



Published in final edited form as:

Dev Med Child Neurol. 2008 December ; 50(12): 918–925. doi:10.1111/j.1469-8749.2008.03150.x.

Outcome tools used for ambulatory children with cerebral palsy: responsiveness and minimum clinically important differences

D Oeffinger, PhD*,
Shriners Hospital for Children (SHC), Lexington, KY

A Bagley, PhD, SHC,
Sacramento, CA

S Rogers, MPH, SHC,
Lexington, KY

G Gorton, BS, SHC,
Springfield, MA

R Kryscio, PhD,
University of Kentucky, Lexington, KY

M Abel, MD,
University of Virginia, Charlottesville, VA

D Damiano, PhD, PT,
Washington University, St. Louis, MO

D Barnes, MD, SHC, and
Houston, TX

C Tylkowski, MD, SHC
Lexington, KY, USA

Abstract

This prospective longitudinal multicenter study of ambulatory children with cerebral palsy (CP) examined changes in outcome tool score over time, tool responsiveness, and used a systematic method for defining minimum clinically important differences (MCIDs). Three hundred and eighty-one participants with CP (Gross Motor Function Classification System [GMFCS] Levels I–III; age range 4–18y, mean age 11y [SD 4y 4mo]; 265 diplegia, 116 hemiplegia; 230 males, 151 females). At baseline and follow-up at least 1 year later, Functional Assessment Questionnaire, Gross Motor Function Measure, Pediatric Quality of Life Inventory, Pediatric Outcomes Data Collection Instrument, Pediatric Functional Independence Measure, temporal–spatial gait parameters, and oxygen cost were collected. Adjusted standardized response means determined tool responsiveness for nonsurgical ($n=292$) and surgical ($n=87$) groups at GMFCS Levels I to III. Most scores reaching medium or large effect sizes were for GMFCS Level III. Nonsurgical group change scores were used to calculate MCID thresholds for ambulatory children with CP. These values were verified by examining participants who changed GMFCS levels. Tools measuring function were responsive when a change large enough to cause a change in GMFCS level occurred. MCID thresholds assess change in study populations over time, and serve as the basis for designing prospective intervention studies.

*Correspondence to first author at Shriners Hospital for Children, 1900 Richmond Road, Lexington, KY 40502, USA. doeffinger@shrinenet.org.

Assessing physical function and quality of life using valid and reliable outcome tools is important for clinical care of individuals with cerebral palsy (CP). One crucial attribute of outcome tools is responsiveness, which establishes the ability to detect change and remain constant with no change.¹⁻⁴ Standardized response mean (SRM)⁵⁻¹³ and effect size¹⁴ have been used to establish responsiveness. Responsiveness of the Pediatric Functional Independence Measure (WeeFIM), Pediatric Outcomes Data Collection Instrument (PODCI), and Gross Motor Function Measure (GMFM-88; GMFM-66) have been reported.^{7,15,16} The responsiveness of other outcome tools commonly used to assess individuals with CP needs to be established.

It is important to understand the minimum clinically important difference (MCID) for outcome tools. MCID is a threshold for determining when meaningful changes occur. MCIDs have been defined and calculated in various ways.^{7,15,17-19} Reported changes frequently reach statistical significance but may not be clinically meaningful.^{20,21} For this study, MCID is the magnitude of change required for an observable difference in function, and is quantified using effect sizes. Small effect sizes may be described as imperceptible to the human eye, medium as being large enough to be seen in normal observation, and large as grossly observable.³

MCIDs for the Pediatric Quality of Life Inventory (Peds-QL),¹⁹ WeeFIM,⁷ and GMFM-88 and -6615 have been reported. Techniques used to obtain these thresholds varied and were obtained from different populations. No MCIDs have been established for the PODCI, temporal-spatial gait parameters, and energy cost during walking (O₂ cost).

This study determined how outcome measures change over 1 year in ambulatory children with CP, with and without surgical intervention. Responsiveness was assessed using SRM and a known change in function defined as a change in Gross Motor Function Classification System (GMFCS) level between assessments. MCIDs were established for the GMFM, PODCI, PedsQL, WeeFIM, O₂ cost, and temporal-spatial gait parameters for ambulatory children with CP. The results provide critical information for designing intervention studies and interpreting results.

Method

This 6-year prospective multicenter study with cross-sectional and longitudinal components was conducted at seven pediatric orthopedic facilities located across the USA (California, Kentucky, Massachusetts, Missouri, Texas, Utah, Virginia) that each treat children from several surrounding states. Institutional Review Board approval was obtained at each site and all participants signed consent, assent as appropriate, and privacy and confidentiality forms. A complete description of the methods was previously reported.²²

TOOLS AND STUDY METHODOLOGY

GMFM Dimensions D (standing) and E (walking, running, jumping), Parent and Child PedsQL, Parent and Child PODCI,²³ FAQ,²⁴ WeeFIM, O₂ cost, temporal-spatial gait parameters, and GMFCS level were collected at baseline and follow-up at least 1 year later. Before the start of the study, local coordinators were trained in GMFCS classification, tool administration, and data collection procedures. Consistency among coordinators was verified. Data were collected into a study-specific database by direct computer entry and reviewed by the project manager for completeness and accuracy.

PARTICIPANTS

Participants were recruited and enrolled for both the cross-sectional and longitudinal study phases. Inclusion criteria were: diagnosis of CP, GMFCS Levels I to III, ages 4 to 18 years,

and ability to complete a gait evaluation. Exclusion criteria were: previous selective dorsal rhizotomy, lower extremity orthopedic surgery within a year, botulinum toxin A injections within 6 months, or a currently operating Baclofen pump.

