Journal of Pediatric Orthopaedics
15:244–253 © 1995 Raven Press, Ltd., New York

Measuring Pediatric Physical Function

Nancy L. Young, B.Sc.P.T., M.Sc., and James G. Wright, M.D., M.P.H., F.R.C.S.C.

Division of Orthopaedic Surgery and Clinical Epidemiology Unit, The Hospital for Sick Children, Toronto, Ontario, Canada

Summary: Most pediatric orthopaedic interventions are intended to improve or preserve physical function, yet their outcomes have been assessed using primarily surrogate measures (e.g., radiographic indices) that may not accurately represent patients' function. Physical function may be more appropriately measured with activity-based scales, but these have been infrequently applied in surgical studies. The purpose of this study was to identify existing activity-based physical-function scales appropriate for pediatric orthopaedics, to present criteria useful for scale selection, and to discuss the special problems of measuring physical function in children. Twenty-one scales relevant to pediatric orthopaedics are described according to their target population, purpose, method of administration, content, and quality of standardization. These scales have been further classified according to a new taxonomy. The unique aspects of measuring physical function in children are discussed and include the effect of age and development, method of reporting, and question formats. Standardized measures of physical function based on physical-activity ability exist and should be used more frequently to assess pediatric orthopaedic interventions. **Key Words:** Measurement—Orthopaedics—Outcome assessment—Physical function.

Most pediatric orthopaedics treatment strategies are intended either to reduce existing physical disability [which has been defined as limitation in physical activity (65)] or to prevent future physical disability. Orthopaedic surgeons, however, have relied primarily on surrogate measures of physical function (e.g., range of motion and radiographic measurement) to assess the effectiveness of interventions. Surrogate measures are popular because they are easy to use, bear immediate relevance, and appear to be reliable. However, surrogate measures are not necessarily reliable (69), and more important, may not accurately represent changes in physical disability [e.g., despite radiographic curve correction in children with spina bifida, walking ability may decrease (40)]. Thus, although surrogate measures may be important in day-to-day clinical management, physical-function measures based on activities are arguably the most cogent outcomes of orthopaedic interventions.

Many activity-based physical-function scales capable of assessing the effects of clinical management have been developed in the rehabilitation field but have not been widely used by orthopaedic surgeons. Surgeons' infrequent use of such scales may be attributed to a lack of awareness of the existence of appropriate scales, difficulty obtaining scales, or unfamiliarity with the factors to be considered in selecting scales. In addition, because measurement of pediatric function has many special problems, such as the impact of growth and development, which few of the existing scales have completely addressed, surgeons may have deemed these scales inadequate for inclusion in clinical practice or research.

The purpose of this article is to aid surgeons in using activity-based measures by (a) cataloguing activity-based pediatric physical-function measures with direct application to pediatric orthopaedic practice; (b) explaining the options relevant to selection of appropriate scales; and (c) discussing some of the special problems of measuring physical function in pediatrics.

Physical function is defined in this article as the ability to use the musculoskeletal system to interact with the environment in a purposeful way for the performance of activities of daily living, mobility (e.g., manual dexterity, transfers, ambulation), and leisure activities (16,42). Physical function is a distinct subcomponent of more global health measures such as functional status (16), health status (1,42, 52,56), and quality of life (42,63). Because of the nature of orthopaedic interventions, the physical-function components of patients' status is the most

Address correspondence and reprint requests to Dr. J. Wright, Division of Orthopaedic Surgery and Clinical Epidemiology Unit, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8.

Dr. J. Wright is the recipient of a Medical Research Council Scholarship.

likely to be affected and therefore is the focus of this article.

Standardization refers to the reliability, validity, and the responsiveness or discriminative ability of a measure. Reliability (reproducibility or consistency) is the degree to which the scale will yield similar answers when the measure is repeated (60). Validity (or accuracy) is the extent to which the scale measures what it is intended to measure. Scales intended to evaluate change must also be able to detect clinically important change, which is termed responsiveness (or sensitivity). Scales intended to measure patients' status must be able to classify subjects correctly, which is termed discrimination and is quantified in terms of specificity and sensitivity (51).

METHODS

Physical-function measures were identified for review from three sources: a computerized literature search, the experience and resources of authors and colleagues, and the reference lists of publications identified using the first two strategies. The computer searches were conducted using the MEDLINE Index for the years 1984 to March 1993 and the allied health literature from 1984 to February 1992.

