the GMFM section E tools were predictive of lower levels on the GMFCS (i.e. higher levels of functioning).

Due to the large number of missing observations not included in the model, the simple regression of GMFCS level on GMFM section E was considered. The sample size in this case was 761, the significance of GMFM section E as a predictor of GMFCS was the same ($p<0.0001$), and the positive relationship between GMFM section E score and GMFCS was maintained.

**Cluster analysis for predicting GMFCS level**

Cluster analysis was used to determine whether the participants could be grouped according to GMFM section E as a way of discerning functional status. Based on the GMFM section E scores, three clusters of participants were formed: high-functioning, medium-functioning, and low-functioning. Results of the cluster analysis are presented in Table VI. Given the range of GMFM section E scores for each cluster, there is a clear separation of groups. The association between GMFCS level and cluster membership was investigated with a $\chi^2$ test of association, which was significant ($p<0.0001$). The extremely small $p$ value indicates that groupings made based on GMFM section E score are clearly associated with the GMFCS level.

**Discussion**

**Descriptive statistics, analysis of variance, and confidence intervals**

Since outcome tools are still relatively new in the field of orthopaedics, many clinicians are still trying to determine what item scores mean and how to incorporate this information into their clinical practice. This study is the largest to date and only the second multicenter project to investigate outcome tools. It provides clinicians with a large data set to compare with individual data at their local facility and can

<table>
<thead>
<tr>
<th>Cluster descriptor</th>
<th>Low Functioning</th>
<th>Medium Functioning</th>
<th>High Functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in cluster</td>
<td>176</td>
<td>232</td>
<td>353</td>
</tr>
<tr>
<td>GMFM section E, mean (SD)</td>
<td>20 (9.1)</td>
<td>61.6 (10)</td>
<td>90.7 (6.6)</td>
</tr>
<tr>
<td>Range for GMFM section E</td>
<td>0–41</td>
<td>42–76</td>
<td>77–100</td>
</tr>
<tr>
<td>Most likely GMFCS level</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>% in the corresponding GMFCS level correctly classified in the cluster</td>
<td>72.7</td>
<td>68.1</td>
<td>84.7</td>
</tr>
</tbody>
</table>

**Figure 1:** Relationship of cluster membership to GMFCS level ($n=761$) by percentage. Note remaining 0.3% in GMFCS level I fell into the low-functioning group.

- **Low functioning**
- **Medium functioning**
- **High functioning**
offer insight into the scores obtained in their practice. The descriptive statistics and one-way ANOVAs presented in this study provide an overview of how a large group of participants with CP classified as GMFCS levels from I to III, perform on these outcome tools. These data are representative of children with CP from a large geographic portion of the US. While this study is cross-sectional in nature and the data do not denote change over time, they provide an indication of current function. The 95% confidence intervals for each outcome tool per GMFCS level provide an estimate of the range of average values that would be expected in patients with CP similar to the patients included in this study.

Based on the differences found among the group means of the different GMFCS levels, we can conclude that the groups were indeed different from one another. The differences among the GMFCS levels for the functional measures obtained from the GMFM, the PODCI, temporal–spatial gait data, and oxygen cost were all significant and substantiate that these outcome tools successfully measure differences in functional abilities. The study results also illustrate that the GMFCS successfully classifies groups of individuals with different functional levels.

Consistent with the results from a study done by Kennes and colleagues (2002), the three PODCI measures of Comorbidity, Pain and Comfort, and Expectations did not show statistically significant differences among GMFCS levels. The lack of significant differences seen among the GMFCS levels for the PODCI measure of comorbidity is likely to be due to the lack of variability in the data, which may, in turn, be an indication that this measure of function is either truly unrelated to GMFCS level or that the study lacked sufficient power to detect a subtle difference in comorbidity scores among the GMFCS levels. The measure of Expectations is associated with questions related to treatments and is really not appropriate to this cross-sectional design. It is important to note that the PODCI measures of Pain and Comfort, Happiness, Satisfaction, and Expectations have not been validated. These measures do not have discrimination power due to their design, there are either two few or not specific enough questions to measure the domain.

COMPARISON OF STUDY FINDINGS WITH NORMATIVE DATA
The comparisons of the scores obtained on the PODCI in this study with published PODCI normative data (see Table III) for control individuals without disabilities, indicated the participants with CP scored significantly lower on all areas with the exception of Happiness. This indicates that this outcome tool is sensitive enough to discern the decreased level of function in those with CP, relative to the level of functioning measured in normally developing control participants as reported in the literature. These comparisons with published normative values can assist clinicians in interpreting the assessment tool scores obtained by their patients.