Of 562 participants who completed baseline assessments, 387 completed the follow-up evaluation (68.7%). Of those who did not complete the follow-up evaluation, 95 were unable to be contacted (16.9%), 29 declined (5.2%), nine did not attend for their follow-up (1.6%), four were no longer ambulatory (0.7%), two had surgery less than 1 year from baseline at study completion (0.4%), and 36 for other reasons (6.4%). Six (1.6%) were excluded from analysis because of incomplete data, resulting in a final sample of 381. There were no differences at baseline between those who completed the follow-up and the 175 who did not for age, height, weight, type of involvement, sex, GMFCS level, birth history, and ethnicity. Of the 381 participants, there were 230 (60%) males, 151 (40%) females; 174 (46%) GMFCS Level I, 132 (34%) Level II, and 75 (20%) Level III; 265 (69%) diplegic and 116 (31%) hemiplegic; and predominately Caucasian (83%). Mean age at baseline was 11 years (SD 4y 3mo, range 4y 3mo–18y 4mo) and at follow-up was 12 years 5 months (SD 3y 3mo, range 5y 2mo–20y 6mo). The mean time between assessments was 1 year 5 months (SD 5mo). Demographics are reported in Table I. Individual treatment plans prescribed by the participants' physicians were followed between assessments; 87 participants (23%) had orthopedic surgery during the study period.

ANALYSIS

Primary analysis—The study sample was separated by GMFCS level and by those with surgical intervention between assessments ($n=87$; GMFCS I=32, II=35, III=20) and those without ($n=292$; GMFCS I=141, II=96, III=55). The nonsurgical group represents changes over time with standard care excluding orthopedic surgery. The surgical group represents changes over time due to orthopedic surgery, which were expected to exceed the nonsurgical group's changes.

Descriptive data—Change scores were calculated for all tools by GMFCS level as follow-up minus baseline score. Mean and SD of the outcome tools' change scores were calculated for the surgical and nonsurgical groups.

SRM—Effect sizes are a standardized unit-free measure, calculated as: mean/SD.⁴ Guidelines for interpreting effect size magnitudes were introduced by Cohen²⁵ as: 0.2 small, 0.5 medium, and greater than 0.8 large. The use of Cohen's effect sizes is based on the assumption that the study groups are independent, with equal samples size and common within-population SD.²⁶

The effect size for repeated measures of independent samples was renamed the standardized response mean (SRM) by Liang et al.¹² and is stated as (Formula 1):

$$\text{SRM}(\text{independent samples}) = \text{mean change score} / \text{SD of change score.}$$

To apply the SRM to dependent samples, one must account for pooled SDs and correlations between measures. The SRM for dependent samples is referred to as the adjusted standardized response mean (SRM_a). The factors of $\sqrt{2}$ and $\sqrt{(1-r)}$, where r is the correlation between repeated measures, are added to the equation to account for pooled samples and correlations respectively.²⁶ The resulting equation for SRM_a is (Formula 2):

$$\text{SRM}_a(\text{dependent samples}) = \text{SRM}(\text{independent samples}) / \sqrt{2} / \sqrt{(1-r)}$$

MCID—MCID²⁷ may be subjectively defined based on patient or family rating, or by clinicians from training and personal experience. Some investigators have proposed objective techniques to define MCID.^{19,26}

The MCIDs of study tools were calculated based on non-surgical group data. This non-operative treatment group was used to estimate the population SD for change and the correlation between baseline and follow-up scores.

The MCID is based on the assumption that the mean change score needed to obtain a medium or large effect size is clinically meaningful. This was substantiated by Portney and Watkins,³ who stated that medium effect sizes are observable and large effect sizes are grossly observable. Based on this assumption, the MCID equation was derived as follows (Formula 3):

$$\text{Mean change score} = \text{effect size (independent samples)} \times \text{SD of change score.}$$

Substituting MCID as the mean change score (Formula 4):

$$\text{MCID} = \text{effect size (independent samples)} \times \text{SD of change score}$$

where any desired effect size can be substituted into the equation.

Because the MCID equation is based on independent samples, equation (Formula 2) was solved for SRM (independent samples; Formula 5):

$$\text{SRM (independent samples)} = \text{SRM}_a \text{ (independent samples)} \times \sqrt{2} \times \sqrt{(1 - r)}$$

where SRM_a is the desired effect size.

After substitutions, the final MCID equation is (Formula 6):

$$\text{MCID} = \text{desired effect size (dependent samples)} \times \sqrt{2} \times \sqrt{(1 - r)} \times \text{SD of change score}$$

From here, 0.5 was substituted for medium (observable) effect size and 0.8 for large (grossly observable) effect size. An example calculation for the GMFM-66 MCID at GMFCS Level I is:

$$\text{MCID for GMFM - 66 at GMFCS Level I} = 0.05 \times \sqrt{2} \times \sqrt{(1 - 0.81)} \times 5.4 = 1.7$$

Secondary responsiveness analysis—Secondary analysis to test the responsiveness of study tools and newly established MCID was completed based on individuals who changed GMFCS levels between study assessments. The GMFCS is a standardized classification system designed to reflect differences in gross motor function that are meaningful in the daily lives of children with CP and their families²⁸ and are clinically meaningful.²⁹ The GMFCS has proven stable over time; 73% children remained in the same level over time and 87.5% were classified in the same level as previous visits, with 11.7% reclassified by one GMFCS level.²⁸ Reclassification to a higher level of ability is valid for children who demonstrate considerable improvement in gross motor function.²⁸ The GMFCS has also been highly correlated (−0.91) to the functional capacity measure of

GMFM-88.³⁰ Previous work by the Functional Assessment Research Group demonstrated clear functional differences among GMFCS Levels I to III on standardized outcome tools.²²

GMFCS levels at baseline and follow-up were verified for 377 of the 381 participants. Thirty-one (8%) individuals had changes in function, defined as a verified change in GMFCS level between assessments, and were placed into an Improved ($n=18$) or Declined ($n=13$) group, and those who stayed the same in a No Change group ($n=346$). Child POD-CI and PedsQL responses were excluded because of small sample sizes. t -tests determined if change scores were significantly different from zero ($p \leq 0.05$). Comparisons among groups were made using analysis of variance (ANOVA) with post-hoc Tukey tests for results $p \leq 0.05$. Change scores were compared with MCID values derived from the nonsurgical group ($n=292$).