Scales were included in the review if the primary intent of the scale was to quantify activity-based physical function. Scales were excluded if they were developed for adult populations without specific documentation of pediatric application or were developmental milestone inventories (assuming a "normal" developmental sequence). Developmental scales have been extensively reviewed elsewhere (8,12–14,20–22,27,43,44,58,61,62); however, one example has been included in the Appendix to promote an understanding of where these scales fit into the taxonomy.

RESULTS

The literature review identified 147 articles. The vast majority of the scales focused on developmental assessment (e.g., Hughes Basic Gross Motor Assessment), psychiatric or behavioral assessment (Child Behavior Check List), and intellectual aptitude (Weschler Scales, Illinois Test of Psycholinguistic Abilities), as has been previously reported (32). Twenty-one scales relevant to pediatric orthopaedics remained after exclusions: 13 were developed primarily for a pediatric population; four are adult scales modified for a pediatric population; and four are adult scales that have not been modified for children but have been used to measure pediatric physical function. These 21 scales are presented in the Appendix; however, the Appendix is not expected to be exhaustive.

For ease of use, the Appendix is subdivided into three sections: scales for children with neurological impairments, scales for children with arthritis, and scales that are not specific to any disease (also called generic scales). Within each section, the scales are listed according to a taxonomic classification. Column 1 lists names of the scales and provides references intended to assist the reader in the further investigation of pertinent scales. In several cases, a publication of the original scale development was not found; thus, the reference cited may be not the original author but a secondary source that presents a sufficient degree of detail for those interested in pursuing this scale further. Columns 2 to 7 are intended to aid surgeons in choosing among the available scales by describing the scales according to criteria that should be considered in the selection of appropriate scales and are described in detail in the following paragraphs.

Purpose of the measure

Column 2 discusses the intended use of the measure. Scales may be *discriminative* (distinguish between groups of patients), *evaluative* (detect change), or *predictive* (forecast the results of subsequent evaluations) (28). In clinical practice, surgeons are most often interested in evaluating the effect of interventions and thus predominantly seek evaluative scales. A scale developed for one purpose is not necessarily valid if used for a different purpose or population.

Domains, format, and scoring

Column 3 lists the functional domains that the scale covers, the format of the questions (e.g., visual analog or categorical), and how the questions are scored and aggregated. In some instances, full information was not available.

Population

Column 4 describes the population(s) on which the scale has been applied. Note that a scale's discriminative ability may be specific to the age or disease group for which it was developed. The measure must also be capable of scoring the full range of ability and disability expected in the study population.

Method of administration

Column 5 describes the details of administering the measure, including by whom (e.g., clinician or self-administered), how (e.g., observation or interview), to whom (e.g., child, parent, or proxy reporter), and the time and special equipment requirements.

Standardization of the measure

Column 6 contains details of the published reliability, validity, and responsiveness testing of the scales. The specifics of each scale's standardization testing can be found in the references cited in the Appendix and should be reviewed before selecting a scale (9,46,60,68,69). One must seek a scale that offers sufficient standardization, bearing in mind that few, if any, scales have been completely standardized and that use of untested or inappropriate outcome variables will decrease the likelihood of detecting clinically important differences.

Taxonomic classification

Column 7 classifies the scales according to a new taxonomy (organizational framework), and the classification number is recorded in brackets in the Appendix. The taxonomy was created to simplify functional scale selection for surgeons and clinical researchers and is shown in Table 1. Under the taxonomy, scales are first classified by whether they are direct or indirect measures and then as capability or performance measures.

Direct measures are scored on the basis of personal observation of an activity or behavior. Indirect measures are scored on the basis of reported activity or behavior. Capability measures assess what the child can do. Performance measures assess what the child does do. Finally, indirect measures are further subdivided according to method of administration into self-report scales and interviewer-administered scales.

Examples of physical-function measures of each of the four main types are as follows. The play-performance scale (31), which requires that a clinician directly observe skills the child uses spontaneously in a normal play environment, is a direct-performance measure (1.1). Clinical gait assessment requires a clinician's direct observation of what the child is able to do when it is demanded of him/her, and thus is a direct-capability measure (1.2). The Klein–Bell ADL scale is an indirect-performance measure, because it is scored based on report of previously observed spontaneous activity in a normal environmental context (2.1). Finally, surveys of what the child can do under hypothetical circumstances (often ideal circumstances) are indirect-capability measures (2.2).

Direct methods may be considered more valid because they eliminate the biases of the reporter but may be sensitive to environmental changes and are rarely practical. Indirect methods offer enhanced feasibility, may have greater consistency of administration, and if self-administered, eliminate interviewer bias. The difficulties inherent in self-report measures include uncertain comprehension or interpretation of the questions and response bias (59).

