Published in final edited form as:

J Pediatr Orthop. 2008 March ; 28(2): 192–198. doi:10.1097/BPO.0b013e3181652185.

Analysis of the Pediatric Outcomes Data Collection Instrument (PODCI) in Ambulatory Children with Cerebral Palsy Using Confirmatory Factor Analysis and Item Response Theory Methods

Diane D. Allen, PhD, PT,

Health and Disability Research Institute, School of Public Health, Boston University, Boston, MA

George E. Gorton, BS,

Shriners Hospitals for Children, Springfield, MA

Donna J. Oeffinger, PhD,

Shriners Hospitals for Children, Lexington, KY

Chester Tylkowski, MD,

Shriners Hospitals for Children, Lexington, KY

Carole A. Tucker, PhD., PT, PCS, and

Shriners Hospitals for Children, Philadelphia, PA

Stephen M. Haley, PhD, PT

Health and Disability Research Institute, School of Public Health, Boston University, Boston, MA

Abstract

Background—Traditional use of the Pediatrics Outcomes Data Collection Instrument (PODCI) assumes that all items have the same structure, are measuring the intended constructs, and assess the right levels of function to show change following orthopedic or neurologic intervention. Item response theory (IRT) methods can statistically account for inherent differences in PODCI item characteristics and thus reveal attributes of the measure important to effectiveness research. Our study uses IRT methods to determine whether PODCI items: fit the projected dimensional structure of the PODCI, assess function on each dimension at the right level for a population of ambulatory children with cerebral palsy (CP), and reveal changes following intervention in this population.

Methods—Proxy-reported PODCI questionnaires for 570 ambulatory children with CP were randomly divided into two groups for model creation and model testing using exploratory and then confirmatory factor analysis. The resulting model was compared to the projected dimensional structure, tested for fit of individual items, and examined for gaps and ceiling effects. Response changes at one year were compared between those with (n=91) and without (n=284) surgical intervention using paired *t*-tests.

Results—Factor analysis reduced the projected dimensions from five to four for this population, resulting in dimensions for mobility, upper extremity function (UEF), comfort and general health (CGH), and self-worth. All but three items fit their respective dimensions; ceiling effects were noted in all dimensions. Responses showed changes in the CGH, mobility, and UEF dimensions in those who had surgery; in those children who did not have surgery, only the UEF responses changed.

Conclusions—The PODCI can show change following intervention when data are analyzed using IRT methods. Ceiling effects in all dimensions may limit the amount of change the PODCI can show in a population of ambulatory children with CP.

Level of Evidence: Level II—This was a retrospective investigation of a diagnostic tool, the PODCI, using a randomized cross-sectional design for model development, and a case-control design to assess sensitivity to change.

Introduction

The Pediatric Outcomes Data Collection Instrument¹ (PODCI) can document functional status in children and adolescents before and after intervention. Traditional PODCI scoring, however, is based on assumptions of item consistency and construct uniformity, inaccuracies that can limit the instrument's meaningfulness. Analysis using item response theory (IRT) methods can expose assumptions and promote interpretation of PODCI results despite different numbers of item response choices, ways of asking questions, sub-types of function addressed, and levels of ability assessed.^{2, 3} Using IRT methods, we can assess the PODCI items and constructs for their overall value in documenting effective health care in a particular population.

The PODCI was developed specifically to assess changes following pediatric orthopedic interventions for a broad range of diagnoses.¹ These diagnoses include ambulatory children with cerebral palsy (CP), having Gross Motor Function Classification System (GMFCS)⁴ levels of I to III, indicating motor function that is minimally to moderately impaired. The PODCI items focus on function and quality of life of the child, attributes that might change with surgical intervention. Reliance on changes in the PODCI to assess effectiveness of intervention, however, requires an understanding of its items and constructs.