Results

PRIMARY ANALYSIS

Descriptive—Table II reports the means and SDs of change scores for nonsurgical and surgical groups by GMFCS level. SDs of change were large compared with magnitude of mean changes.

Standardized response mean (SRM_a)—The values of SRM_a of the nonsurgical and surgical groups by GMFCS level are reported in Table II. For the GMFCS Level I nonsurgical group, the only score exceeding a medium effect size was Parent PODCI Upper Extremity. No tools achieved medium effect size for GMFCS Level II. For the GMFCS Level III non-surgical group, GMFM Dimension E achieved a medium effect size and Dimension D a large effect size. For the GMFCS Level I surgical group, Parent PODCI Global Function and Parent PODCI Upper Extremity achieved a medium effect size and Parent PODCI Transfers achieved a large effect size. For the GMFCS Level II surgical group, only WeeFIM Self Care reached a medium effect size. For the GMFCS Level III surgical group, six outcome scores reached a medium effect size: GMFM-66, Parent PedsQL School Functioning, Parent PODCI Global Function, Parent PODCI Comfort/Pain, Parent PODCI Satisfaction, and Parent PODCI Transfers. For the GMFCS Level III surgical group, Parent PedsQL Social Functioning and WeeFIM Cognition reached a large effect size.

Minimal clinical important difference—Minimum change scores needed for a MCID on the outcome tools at medium and large effect sizes are reported (Table III). Change scores exceeding MCIDs for a medium effect size are shown in bold type in Table II.

SECONDARY RESPONSIVENESS ANALYSIS

In the Improved group, 12 participants changed from GMFCS Level II to I and six from III to II. In the Declined group, five changed from GMFCS Level I to II and eight from II to III. In the Improved group, 9 of 18 had surgery between assessments. In the Declined group, 6 of 13 had surgery between assessments. In the No Change group, 72 of the 274 had surgery between assessments.

Change scores significantly different from zero ($p < 0.05$) are reported in Table IV. Those with significant differences ($p < 0.05$) between groups (No Change vs Declined, Improved vs Declined) are highlighted. With the exception of PODCI Satisfaction, all change scores for the Declined group in Table IV exceeded MCID for a medium effect size, with six of ten exceeding the thresholds for large effect size. For the Improved group, four of ten exceeded MCID for a medium effect size, and three exceeded MCID for large effect sizes. For the No

Change group, change scores did not exceed the medium effect size for any tool. Overall, scores increased for those with an improved GMFCS level, decreased for those with a declined level, and varied little for those with no change. Change scores were larger in the Declined group than the Improved group.

Discussion

This study sought to evaluate change and responsiveness of outcome tools assessing ambulatory children with CP over time. Although the tools are valid and reliable,^{23,24,31–33} assessments of responsiveness have been limited to intervention studies, small sample sizes, and a limited number of tools.^{7,15,16} As MCID has not been established for many tools, interpretation is often in the context of anecdotal experience. A systematic method of defining MCIDs for ambulatory children with CP is presented in this study.

The responsiveness of outcome tools is difficult to assess because changes must be compared with a known change in function. However, no criterion standard of change exists. In the absence of a known change, responsiveness has been assessed using statistical techniques including SRM^{5–8,10–12} in studies of childhood disabilities, including CP.⁷

The magnitude of the SRM_a is affected by both the change scores and the variability of the changes. Heterogeneity in diagnosis and severity level in CP results in highly variable baseline and change scores. In this study, inclusion criteria were broad and no constraints were placed on participants during the study duration, resulting in a diverse sample. Owing to high variability, large change scores are needed to obtain large SRM_a values.

The natural history of children with CP is characterized by a gradual deterioration of gait function over 1 year 6 months to 5 years.^{34–36} Therefore, minimal changes were expected in a 17-month period for the nonsurgical group. This was supported by the nonsurgical group's trivial to small SRM_a values at all GMFCS levels. Orthopedic surgery is expected to maintain or improve function.^{21,36} Surgical intervention varied from isolated releases to multilevel bony and soft-tissue procedures. In the surgical group, SRM_a values were at trivial or small levels for most tools and did not demonstrate change for all GMFCS levels. Most SRM_a values that reached medium and large effect sizes were noted for GMFCS Level III. Small SRM_a values may be due to maintenance of function or the surgical group's variability of change scores. Other factors that may contribute to small SRM_a values for both groups in this study are limited tool resolution, no change occurring, or the tools not measuring the appropriate factors of change.

The transition from SRM_a to the calculation of MCID is a novel approach. MCID has been reported for PedsQL,¹⁹ Wee-FIM,⁷ and GMFM-66.^{15,18} The MCIDs in this study are higher than previously reported,^{7,15,19} probably because of the larger sample size with greater variability in change data. Based on the investigators' clinical expertise, the calculated MCIDs are consistent with anecdotal reports. MCID based on personal experience and observation may not appreciate the variability of the change data and so may be overly optimistic.

MCIDs established in this study are based on longitudinal data from a large sample of ambulatory children with CP, from seven geographic regions, and with no surgical intervention during the study period. This is more generalizable than previously reported MCIDs, which were based on a variety of diagnostic populations, from single geographic areas, and primarily collected cross-sectionally.^{7,15,19} These data define expected change over one year for other studies with the same population. If changes exceed the reported MCIDs, a clinically meaningful change is likely to have occurred.