Capability measures have the advantage of potentially measuring all children in a consistent setting and being able to determine their best ability, but findings may not relate to their community function. Performance measures may better reflect their usual physical function by taking into account the child's usual social, environmental, and emotional settings, but improvement in performance may lag behind improvement in physiologic parameters and capability (e.g., a child may have the required range of motion to climb stairs and be capable but not yet perform the activity at school).

TABLE 1. *Physical function measure taxonomy*[a]

1. Direct measurement (clinical observation)
1.1 Performance based (does do)
1.2 Capability based (can do)
2. Indirect measurement (report of parent, patient, or proxy)
2.1 Performance based (does do)
2.11 Interviewer administered
2.12 Independent/self-administered
2.2 Capability based (can do)
2.21 Interviewer administered
2.22 Independent/self-administered

[a] Scales must be selected on the basis of their purpose and population. It was not feasible to put this into the taxonomy, as there is considerable overlap in purposes and populations.

DISCUSSION

Physical-function measures are infrequently used by pediatric orthopaedists, possibly because of lack of awareness of existing scales, limited availability of the scales, difficulties in making appropriate selections, or because of the special problems inherent in the measurement of pediatric function. The identification and description of 21 scales that are potentially appropriate for pediatric orthopaedics attempt to address the first two obstacles. We hope the taxonomy presented in this article and the discussion of criteria for scale selection will aid surgeons in choosing among the available scales. Finally, the special challenges of measuring physical function in children must be addressed. Clinical application of physical-function scales should offer substantial benefits to research, provided there is some recognition of the effects of age, growth and development, the impact of the parent in reporting, and the framing of questions and response options.

Age, growth, and development

Prime considerations when evaluating the appropriateness of a pediatric scale for a specific population are the age for which the scale is applicable and the effect of development on sequential scores. Due to development, age has a distinct impact on ability to perform certain activities and on their relative importance. For example, tricycle riding is an important part of physical function at age 4 but not at age 8, even though the motor skills required are still present.

At least two methods may accommodate for the effects of age. First, a comprehensive scale may be developed that covers physical function across all age groups, such as the Rand Health Insurance Scale. This method is simple because only one scale is required for all children, but it may not be responsive to clinically important change. A variation of this method is to have a single scale but correct for age or stage of development by dividing the child's score by the maximum potential score for the child's specific age. For example, a 4-year-old might have a maximum score of 100 points, whereas on the same scale, a 5-year-old might have a maximum of 120 because of advanced motor skill level. If a child scores 80 at age 4, has surgery, and subsequently scores 96 at age 5 (both standardized scores are equal to 80% of expected), then no improvement can be attributed to the intervention beyond that which would have occurred with development. (Note, however, that the absence of a decline may be a clinically important finding indicative of success.) The process of age-adjusted scores requires normative data on the populations in question, and expected scores for disabled children are rarely available.

The second option is to use scales that are appropriate for limited age groups. This, however, requires multiple scales to accommodate various age groupings and makes measuring the effect of an intervention in children who cross over into a new age category during the trial very difficult. Thus age-specific scales are not recommended unless a translation between scales for different age groups has been clearly determined.

Self, parent, and proxy report

If the decision has been made to evaluate physical function indirectly, then the source of information must be selected: child, parent, or a proxy. When the focus of the intervention and research is the child, then the child should be the source of information. Parent report is required for patients whose communicative capacity is impaired by age, illness, or cognitive ability. Alternatively, proxy report can be used and may be advantageous when strong parental bias is suspected.

Context

The environmental conditions are particularly important when measuring physical function because they define whether capability or performance is being measured and affect the outcome (e.g., the degree of motivation, environmental distractions, and the presence of parents may significantly affect children's physical function). Additionally, physical function can be measured in multiple ways depending on the wording of the questions. Questions may ask about quality or quantity of function, each potentially yielding a different outcome. For example, physical function can be measured on a scale of independence, which can be affected by physical function, availability of supports, and willingness to accept assistance. The social construction of childhood is such that most children have readily available supports and may also be willing to accept assistance; therefore, independence measures may overestimate children's disability. Thus physical-function measures should not be adopted without consideration of contextual issues. Children's ability to comprehend certain question formats (such as visual analog scales) may also change as a function of age and requires consideration.

In summary, this article has addressed the problems of availability, difficulties in selecting appropriate measures, and conceptual and methodological issues unique to measuring physical function in children. Appropriate scales can be selected using the references, standards, and taxonomy provided. Clinicians are encouraged to include activity-based function outcome measures in clinical and research practice, provided that they evaluate the existing scales carefully with regard to population, purpose, and standardization.