The items within the PODCI vary greatly. Items have different weights, with possible scores ranging from three (for response choices indicating “often,” “sometimes,” or “rarely or never”) to six (for response choices indicating “none,” “very mild,” “mild,” “moderate,” “severe,” or “very severe”). For most items a lower number indicates higher functioning or a more positive quality of life, but for some items, a higher number indicates more positive attributes. The PODCI authors designate seven sub-constructs or dimensions within the instrument: upper extremity function, transfers and mobility, physical function and sports, comfort (lack of pain), happiness, satisfaction, and expectations. A few items address general health and are not scored with a designated dimension. Functional items range in difficulty from getting up from a chair to participating in competitive sports. With multiple weights, constructs, and levels of function represented, scoring may not mean what the user assumes it does.

Despite the item differences, previous studies with the PODCI have accumulated evidence of its psychometric properties. Based on 470 parent and adolescent respondents to the PODCI, Daltroy et al.¹ provided evidence of test-retest reliability (with a 1-2 day interval), internal consistency, and discriminant validity of the various dimensions. PODCI responses indicated lower levels of function for children with upper extremity amputation⁵ or with scoliosis⁶ compared to children without a diagnosed pathology. The global function and comfort subscale (a combination of the four functional dimensions) correlates moderately strongly with performance in gait measures for children with cerebral palsy.⁷ The PODCI dimensions of transfers and mobility and physical function and sports correlate moderately strongly with functional level as classified by GMFCS levels of I, II, or III.⁸ Children or their proxy respondents generally respond at the ceiling of most functional and quality of life dimensions if they have no diagnosed pathology.⁹ Ceiling effects have also been noted with the PODCI in some populations with pathology.¹⁰ Change was minimal across a total group of 316 children with orthopedic impairments from baseline to 9 months later, but was significant across all dimensions for a subgroup of 113 children who were judged “likely to improve.”¹

The PODCI showed improvements in the global function and comfort sub-scale and the physical function and sports dimension following muscle tendon lengthening surgery for children with CP.¹¹

Judging by this previous work, the PODCI has potential usefulness in documenting the effectiveness of intervention in ambulatory children with CP. None of the previous work, however, examined the characteristics of individual items or showed the relationship of items to constructs within the PODCI. The intent of this analysis is to provide evidence of the usefulness of the PODCI as a measure of effective intervention for ambulatory children with CP. The specific purposes of this paper are:

- a. to determine the best-fitting model of constructs underlying the items within the PODCI for a population of ambulatory children with CP,
- b. to determine if gaps in item content or ceiling or floor effects exist for this population, and
- c. to examine changes in PODCI responses on particular dimensions following intervention.

Materials and Methods

The Functional Assessment Research Group (FARG) of the Shriners Hospitals for Children collected data from children with CP at seven pediatric orthopedic hospitals during a six year prospective investigation of the applicability of outcome instruments with the CP population. 12 PODCI questionnaires were completed by a parent or guardian for 570 children. Of those 570, 375 also completed PODCI questionnaires one year after baseline or one year after surgical intervention. Ninety-one children had surgical intervention following their baseline analysis. The ages of the children were 4 to 19 years (mean = 10.8, standard deviation = 3.6) (Figure 1). GMFCS levels were I to III (Figure 2). We performed secondary analyses of these data; all data were stripped of names or identifying information prior to use in this study. This study was approved by the Boston University Institutional Review Board.

To examine the construct and sub-constructs or dimensions of the PODCI, we first reviewed the Scoring Algorithm for version 2.0 (1997) of the PODCI.¹³ This Scoring Algorithm presents the authors' allocation of items to each of the dimensions, as well as the weighting of individual items and equations used for calculating dimension scores. As per the scoring algorithm, we counted as "missing" those items that did not apply to a particular child who was too young, did not attend school, or for whom school was not in session.

To examine the dimensions of the PODCI, we randomly split the initial data set of 570 cases into 2 groups of 285 each. First, the data set for group 1 was used to create a model through exploratory factor analysis that assigned items to dimensions or factors. Then, the data set for group 2 was used to test the chosen model through confirmatory factor analysis.