To establish tool responsiveness further, children who changed GMFCS levels between study assessments were identified and change scores analyzed. The GMFCS is a widely accepted, valid, and stable tool for classifying severity level of children with CP ($\kappa=0.75$; children younger than 2y).^{28,29} In this study, outcome tools were responsive when compared with improvements and declines in function. Changes in functional subscales were most consistent with changes in GMFCS level.

Larger change scores in the Declined group may reflect the subtle functional differences between GMFCS Levels I and II, compared with II and III. In the Improved group, most changed from II to I whereas those in the Declined group changed from Level II to III. Distinctions between Levels I and II are that children in level II have limitations in the ease of performing movement transitions, walking outdoors and in the community, and the ability to perform gross motor skills such as running and jumping.²⁹ For Levels II and III, children in the latter need assistive devices to walk, whereas children in the former do not require them after age 4.²⁹

This study provides clinicians and researchers with data that can directly apply to clinical practice. Because data for all tools were collected on the same population, comparisons can be made between change scores on different tools. The results apply to changes in group means; therefore clinicians should not interpret the results at the individual level. MCID thresholds can assist clinicians and researchers in moving beyond interpreting findings based strictly on statistical significance. Effect sizes are a useful statistical technique because they are operationally defined, independent of the sample size, and unit-free.³⁷ However, they can be difficult to translate into the clinical setting. The conversion of an effect size (SRM_a) to an MCID makes the measure more useful because they are in units of the tool score. MCID thresholds for both medium and large effect sizes were reported based on Cohen's operational definition, where a medium effect size is of moderate clinical importance and a large effect size is of crucial clinical importance.³⁷ If a change score exceeds MCID, it is likely that change is of clinical importance. For researchers, the results of this study can: help determine if statistically significant results are also clinically relevant; be used for power calculations; and assist in tool selection for future studies.

LIMITATIONS

GMFCS level was used as the measure of change in function because there is no criterion standard for functional change. Examining differences between Improved, Declined, and No Change groups was limited because of small sample sizes in the Improved and Declined groups. Small sample sizes in the Improved group may have been because children in GMFCS Level I cannot, by definition, improve.

SRM_a and MCID values are based on data from the non-surgical group, including individuals of different ages. The SRM_a and MCID values may differ based on age; however, age was not a significant predictor of changes in outcome scores from other analyses of this study.

The short-term follow-up of the study participants may have limited the magnitude of changes seen for either non-surgical or surgical treatments. Although large effect sizes are more likely to be clinically relevant than small ones, this is not always true.³⁸ Ideally, clinical relevance would also include an external standard.

Conclusion

This study refined the understanding of how ambulatory children with CP function, by examining change scores over time in a large sample. Outcome tools demonstrated few

changes beyond a small effect size for either the surgical or nonsurgical group. Tools were responsive when a change in function occurred large enough to cause a change in GMFCS level. A systematic method of defining MCID was established using the variability of the change scores in the nonsurgical group. These threshold values can be used to assess the change in study populations over time. They can also serve as the basis for designing prospective intervention studies.

Acknowledgments

This work was funded by Shriners Hospitals for Children, Clinical Outcomes Study Advisory Board Grant no. 9140 'A cross-sectional and longitudinal assessment of outcome instruments in patients with ambulatory cerebral palsy'. The authors acknowledge the following investigators for their contribution to this article and the Functional Assessment Research Group (FARG) project: Judi Linton MS PT (SHC Houston, TX); Elroy Sullivan PhD (SHC Houston, TX); Mark Romness MD (UVA); and Diane Nicholson PhD PT (SHC Salt Lake City, UT). The authors also acknowledge from each site the participation of the FARG study coordinators, and the patients and their families.

List of abbreviations

FAQ	Functional Assessment Questionnaire
GMFCS	Gross Motor Function Classification System
GMFM	Gross Motor Function Measure
GMFM-66	Gross Motor Function Measure Score calculated using the Gross Motor Ability Estimator Program
ICF	International Classification of Functioning, Disability and Health
MCID	Minimum clinically important difference
O₂ cost	Energy (oxygen) cost during walking
PedsQL	Pediatric Quality of Life Inventory
PODCI	Pediatric Outcomes Data Collection Instrument
SRM	Standardized response mean
SRM_a	Adjusted standardized response mean
WeeFIM	Pediatric Functional Independence Measure

References

1. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27–36. [PubMed: 3972947]
2. Garratt A, Hutchinson A, Russell I. Network for Evidence-Based Practice in Northern and Yorkshire (NEBPINY). The UK version of the Seattle Angina Questionnaire (SAQ-UK): reliability, validity and responsiveness. *J Clin Epidemiol* 2001;54:907–15. [PubMed: 11520650]
3. Portney, L.; Watkins, MP. *Foundations of Clinical Research: Applications to Practice*. 2. Upper Saddle River, NJ: Prentice Hall Health; 2000.
4. De Bruin AFDJ, De Witte LP, Stevens FC, Philipsen H. Assessing the responsiveness of a functional status measure: the Sickness Impact Profile versus the SIP68. *J Clin Epidemiol* 1997;50:529–40. [PubMed: 9180645]
5. De Smet L. Responsiveness of the DASH score in surgically treated basal joint arthritis of the thumb: preliminary results. *Clin Rheumatol* 2004;23:223–24. [PubMed: 15168149]
6. MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. *J Hand Ther* 2004;17:18–23. [PubMed: 14770134]