Future research will be required to determine the relationship between performance and capability, the agreement between parents and children, and the preferred context to measure physical function. These issues do not have a single correct answer, but none of the difficulties precludes the use of these activity-based measures. Finally, because orthopaedic interventions are intended to improve (or maintain) function, evaluations of surgical therapy should include measures of physical function, which can then be interpreted on the basis of clinical expectations.

REFERENCES

1. Bergner M, Rothman ML. Health status measures: an overview and guide for selection. *Annu Rev Public Health* 1987; 8:191–210.
2. Boyce W, Gowland C, Russell D, Goldsmith C, Rosenbaum P, Plews N, Lane M. Consensus methodology in the development and content validation of a gross motor performance measure. *Physiother Can* 1993;45(2):94–100.
3. Boyce WF, Gowland C, Hardy S, Rosenbaum PL, Lane M, Plews N, Goldsmith C, Russell DJ. Development of a quality-of-movement measure for children with cerebral palsy. *Phys Ther* 1991;71(11):820–32.
4. Cadman D, Boyle MH, Offord DR, Szatmari P, Rae-Grant NI, Crawford J, Boyles J. Chronic illness and functional limitation in Ontario children: findings of the Ontario child health study. *Can Med Assoc J* 1986;135:761–7.
5. Coulton CJ, Zborowsky E, Lipton J, Newman AJ. Assessment of the reliability and validity of the Arthritis Impact Measurement scales for children with juvenile arthritis. *Arthritis Rheum* 1987;30:819–24.
6. Eisen M, Donald CA, Ware JE, Brook RH. *Conceptualization and measurement of health for children in the health insurance study*. Report Number R-2313-HEW. Santa Monica, CA: Rand Corporation, 1980.
7. Eisen M, Ware JE, Donald D. Measuring components of children's health status. *Med Care* 1979;17:902–21.
8. Ellison PH, Browning CA, Larson B, Denny J. Development of a scoring system for the Milani-Comparetti and Gidoni method of assessing neurologic abnormality in infancy. *Phys Ther* 1983;63:1414–23.
9. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.
10. Feldman AB, Haley SM, Coryell J. Concurrent and construct validity of the Pediatric Evaluation of Disability Inventory. *Phys Ther* 1990;70:602–10.
11. Fife SE, Roxborough LA, Armstrong RW, Harris SR, Gregson JL, Field D. Development of a clinical measure of postural control for assessment of adaptive seating in children with neuromotor disabilities. *Phys Ther* 1991;71:981–93.
12. Gans BM, Haley SM, Hallenborg SC, Mann N, Inacio CA, Faas RM. Description and interobserver reliability of the Tufts Assessment of Motor Performance. *Am J Phys Med Rehabil* 1989;67:202–10.
13. Gesell Institute of Human Development. *The Gesell Pre-*