CORRELATIONS WITH GMFCS
The primary purpose of this study was to investigate the relationship between the level of condition severity of children with ambulatory CP, defined by GMFCS level, and various outcome tools. Before this study, Daltroy and coworkers (1998) investigated the relationship between level of severity and the PODCI to establish validity of the PODCI tool. Abel and colleagues in a study (2003) presented at the Pediatric Orthopaedic Society of North America in 1999 reported that perception of functional severity showed fair to good correlation with the GMFM and that these scores were consistent with levels of severity. Both Daltroy and Abel and their respective colleagues used the term ‘level of severity’ to refer to the differences in scores between children with hemiplegic CP and diplegic CP. Previous studies also investigated the validity aspect of the outcome tools, but did not address the specific relationship between a standardized classification of severity in CP, the GMFCS, to functional measures of GMFM, PODCI scores, temporal–spatial gait parameters, and oxygen cost.

Based on the findings from the present study, moderate to strong relationships (rho≥0.50) do exist between the GMFCS level and scores obtained on outcome measures used to assess functional abilities. These newly established associations provide evidence that the assessed outcome tools can distinguish between children with different levels of CP and clinicians should feel comfortable using these tools to assess functional levels in children with ambulatory CP.

The PODCI measure of Global Function is an average score of the four PODCI measures of Upper Extremity and Function, Transfer and Mobility, Sports and Physical Function, and Pain and Comfort. The correlation of this measure with GMFCS level (rho=–0.44) reflects the effect of combining measures that have strong associations with GMFCS (Transfer and Mobility, Sports and Physical Function) with those that have weaker associations (Upper Extremity and Physical Function, Pain and Comfort) resulting in the Global Function variable only moderately related to GMFCS. Although the PODCI Upper Extremity and Physical Function measure does assess a functional aspect of performance and thus might be expected to correlate strongly with GMFCS, the correlation is modest (rho=–0.28). This could be explained by the fact that there is limited emphasis placed on upper extremity function in the GMFCS guidelines.

The magnitude of Spearman’s rank correlation is a function of how well the ordered rankings of the two variables agree. In this study, there are a few correlations that are relatively small in magnitude but are still statistically significant. This is likely due to relatively large sample sizes for many of the comparisons. The most notable example of a measure that has a low degree of association with GMFCS level is the PODCI measure of Satisfaction (rho=–0.17, p<0.005). The modest level of correlation is consistent with what would be expected for this relationship between measures of physical function and measures of mental well-being.

The low (and non-significant) correlations reported between the GMFCS and each of the PODCI measures of Pain and Comfort, Expectations, and Happiness should be interpreted with caution due to the nature of these measures. As previously stated, these measures have not been validated and do not have discrimination power because of their design. The non-significant correlation between the GMFCS and PODCI measure of comorbidity is likely due to the lack of spread in the scores across participants. Based on these findings, the PODCI should not be used to determine functional differences based on these measures.

CORRELATION AMONG OUTCOME TOOLS
A secondary purpose of this study was to examine the relationships among the different outcome tools. Previous work in the area by other investigators examined relationships among outcome measures in pediatric samples. Abel and colleagues
presented data in 1999 at the Pediatric Society of North America meeting that examined the correlation between the PODCI Global Function score and the GMFM and temporal–spatial gait data. They found a correlation between PODCI global scores and the GMFM ($r=0.70$) and between the PODCI and gait velocity ($r=0.40$) and stride length ($r=0.40$). When Damiano and Abel (Damiano and Abel, 1996) examined the relationship between the GMFM and gait analysis, they reported correlations ranging from $r=0.60$ to $0.79$ between the GMFM and temporal–spatial parameters. Drouin and colleagues (Drouin et al. 1996) looked at the relationship between gait velocity and and the GMFM. For the total group of children, significant linear relationships were obtained between gait velocity and the GMFM sections D and E scores ($r=0.91$, $r=0.93$ respectively; $p<0.0001$). McCarthy and coworkers (2002) examined the correlations between the PODCI and GMFM as part of a larger study evaluating the reliability and validity of pediatric tools for measuring health and well-being of children with spastic CP. They reported $r$ values of 0.88 and 0.80 for the correlation between the GMFM and the PODCI measures of Transfer and Mobility and Upper Extremity Involvement respectively.

These previous studies primarily used correlations to establish reliability and validity of the outcome tools. The present study investigated correlations among the different outcome measures with the focus of examining what relationships existed among the various types of functional assessments, and to what degree. The data previously reported were based on sample sizes of 30 to 115 participants, while sample sizes in this study range from 226 to 1047 participants. The correlations found between the GMFM sections D and E scores and other outcome measures show that a clinician-observed assessment of function is related to the technical measures of gait velocity and oxygen cost, as well as to the parent report of Transfer and Mobility, Sports and Physical Function, and Global Function as measured by the PODCI.