7. Ottenbacher KJ, Msall ME, Lyon N, et al. The WeeFIM instrument: its utility in detecting change in children with developmental disabilities. *Arch Phys Med Rehabil* 2000;81:1317–26. [PubMed: 11030496]
8. Salaffi F, Stancati A, Neri R, Grassi W, Bombardieri S. Measuring functional disability in early rheumatoid arthritis: the validity, reliability and responsiveness of the Recent-Onset Arthritis Disability (ROAD) index. *Clin Exp Rheumatol* 2005;23(Suppl 39):S31–42. [PubMed: 16273783]
9. Wallace D, Duncan PW, Lai SM. Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *J Clin Epidemiol* 2002;55:922–28. [PubMed: 12393081]
10. Williams G, Robertson V, Greenwood K, Goldie P, Morris ME. The concurrent validity and responsiveness of the high-level mobility assessment tool for measuring the mobility limitations of people with traumatic brain injury. *Arch Phys Med Rehabil* 2006;87:437–42. [PubMed: 16500181]
11. Goldie P, Matyas T, Evans O. Deficit and change in gait velocity during rehabilitation after stroke. *Arch Phys Med Rehabil* 1996;77:1074–82. [PubMed: 8857890]
12. Liang M, Fossel A, Larson M. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632–43. [PubMed: 2366602]
13. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369–78. [PubMed: 7490600]
14. Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. *Med Care* 1989;27:S178–89. [PubMed: 2646488]
15. Wang HY, Yang YH. Evaluating the responsiveness of 2 versions of the gross motor function measure for children with cerebral palsy. *Arch Phys Med Rehabil* 2006;87:51–56. [PubMed: 16401438]
16. Damiano DL, Gilgannon MD, Abel MF. Responsiveness and uniqueness of the pediatric outcomes data collection instrument compared to the gross motor function measure for measuring orthopaedic and neurosurgical outcomes in cerebral palsy. *J Pediatr Orthop* 2005;25:641–45. [PubMed: 16199947]
17. Bowen TR, Lennon N, Castagno P, Miller F, Richards J. Variability of energy-consumption measures in children with cerebral palsy. *J Pediatr Orthop* 1998;18:738–42. [PubMed: 9821128]
18. Russell, D.; Rosenbaum, P.; Avery, L.; Lane, M. Clinics in Developmental Medicine No. 159. London: Mac Keith Press; 2002. Gross Motor Function Measure (GMFM-66 & GMFM-88) User's Manual.
19. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr* 2003;3:329–41. [PubMed: 14616041]
20. Carney BT, Oeffinger D, Gove NK. Sagittal knee kinematics after rectus femoris transfer without hamstring lengthening. *J Pediatr Orthop* 2006;26:265–67. [PubMed: 16557147]
21. Gage JR, Fabian D, Hicks R, Tashman S. Pre- and postoperative gait analysis in patients with spastic diplegia: a preliminary report. *J Pediatr Orthop* 1984;4:715–25. [PubMed: 6511900]
22. Oeffinger D, Tylkowski C, Gorton G, et al. Outcome tools commonly used in ambulatory children with cerebral palsy: part I – descriptive characteristics. *Dev Med Child Neurol* 2007;49:172–80. [PubMed: 17355472]
23. Daltroy LH, Liang MH, Fossel AH, Goldberg MJ. The POSNA pediatric musculoskeletal functional health questionnaire: report on reliability, validity, and sensitivity to change. Pediatric Outcomes Instrument Development Group. Pediatric Orthopaedic Society of North America. *J Pediatr Orthop* 1998;18:561–71. [PubMed: 9746401]
24. Novacheck TF, Stout JL, Tervo R. Reliability and validity of the Gillette Functional Assessment Questionnaire as an outcome measure in children with walking disabilities. *J Pediatr Orthop* 2000;20:75–81. [PubMed: 10641694]
25. Cohen, J. Statistical power analysis for the behavioral sciences. 2. Manwah, NJ: Lawrence Erlbaum; 1988.
26. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating

- magnitude of intervention-related change in health services research. *Int J Integr Care* 2002;2:e15. [PubMed: 16896390]
27. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–78. [PubMed: 10335749]
 28. Palisano RJ, Cameron D, Rosenbaum PL, Walter SD, Russell D. Stability of the Gross Motor Function Classification System. *Dev Med Child Neurol* 2006;48:424–28. [PubMed: 16700931]
 29. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;39:214–23. [PubMed: 9183258]
 30. Palisano RJ, Hanna SE, Rosenbaum PL, et al. Validation of a model of gross motor function for children with cerebral palsy. *Phys Ther* 2000;80:974–85. [PubMed: 11002433]
 31. McCabe MA, Granger CV. Content validity of a pediatric functional independence measure. *Appl Nurs Res* 1990;3:120–22. [PubMed: 2400209]
 32. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther* 2000;80:873–85. [PubMed: 10960935]
 33. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. *Med Care* 1999;37:126–39. [PubMed: 10024117]
 34. Bell KJ, Ounpuu S, DeLuca PA, Romness MJ. Natural progression of gait in children with cerebral palsy. *J Pediatr Orthop* 2002;22:677–82. [PubMed: 12198474]
 35. Johnson D, Damiano D, Abel M. The evolution of gait in childhood and adolescent cerebral palsy. *J Pediatr Orthop* 1997;17:392–96. [PubMed: 9150031]
 36. Gough M, Eve LC, Robinson RO, Shortland AP. Short-term outcome of multilevel surgical intervention in spastic diplegic cerebral palsy compared with the natural history. *Dev Med Child Neurol* 2004;46:91–97. [PubMed: 14974633]
 37. Hojat M, Xu G. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ Theory Pract* 2004;9:241–49. [PubMed: 15316274]
 38. Jacobson N, Truax P. Clinical Significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psych* 1991;59:12–19.