Factor analyses were performed with the software Mplus,¹⁴ using the unweighted least squares (ULS) method, Promax rotation (allowing correlation between the factors), and specification for categorical (as opposed to continuous) data. For this study, all items relating to function, general health, or quality of life (excluding parental expectations or satisfaction) were initially included in the factor analyses.

Several analyses were performed and the "best" model was chosen based on the following criteria: all factors have an eigenvalue greater than 1, the model explains a high percentage of the variability in the data, the model contains the smallest number of factors possible while grouping items together logically by content, and no factor contains fewer than four items.

Once the “best” model was identified in the group 1 data set, it was tested in the group 2 data set with confirmatory factor analysis. The model fit in the group 2 data set was examined via the root mean square error of approximation (RMSEA). An RMSEA of 0.05 was considered a good fit, and over 0.1 was considered a poor fit. Also, the item groupings were compared between groups 1 and 2 to be sure that groupings in the exploratory phase remained the same in the confirmatory phase. Once confirmed, the new model was compared to the groupings of items designated as dimensions by the PODCI Scoring Algorithm.¹³

After the model was confirmed, we used IRT methods to analyze each of the new factors or dimensions separately. IRT methods allow us to create a scale for each dimension, like upper extremity function, so that low and high functional abilities are at each end of the scale. We can estimate the location of each child’s functional level on the scale, and each item response choice on the same scale, whether easy or hard at the corresponding level of functional ability (Figure 3). The position of the child and item response levels on the scale is determined mathematically based on the probability of responding with the selected item response choices for a child with that level of functional ability. For example, when responding for a child with a low level of function, the probability is low that the respondent will mark “easy” to a question about lifting heavy books, a difficult item for most children. We used data from the initial 570 children and specified one-parameter IRT models in ConQuest¹⁵ software to assess how the items functioned within each dimension.

With IRT methods, estimations of child and item response locations all have an associated measurement error. A smaller error indicates greater certainty in estimating a child’s location and depends on the number and proximity of item response levels on the scale (see Figure 3). When no item responses exist close to the scale levels of particular children, designated as gaps or floor or ceiling effects, larger measurement errors can diminish the ability of the instrument to locate children accurately or to record their changes with intervention. Once we estimated the locations, we compared a histogram of child and item response locations on each dimension to examine mismatches or gaps that might indicate insufficient item responses for accuracy at any place on the scale.

To assess item functioning further, we also examined item “fit,” an indication of association with the other items in each dimension. We designated items as fitting well if the weighted mean square provided by ConQuest was between 0.75 and 1.34 and the *t* statistic was between -1.96 and +1.96.³

To assess sensitivity to changes related to intervention, we compared the initial and one-year data for the children without (n=284) and with (n=91) surgical intervention. This analysis was performed in ConQuest, using a multidimensional IRT model incorporating the four dimensions of the new model. Paired *t*-tests revealed any differences between initial and one-year child estimated locations on any of the dimensions for the no-intervention and intervention groups. Since we expected change in the direction of improvement in function and quality of life, we used one-tailed tests. We set α at 0.0125 to maintain the analyses’ overall α at .05 while four dimensions were examined for each child.

Results

Initial exploratory factor analysis divided items into nine factors but the last several factors had only two to three items in each. With subsequent models we forced successively fewer factors. Some items loaded with no other groups of items: the two items asking about use of assistive devices, and the item regarding health one year ago. We omitted these three items from further analyses. The four-factor model was chosen for testing in the group 2 data set. The four-factor model had all factors with eigenvalues greater than 1, explaining 58% of the

variability in these data, and items fit logically together in their groups although not necessarily along the same groupings as in the scoring algorithm. The fit of the four-factor model in the group 2 data set showed an RMSEA of 0.085. Other indices of fit indicated marginal (Comparative Fit Index or CFI of 0.85) to acceptable (Tucker Lewis Index or TLI of 0.92) fit of the model to the group 2 data set. All items in the group 2 data set loaded on the same factors as for the group 1 data set and contributed significantly to their respective factors.