**Predicting GMFCS Level using Outcome Tools**

The present study examined whether any of the outcome variables could predict the GMFCS level of a child with ambulatory CP. The GMFM section E score was found to be the best (and only significant) predictor of the child’s GMFCS level. The clustering showed a separation of participants into groups based on the GMFM section E score that is very consistent with GMFCS assignment. Of the 761 participants who had complete information on both GMFM section E and GMFCS, 584 were classified into clusters that agreed exactly with their GMFCS level. Based on these findings, we can conclude that the GMFM section E score can be used to predict GMFCS level relatively accurately (with an overall correct classification rate of 76.6%). Using these data, an expected range of GMFM section E scores for each GMFCS level was established. These ranges can be used in the clinical setting to provide a good approximation of a child’s GMFCS level when the GMFM section E score is known. One application for this new information is to assist clinicians in grouping patients for retrospective studies when the GMFM section E score is known but the GMFCS level is not.

As it is more difficult to distinguish between a child who is at level I versus a child who is at level II (Palisano et al. 1997) it is not surprising that the lowest degree of agreement between the two methods of classification was for the middle group of individuals. Only 68.1% of those with a GMFCS level II were placed in the middle-functioning group by the clustering procedure. For the 65 participants in the GMFCS level II group who were not included in the middle cluster of functioning, the likelihood of being classified at the high level (GMFCS III) was four times the likelihood of being classified at the low level (GMFCS I) of functioning. In all, for the 177 participants (25.2% of all participants) whose cluster membership did not agree with their GMFCS level, 113 of them (63.8% of these 177 participants) were classified at a higher GMFCS level than would be suggested by the GMFM section E score.

**Limitations**

This study is not without limitations. The primary limitation is that as the data were collected retrospectively from multiple sites, we were unable to control that all participants completed all four evaluations. Therefore, the data reported do not represent the same sample for each tool. This is more of a limitation when comparing across tools than between the GMFCS and the outcome tools. Another limitation due to the retrospective design is the inability to determine the statistical power of the tests used.

Another potential limitation is the assumption that all tests were administered in a standardized manner. We believe that the standardized training of clinicians administering the GMFM and the lack of sensitivity of the temporal–spatial gait parameters to clinician variability, controls for this limitation. This is substantiated by data obtained from a standardization assessment performed across 12 different Shriners Hospital motion analysis laboratories. As reported by Gorton and co-investigators (2001) at the Gait and Clinical Movement Analysis Meeting, a 1.1% coefficient of variation was found for gait velocity, stride length, and cadence on the same normally developing individual, across 24 clinicians and 12 laboratories. We interviewed the investigators at each site to determine if there were any procedural differences that could impact the data. Based on these conversations, we are confident that the data collection procedures were similar across sites. However, the overlap seen in the data between GMFCS levels I and II may reflect the lack of standard agreement in this retrospective study and with more attention to interinstitutional standardization, agreement could be better.

This study looked at children with CP who were ambulatory and so considered only those functioning at GMFCS levels I to III. Our prediction that those with GMFM section E scores of 0 to 41 be classified as GMFCS level III is based on the a priori knowledge that this group was ambulatory. In general, those with GMFM scores in the low function range may also be non-ambulatory, and thus may function at lower than GMFCS level III, but that population was not considered in this study. Thus, there are two things that need to be known to use the prediction: (1) the score and (2) that they are ambulatory.

One additional limitation is the use of 95% confidence intervals of the means to quantify the variability of the average score for each assessment tool at each GMFCS level. A more clinically useful technique would be to develop intervals that contain the majority of individual scores (as opposed to sample means) with a stated level of confidence. Ideally, we would like to be able to predict a range of scores for each tool per GMFCS level. In a prospective study that is currently in progress, prediction models will be developed in an attempt to create these confidence intervals for the individual scores. Such intervals could be clinically useful in predicting future physical functioning in children with CP.
Conclusion
The information gained from this study is applicable to both clinician, and families. The study results can help clinicians have a better understanding of the scores obtained on the outcome tools currently used in their practices. This improved understanding is reflected in the care provided to the patients and families.

The findings across GMFCS levels reported in this study have implications for decision making and interpretation of intervention outcomes in children with ambulatory CP. These data also establish justification for using the GMFCS as a classification system in clinical studies requiring grouping of similar functional abilities.

Future work currently underway includes both cross-sectional and longitudinal observations for the same individuals across all tools and will expand to include the Functional Assessment Questionnaire (Novacheck 2001) and the Pediatric Quality of Life measure (Varni 2002).

References


List of abbreviations
GMFCS: Gross Motor Function Classification System
GMFM: Gross Motor Function Measure
ICF: International Classification of Functioning, Disability and Health
NCMRR: National Center for Medical Rehabilitation Research
PEDI: Pediatric Evaluation of Disability Index
PODCI: Pediatric Outcomes Data Collection Instrument