Table I

Longitudinal participant demographics ($n=381$), overall and by Gross Motor Function Classification System (GMFCS) level^a

Demographic category	Overall, n (%)	GMFCS Level I, n (%)	GMFCS Level II, n (%)	GMFCS Level III, n (%)
Sex				
Male	230 (60)	114 (66)	74 (56)	42 (55)
Female	151 (40)	60 (34)	58 (44)	33 (45)
Total	381 (100)	174 (100)	132 (100)	75 (100)
Race/ethnicity				
African-American	27 (7)	12 (7)	8 (6)	7 (9)
Caucasian	315 (83)	150 (86)	113 (86)	52 (70)
Hispanic	27 (7)	7 (4)	8 (6)	12 (16)
Other	12 (3)	5 (3)	3 (2)	4 (5)
Total	381 (100)	174 (100)	132 (100)	75 (100)
Diagnosis ^a				
Diplegia	265 (69)	85 (49)	105 (80)	75 (100)
Hemiplegia	116 (31)	89 (51)	27 (20)	0 (0)
Total	381 (100)	174 (100)	132 (100)	75 (100)
Age ^a (y), <i>n</i>				
Mean	11.0	11.2	10.6	11.3
SD	3.4	3.4	3.3	3.4
Range	4.1–18.3	4.2–18.3	4.1–17.4	4.8–18.3
Height (cm), <i>n</i>				
Mean	138	140.6	135.7	135.8
SD	19.1	19.5	19.1	17.7
Range	90–180	90–180	96–180	102–171
Weight (kg), <i>n</i>				
Mean	38.5	39.5	37.6	37.8
SD	17.5	17.2	17.6	18
Range	12–120	13–101	12–120	15–114
Time between assessments (y)				
	1.4	1.4	1.3	1.4
	0.4	0.4	0.3	0.5

^aGMFCS level, diagnosis, and age are as reported at the baseline visit.

Table II

Mean, SD, and adjusted standardized response means (SRM_a) of outcome tool change scores for nonsurgical (n=292) and surgical (n=87) groups by Gross Motor Function Classification System (GMFCS) level

Outcome tool	Non-surgical						Surgical					
	GMFCS I		GMFCS II		GMFCS III		GMFCS I		GMFCS II		GMFCS III	
	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a
GMFM Dimension D	-0.1 (5.3)	-0.02	1.5 (7.4)	0.23	2.8 (8.6)	0.90	-1.5 (4.3)	-0.42	-2.9 (11.2)	-0.25	3.1 (12.6)	0.41
GMFM Dimension E	-0.2 (8.3)	-0.03	1.0 (9.3)	0.18	2.5 (9.0)	0.69	0.3 (5.9)	0.04	-2.5 (10.5)	-0.36	-0.4 (14.2)	-0.04
GMFM-66	0.2 (5.4)	0.07	0.8 (3.5)	0.43	0.5 (3.5)	0.34	-1.7 (5.2)	-0.42	0.4 (3.7)	0.22	1.0 (3.5)	0.68
O ₂ cost (ml O ₂ /kg/min)	-0.01 (0.09)	-0.07	-0.02 (0.18)	-0.08	0.04 (0.17)	0.35	0.00 (0.06)	0.01	-0.03 (0.1)	-0.26	-0.06 (0.27)	-0.22
Parent PedsQL Emotional Functioning	2.2 (15.6)	0.17	0.4 (13.8)	0.04	-2.4 (14.7)	-0.14	0.8 (17.8)	0.04	3.2 (14.2)	0.28	0.3 (11.5)	0.03
Parent PedsQL Physical Functioning	-0.4 (16.5)	-0.03	1.8 (16.9)	0.11	-0.7 (18.9)	-0.03	7.1 (19.5)	0.33	2.9 (14.9)	0.26	5.6 (17.2)	0.36
Parent PedsQL School Functioning	1.3 (17.6)	0.17	-1.6 (16.7)	0.24	-1.4 (19.6)	0.24	0.3 (15.7)	0.02	1.5 (15.9)	0.10	7.3 (15.6)	0.54
Parent PedsQL Social Functioning	2.4 (15.8)	0.08	4.1 (17.3)	-0.11	3.8 (15.2)	-0.06	4.4 (19.6)	0.23	-0.6 (20.0)	-0.03	8.0 (13.3)	0.93
Parent PODCI Expectations	-0.2 (25.2)	-0.01	-3.8 (22.4)	-0.15	-4.0 (22.9)	-0.15	5.0 (25.9)	0.15	-1.3 (22.5)	-0.05	-3.7 (28.2)	-0.11
Parent PODCI Global Function	1.1 (9.9)	0.13	3.6 (10.8)	0.37	1.0 (8.5)	0.18	4.3 (8.6)	0.73	1.1 (13.0)	0.09	6.9 (11.6)	0.72
Parent PODCI Happiness	-1.6 (22.0)	-0.06	1.4 (16.8)	0.09	-0.6 (13.7)	-0.06	3.0 (14.2)	0.23	2.2 (18.4)	0.13	1.5 (19.1)	0.09
Parent PODCI Comfort/Pain	-0.9 (21.2)	-0.04	5.8 (21.2)	0.28	-1.9 (21.7)	-0.08	4.8 (19.7)	0.25	1.9 (28.2)	0.06	12.9 (23.3)	0.56
Parent PODCI Satisfaction	2.3 (32.0)	0.07	1.3 (28.1)	0.05	5.2 (28.7)	0.21	12.1 (34.7)	0.34	7.9 (37.3)	0.16	20.0 (37.7)	0.52
Parent PODCI Sports	1.0 (12.4)	0.11	2.5 (13.9)	0.21	2.1 (13.2)	0.18	3.6 (12.7)	0.37	0.3 (14.5)	0.03	3.4 (16.5)	0.24
Parent PODCI Upper Extremity	3.9 (10.0)	0.54	3.9 (11.6)	0.47	1.6 (9.8)	0.31	4.0 (9.5)	0.78	1.3 (12.8)	0.15	3.6 (12.8)	0.42
Parent PODCI Transfers	0.6 (10.8)	0.05	2.5 (10.3)	0.28	2.3 (12.1)	0.27	4.6 (7.9)	0.83	0.9 (11.5)	0.09	7.9 (13.2)	0.79
WeeFIM Cognition	-0.8 (9.1)	-0.10	0.2 (9.1)	0.03	0.4 (7.1)	0.07	-1.3 (8.8)	-0.12	0.7 (8.3)	0.08	2.6 (6.7)	0.93
WeeFIM Mobility	-0.1 (4.2)	-0.03	0.7 (8.2)	0.08	0.8 (8.8)	0.15	1.0 (5.1)	0.21	1.4 (6.0)	0.38	4.2 (11.3)	0.47
WeeFIM Self Care	0.8 (7.3)	0.14	2.0 (11.0)	0.25	3.0 (10.4)	0.46	2.7 (8.3)	0.44	5.7 (12.8)	0.50	1.2 (15.4)	0.09
Cadence (% normal)	0.3 (10.8)	0.03	0.2 (14.0)	0.01	3.5 (14.8)	0.34	-3.2 (11.5)	-0.23	1.1 (15.4)	0.07	2.5 (28.0)	0.07
Stride length (% normal)	-1.8 (9.9)	-0.21	-1.1 (10.7)	-0.14	-3.5 (13.2)	-0.36	-0.1 (9.5)	-0.01	-7.2 (15.0)	-0.49	-1.0 (15.5)	-0.07
Velocity (% normal)	-0.8 (16.5)	-0.04	0.5 (16.2)	0.04	-0.9 (15.2)	-0.08	-3.0 (16.3)	-0.15	-5.8 (16.5)	-0.37	2.4 (20.1)	0.13
Child PedsQL Emotional Functioning	4.0 (21.5)	0.17	0.7 (22.4)	0.02	-0.3 (18.9)	-0.02	Insufficient number per group					
Child PedsQL Physical Functioning	0.1 (16.9)	0.01	-0.9 (19.4)	-0.04	-1.1 (18.8)	-0.06						
Child PedsQL School Functioning	3.5 (24.2)	0.09	0.7 (17.6)	0.08	0.2 (16.7)	-0.04						