Factors obtained in this sample slightly differed from the item groupings designated in the PODCI Scoring Algorithm (Table 1). Transfers and basic mobility items loaded together with sports and physical function items. Upper extremity items remained separate from the other mobility items. "Putting on socks" loaded with upper extremity function (UEF) rather than with basic mobility items. The happiness sub-scale identified by the PODCI Scoring Algorithm became part of a self-worth factor because items about making friends, getting together with friends, and being full of pep loaded together with the happiness items. Pain and general health items loaded together in one factor that we re-named comfort and general health (CGH).

When analyzing the factors as separate dimensions in ConQuest, three items did not "fit" within the designated boundaries for their dimensions (bike/trike, participate in gym, miss school). When comparing item response levels and children's locations, there were gaps in the item response histograms; no item responses for any item were available to estimate the corresponding children's functional levels with certainty (Figures 4-7). Lack of available item responses at the high end of the Mobility, Self-Worth, and UEF scales indicate a ceiling effect, with less certainty for locating the highest functioning children.

Using estimates of the children's functional levels from IRT analyses, statistically significant changes were noted in the locations of the children in the surgical intervention group between baseline and one year with improvement in the CGH ($t = -2.77, p = .003$), Mobility ($t = -3.42, p = .0005$), and UEF ($t = -2.38, p = .01$) dimensions. Statistically significant changes were noted in the no-intervention group between baseline and one year with improvement in the UEF dimension ($t = -5.05, p < .00001$). In contrast, when using the PODCI Scoring Algorithm to determine if changes occurred in a one year period, changes were noted in both the upper extremity and transfers and basic mobility dimensions for children without intervention ($t = 5.72$ and 2.59 respectively) and only in the transfers and basic mobility dimension for children with surgical intervention ($t = 2.81, p < .004$).

Discussion

In this data set of ambulatory children with CP, the best-fitting model of constructs underlying the items within the PODCI showed different item groupings than published in the PODCI Scoring Algorithm. Some items that did not associate with the main factors in the original testing of the PODCI (e.g., general health, feel sick and tired, full of pep, make friends) showed associations within the new item groupings, and a few others (the two assistive device items and health a year ago) did not associate with the new groupings. When testing the new groupings, three items did not fit in their dimensions (miss school, bike/trike, and participate in gym). Inspection of these items within their new dimensions revealed substantive reasons why they might not fit: missing school does not fit with UEF items in content even though the factor analysis suggested it belonged in that factor; riding a bicycle or tricycle or participating in gym could have many meanings to respondents since various adaptations might make it possible for children to ride or participate even though the children have vastly different functional levels.

The new dimensions of CGH, Mobility, Self-Worth, and UEF all consist of statistically associated items consistent with a best-fitting model. Inspection of items within the new

dimensions reveals provocative substantive as well as statistical associations. For instance: mobility seems to be a single construct consisting of transfers, locomotion and sports activities lying on a continuum of easier to harder tasks; general health is associated with lack of pain; and making friends is associated with feelings of happiness. While none of these item associations is restricted logically to the population of ambulatory children with CP, the new dimensions proposed in this study cannot be applied to other populations without confirmatory research.

When we examined the new dimensions in comparison with the children in this sample (Figures 4-7), we noted ceiling effects and gaps. Some children's function and quality of life were at the same or higher levels than the top levels of item response possible, with ceiling effects particularly noted for Mobility, Self-Worth, and UEF. The PODCI will not readily show changes for children who start with function close to the top levels (about 5% of this sample), nor will the PODCI show changes well when children's locations are close to gaps in item response levels. Future users of the PODCI wanting to show change at these levels following intervention should consider adding more difficult tasks or response levels to fill in the gaps in the four dimensions.

When examining changes in PODCI responses following intervention, the new dimensions revealed a different sensitivity than that revealed when using the PODCI Scoring Algorithm. Specifically, analyses with IRT methods and these new dimensions proved more sensitive to the types of changes expected (in Mobility, CGH, and UEF) following surgical interventions in this population. Even with the new dimensions, however, the effect size for these changes was small, at 0.19 for UEF, 0.24 for Mobility, and 0.34 for CGH. Alternative explanations for small effect sizes include lack of large functional change following these interventions and the dilution of intervention-related changes over a one-year follow-up period that inevitably includes developmental changes as well (as the change in UEF in the no-intervention group indicates). The intervention group started at levels of function and quality of life below the no-intervention group in all dimensions, perhaps the reason why interventions were considered for this group. It may be that both groups could have shown further improvements if the ceiling effect had not restricted the ability of the PODCI to record higher levels of function and quality of life.