Outcome tool	Non-surgical						Surgical					
	GMFCS I		GMFCS II		GMFCS III		GMFCS I		GMFCS II		GMFCS III	
	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a	Mean (SD)	SRM _a
Child PedsQL Social Functioning	2.1 (20.8)	0.14	2.0 (20.6)	0.04	-0.9 (19.5)	0.01						
Child PODCI Expectations	-0.2 (23.5)	-0.01	-5.3 (20.4)	-0.28	-6.1 (30.5)	-0.18						
Child PODCI Global Function	-0.2 (9.0)	-0.02	-0.3 (12.6)	-0.02	1.3 (9.3)	0.11						
Child PODCI Happiness	1.3 (14.3)	0.09	2.7 (18.2)	0.15	0.4 (14.6)	0.04						
Child PODCI Comfort/Pain	-2.7 (23.0)	-0.10	-2.6 (32.5)	-0.05	1.5 (21.4)	0.06						
Child PODCI Satisfaction	-1.6 (34.7)	-0.04	10.2 (39.6)	0.21	-3.3 (35.6)	-0.09						
Child PODCI Sports	0.3 (13.5)	0.03	-0.5 (17.1)	-0.03	1.4 (17.9)	0.07						
Child PODCI Upper Extremity	0.6 (5.5)	0.12	1.63 (5.92)	0.40	1.4 (9.8)	0.12						
Child PODCI Transfers	-0.3 (5.9)	-0.06	0.28 (9.68)	0.03	0.7 (10.4)	0.05						

Values in bold type indicate change scores that exceed the magnitude for minimum clinically important difference scores required for medium effect size. Values in bold and italic type indicate change scores that exceed the magnitude for minimum clinically important difference scores required for large effect size. GMFM, Gross Motor Function Measure; GMFM-66, Gross Motor Function Measure Score calculated using the Gross Motor Activity Estimator Program; O₂ cost, energy (oxygen) cost during walking; PedsQL, Pediatric Quality of Life Inventory; PODCI, Pediatric Outcomes Data Collection Instrument; WeeFIM, Pediatric Functional Independence Measure.

Table III

Minimum change scores needed on outcome tools for Minimum Clinically Important Difference (MCID) of medium (0.5) or large (0.8) effect size, by Gross Motor Function Classification System (GMFCS) level and overall

Outcome tool	GMFCS I (n=141)		GMFCS II (n=96)		GMFCS III (n=55)		Overall (n=292)	
	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)
GMFMDimension D	2.4	3.8	3.3	5.3	1.5	2.4	1.2	1.8
GMFMDimension E	4.0	6.5	2.8	4.5	1.8	3.0	1.6	2.6
GMFM-66	1.7	2.7	1.0	1.5	0.7	1.2	0.8	1.3
O ₂ cost (ml O ₂ /kg/min)	0.04	0.06	0.11	0.17	0.06	0.09	0.06	0.09
Parent PedsQL Emotional Functioning	6.6	10.5	5.6	9.0	8.5	13.6	6.5	10.5
Parent PedsQL Physical Functioning	7.9	12.7	8.1	13.0	10.7	17.1	7.9	12.7
Parent PedsQL School Functioning	7.1	11.4	8.6	13.8	7.7	12.3	7.7	12.3
Parent PedsQL Social Functioning	7.8	12.4	7.1	11.3	10.7	17.1	8.0	12.8
Parent PODCI Expectations	13.8	22.0	12.7	20.3	12.9	20.6	13.3	21.2
Parent PODCI Global Function	4.5	7.2	4.9	7.9	2.8	4.5	3.8	6.0
Parent PODCI Happiness	12.6	20.1	7.8	12.5	5.2	8.4	9.7	15.6
Parent PODCI Comfort/Pain	11.4	18.3	10.2	16.3	11.8	18.9	11.2	18.0
Parent PODCI Satisfaction	17.3	27.6	12.0	19.2	12.4	19.8	14.4	23.0
Parent PODCI Sports	4.6	7.3	6.0	9.7	5.7	9.0	4.3	6.8
Parent PODCI Upper Extremity	3.6	5.8	4.1	6.5	2.6	4.2	3.4	5.4
Parent PODCI Transfers	5.9	9.4	4.5	7.2	4.1	6.6	4.0	6.4
WeeFIM Cognition	4.0	6.4	3.1	5.0	2.4	3.9	3.4	5.5
WeeFIM Mobility	1.9	3.0	4.4	7.1	2.7	4.3	2.3	3.9
WeeFIM Self Care	2.8	4.5	4.0	6.4	3.3	5.2	3.1	5.0
Cadence (% normal)	5.9	9.5	7.6	12.2	5.2	8.3	5.0	8.1
Stride length (% normal)	4.2	6.7	3.9	6.3	4.9	7.9	3.6	5.8
Velocity (% normal)	8.7	13.9	6.8	10.9	5.5	8.8	5.7	9.1
Child PedsQL Emotional Functioning	12.0	19.3	13.9	22.2	9.4	15.0	12.1	19.3

Dev Med Child Neurol. Author manuscript; available in PMC 2010 November 24.

Outcome tool	GMFCS I (n=141)		GMFCS II (n=96)		GMFCS III (n=55)		Overall (n=292)	
	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)	MCID medium (0.5)	MCID large (0.8)
Child PedsQL Physical Functioning	8.3	13.2	11.3	18.1	9.1	14.6	8.8	14.1
Child PedsQL School Functioning	11.9	19.1	12.0	19.3	9.7	15.4	11.4	18.3
Child PedsQL Social Functioning	12.9	20.6	7.8	12.4	7.7	12.3	10.5	16.8
Child PODCI Expectations	11.2	17.9	9.4	15.1	17.0	27.1	12.0	19.1
Child PODCI Global Function	4.3	6.9	7.6	12.2	5.7	9.1	5.1	8.2
Child PODCI Happiness	6.9	11.0	9.1	14.6	5.1	8.1	7.0	11.2
Child PODCI Comfort/Pain	13.6	21.8	24.1	38.6	13.1	20.9	16.5	26.3
Child PODCI Satisfaction	19.9	31.8	24.4	39.1	18.3	29.3	20.7	33.1
Child PODCI Sports	5.9	9.4	8.9	14.3	9.7	15.5	6.4	10.3
Child PODCI Upper Extremity	2.5	4.1	2.0	3.3	6.0	9.5	3.0	4.8
Child PODCI Transfers	3.0	4.7	5.3	8.5	6.6	10.5	2.0	5.2

GMFCS, Gross Motor Function Measure; GMFCS-66, Gross Motor Function Measure Score calculated using the Gross Motor Activity Estimator Program; O2 cost, energy (oxygen) cost during walking; PedsQL, Pediatric Quality of Life Inventory; PODCI, Pediatric Outcomes Data Collection Instrument; WeeFIM, Pediatric Functional Independence Measure.

Table IV

Mean change score, SD of change, and *p* value of Gross Motor Function Classification System level change groups (Improved, No Change, and Declined)

Tool subscales	Improved		No Change		Declined	
	Mean (SD)	<i>p</i> value	Mean (SD)	<i>p</i> value	Mean (SD)	<i>p</i> value
GMFM Dimension D ^a	3.6 (4.8)	0.01	0.9 (6.9)	0.02	-8.9 (10.1)	0.01
GMFM Dimension E ^a	3.9 (9.4)	ns	0.8 (6.7)	0.02	-12.2 (19.5)	0.04
GMFM-66 ^a	1.4 (2.4)	0.03	0.5 (4.0)	0.02	-4.0 (6.5)	0.05
Parent PODCI Global ^a	5.4 (10.7)	0.05	2.5 (10.2)	<0.001	<i>-5.5 (9.6)</i>	ns
Parent PODCI Transfers ^a	5.5 (13.5)	ns	2.3 (10.5)	<0.001	<i>-5.9 (12.9)</i>	ns
Parent PODCI Sports ^a	5.1 (14.1)	ns	2.2 (12.9)	<0.001	-9.7 (6.6)	<0.001
Velocity (% normal) ^a	-0.1 (13.7)	ns	-0.4 (16)	ns	-13.2 (18.6)	0.03
Stride Length (% normal) ^a	-1.0 (7.8)	ns	-1.8 (11.3)	<0.001	-10.9 (17.0)	0.04
Parent PODCI Satisfaction	13.9 (23.0)	0.02	5.0 (31.5)	<0.001	-9.6 (45.0)	ns
Parent PedsQL Physical	8.6 (14.6)	0.02	1.4 (17.3)	ns	<i>-7.9 (12.5)</i>	0.04

^aThese rows indicate significant changes that differed between the No Change versus Declined, and Improved versus Declined groups. Values in bold type indicate mean change scores that exceed minimum clinically important difference scores for large effect sizes. Values in italic type indicate mean change scores that exceed minimum clinically important difference scores for medium effect sizes. ns, not significant.