In summary, confirmatory factor analysis and IRT methods revealed a modified set of constructs or dimensions underlying the PODCI. Further examination of the new dimensions revealed sensitivity to changes with intervention despite existing ceiling effects and gaps in the ability of the PODCI to assess the population of interest. The construct validity evidence provided here supported different item groupings for ambulatory children with CP than those suggested by the PODCI Scoring Algorithm. The dimensions showing change following surgical intervention included Mobility, UEF, and CGH, despite ceiling effects in Mobility and UEF. Thus, the PODCI remains a useful tool for assessing effectiveness particularly of surgical intervention using these dimensions in this population, but it has limitations not entirely mitigated by using IRT methods to analyze the results. Additional measures, particularly those that assess higher levels of function and quality of life, may improve our ability to record evidence of effectiveness in this population.

Acknowledgements

The authors would like to acknowledge the Functional Assessment Research Group (FARG) for the data for this study. FARG collected the data under a grant from Shriners Hospitals for Children. At the time this study was completed, additional funding sources included grant #RAB8957 from Shriners Hospitals for Children (Dr. Tucker, PI), a post-doctoral award to Dr. Allen in grant # H133P0001 from the National Institute on Disability and Rehabilitation Research, and an Independent Scientist Award (K02 HD45354-01) to Dr. Haley.

References

1. Daltroy L, Liang M, Fossel A, Goldberg M. The POSNA pediatric musculoskeletal functional health questionnaire: report on reliability, validity, and sensitivity to change. *J Pediatr Orthop* 1998;18(5): 561–571. [PubMed: 9746401]
2. Wright, BD.; Masters, GN. Rating scale analysis. MESA Press; Chicago, IL: 1982.
3. Wilson M, Allen DD, Li JC. Improving measurement in behavioral sciences using item response modeling: introducing item response modeling. *Health Educ Res* 2006;21(Supplement 1):i4–i18. [PubMed: 17018765]
4. Palisano R, 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;39:214–223. [PubMed: 9183258]
5. Lerman JA, Sullivan E, Barnes DA, Haynes RJ. The Pediatric Outcomes Data Collection Instrument (PODCI) and functional assessment of patients with unilateral upper extremity deficiencies. *J Pediatr Orthop* 2005;25(3):405–407. [PubMed: 15832164]
6. Lerman JA, Sullivan E, Haynes RJ. The Pediatric Outcomes Data Collection Instrument (PODCI) and functional assessment in patients with adolescent or juvenile idiopathic scoliosis and congenital scoliosis or kyphosis. *Spine* 2002;27(18):2052–2057. [PubMed: 12634568]
7. Tervo RC, Azuma S, Stout J, Novacheck T. Correlation between physical functioning and gait measures in children with cerebral palsy. *Dev Med Child Neurol* 2002;44:185–190. [PubMed: 12005321]
8. Oeffinger DJ, Tytkowski CM, Rayens MK, et al. Gross motor function classification system and outcome tools for assessing ambulatory cerebral palsy: a multicenter study. *Dev Med Child Neurol* 2004;46:311–319. [PubMed: 15132261]
9. Haynes RJ, Sullivan E. The Pediatric Orthopaedic Society of North America pediatric orthopaedic functional health questionnaire: an analysis of normals. *J Pediatr Orthop* 2001;21(5):619–621. [PubMed: 11521030]
10. Vitale MG, Levy DE, Moskowitz AJ, et al. Capturing quality of life in pediatric orthopaedics: two recent measures compared. *J Pediatr Orthop* 2001;21(5):629–635. [PubMed: 11521032]
11. 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;25:641–645. [PubMed: 16199947]
12. Oeffinger DJ, Gorton GE, Bagley AM, et al. Outcome assessments in children with cerebral palsy, Part I: descriptive characteristics of GMFCS Levels I to III. *Dev Med Child Neurol* 2007;49:172–180. [PubMed: 17355472]
13. American Academy of Orthopaedic Surgeons. Pediatric Orthopaedic Society of North America. American Academy of Pediatrics. Shriners Hospitals. Scoring Algorithms for Pediatrics Outcomes Data Collection Instrument. Vol. 2.0 ed. 1997.
14. *Mplus* [computer program]. Version 3.0. Muthen & Muthen; Los Angeles, CA: 2001.
15. ACER ConQuest: Generalised Item Response Modelling Software [computer program]. Version 2.0. ACER (Australian Council for Educational Research) Press; Hawthorn, Australia: 2003.

Figure 1.
PODCI Age Frequency Data, N=570

Figure 2.
PODCI Frequency Data for GMFCS Score (N=568)

Figure 3.

Comparing the Location of Three Children with the Locations of Item Responses on a Scale Representing Functional Status. An example item asks: During the last week was it easy or hard for your child to lift heavy books? “Easy” is the most probable response for children A and C. Child B probably cannot do this activity. Child C’s location is close to the location of an item response so is more certain than child A; the measurement error for the estimated location is smaller for child C.

Figure 4.

Histograms on a Scale Representing Comfort and General Health. The histogram on the left indicates the number of children at each location on the scale. The histogram on the right indicates the cumulative item responses at each location when considering the five items associated with this dimension.

Figure 5. Histograms on a Scale Representing Mobility. The histogram on the left indicates the number of children at each location on the scale. The histogram on the right indicates the cumulative item responses at each location when considering the nineteen items associated with this dimension.

Figure 6.

Histograms on a Scale Representing Self-Worth. The histogram on the left indicates the number of children at each location on the scale. The histogram on the right indicates the cumulative item responses at each location when considering the eight items associated with this dimension.

Figure 7.

Histograms on a Scale Representing Upper Extremity Function. The histogram on the left indicates the number of children at each location on the scale. The histogram on the right indicates the cumulative item responses at each location when considering the ten items associated with this dimension.

Table 1
Four-Factor Model of PODCI Compared to Designations Recommended in PODCI Scoring Algorithm

Four Factor Model	Item	PODCI Designation	Four Factor Model	Item	PODCI Designation
Upper Extremity	Lift heavy books	UE	Mobility	Sit in chair w/o holding	TBM
	Pour a half gallon of milk	UE		In and out of bed	TBM
	Open a jar	UE		On and off toilet/chair	TBM
	Use a fork and spoon	UE		Stand at sink	TBM
	Comb hair	UE		Need help for stand/sit	TBM
	Button buttons	UE		Participate in gym/recess	SPF
	Put on socks	TBM		Walk 1 block	TBM
	Write with a pencil	UE		Climb 1 flight stairs	TBM
	Open door knobs	UE		Bend/pick up from floor	TBM
	Miss school	(Not designated)		On and off bus	TBM
Comfort and General Health	How much did pain interfere with normal activities	PC		Need help to walk/climb	SPF
	How much pain	PC		Run short distance	SPF
	General health	(Not designated)		Walk 3 blocks	SPF
	Feel sick and tired	(Not designated)		Climb 3 flights stairs	SPF
	How much of time did pain interfere with activities	PC		Ride bicycle/tricycle	SPF
Self-Worth	Happy with health	Hap		Outdoor recreation	SPF
	Happy with looks	Hap		Walk more than mile	SPF
	Happy with clothes	Hap		Pick-up games/sports	SPF
	Full of pep	(not designated)		Competitive sports	SPF
	Make friends	(not designated)			
	Happy with body	Hap			
	Get together with friends	SPF			
	Happy that can do the same things as friends	Hap			

UE = upper extremity

TBM = transfers, basic mobility

Hap = happiness

SPF = sports, physical fitness

PC = pain, comfort