school Test for evaluating motor, adaptive, language, and personal-social behavior in children ages 2½ to 6. New Haven, CT: The Gesell Institute of Human Development, 1979.
14. Gowland C, King G, King S, Law M, Letts L, MacKinnon L, Rosenbaum P, Russell D. *Review of selected measures in neurodevelopmental rehabilitation (a rational approach for selecting clinical measures)*. Research Report, 91-2. Hamilton, Ontario: Neurodevelopmental Clinical Research Unit, Chedoke-McMaster Hospitals, 1991.
15. Granger CV, Hamilton BB, Kayton R. *Guide for use of the Functional Independence Measure for Children (WeeFIM)*. New York: Research Foundation, State University of New York, 1989.
16. Haley SM, Coster W, Ludlow LH. Pediatric functional outcome measures. *Phys Med Rehabil Clin North Am* 1991;2 (4):689–723.
17. Haley SM, Coster WJ, Faas RM. A content validity study of the Pediatric Evaluation of Disability Inventory. *Pediatr Phys Ther* 1991;3:177–84.
18. Haley SM, Coster WJ, Ludlow LH, Haltiwanger JT, Andrellow PJ. *Pediatric Evaluation of Disability Inventory (PEDI). Development, standardization and administration manual*. Boston, MA: New England Medical Center Hospitals, 1992.
19. Haley SM, Ludlow LH, Gans BM, Faas RM, Inacio CA. Tufts assessment of motor performance: an empirical approach to identifying motor performance categories. *Arch Phys Med Rehabil* 1991;72:359–66.
20. Haywood KM. *Life span motor development*. Champaign, IL: Human Kinetics Publishers, 1986.
21. Howe S, Levinson J, Shear E, Hartner S, McGirr G, Schulte M, Lovell D. Development of a disability measurement tool for juvenile rheumatoid arthritis. *Arthritis Rheum* 1991;34: 873–80.
22. Hughes JE. *Hughes Basic Gross Motor Assessment (manual)*. Yonkers, NY: GE Miller, 1979.
23. Hutchinson TA, Boyd NF, Feinstein AR, Gonda A, Hollomby D, Rowat B. Scientific problems in clinical scales, as demonstrated in the Karnofsky Index of Performance Status. *J Chron Dis* 1979;32:661–6.
24. Kaplan RM, Bush JW, Berry CC. The reliability, stability, and generalizability of a health status index. *Proceedings of the Social Statistics Sections*. Alexandria, VA: American Statistical Association, 1978:704–9.
25. Karnofsky DA, Burchenall JH. The clinical evaluation of chemotherapeutic agents in cancer. In: McLeod CM, ed. *Evaluation of chemotherapeutic agents*. New York: Columbia University Press, 1949:190–204.
26. Kicklighter RH, Richmond BO. *Children's Adaptive Behavior Scale—revised*. Atlanta, GA: Humanics Ltd., 1983.
27. King-Thomas L, Hacker BJ. *A therapist's guide to pediatric assessment*. Toronto, Ontario: Little, Brown, 1987.
28. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27–36.
29. Klein RM, Bell B. *The Klein–Bell ADL Scale manual*. Seattle, WA: University of Washington Medical School, Health Sciences Resource Center/SB-56, 1979.
30. Klein RM, Bell B. Self-care skills: behavioural measurement with the Klein–Bell ADL scale. *Arch Phys Med Rehabil* 1982;63:335–8.
31. Lansky LL, List MA, Lansky SB, Cohen ME, Sinks LF. Toward the development of a Play Performance Scale for Children (PPSC). *Cancer* 1985;56:1837–40.
32. Lansky SB, List MA, Lansky L, Ritter-Sterr C, Miller DR. The measurement of performance in childhood cancer patients. *Cancer* 1987;60:1651–6.
33. Law M. *Copy of the Klein-Bell ADL Scale with age norms applied*. Hamilton, Ontario: McMaster University, 1992.
34. Law M, Baptiste S, Darswell-Opzoomer A, McColl MA, Polatajiko H, Pollock N. *The Canadian Occupational Performance Measure*. Canada: CAOT Publications, 1991.
35. Law M, Baptiste S, McColl MA, Opzoomer A, Polatajiko H, Pollock N. The Canadian Occupational Performance Measure: an outcome measure for occupational therapy. *Can J Occup Ther* 1990;57:82–8.
36. Law M, Letts L. A critical review of scales of activities of daily living. *Am J Occup Ther* 1989;43:522–8.
37. Law M, Usher P. Validation of the Klein–Bell Activities of Daily Living Scale for children. *Can J Occup Ther* 1988;55: 63–8.
38. Lovell DH, Howe S, Shear E, Hartner S, McGirr G, Schulte M, Levinson J. Development of a disability measurement tool for juvenile rheumatoid arthritis. *Arthritis Rheum* 1989; 32:1390–5.
39. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Med J* 1965;14:61–5.
40. Mazur J, Menelaus MB, Dickens RV, Diog WG. Efficacy of surgical management for scoliosis in myelomeningocele: correction of deformity and alteration of functional status. *J Pediatr Orthop* 1986;6:568–75.
41. McCabe MA, Granger CV. Content validity of a pediatric functional independence measure. *Appl Nurs Res* 1990;3: 120–2.
42. Meenan RF. Health status assessment in pediatric rheumatology. *Rheum Dis Clin North Am* 1987;13:133–40.
43. Milani-Comparetti A, Gidoni AE. Routine developmental examination in normal and retarded children. *Dev Med Child Neurol* 1967;9:631–8.
44. Miller LJ. *Miller Assessment for Pre-schoolers (manual)*. Englewood, CO: The Foundation for Knowledge in Development, 1982.
45. Milstein JM, Cohen ME, Sinks LF. The influence and reliability of neurologic assessment and Karnofsky performance score on prognosis. *Cancer* 1985;56:1834–6.
46. Nunnally JC. *Psychometric theory*. New York: McGraw-Hill, 1978.
47. Orenstein DM, Nixon PA, Ross EA, Kaplan RM. The quality of well-being in cystic fibrosis. *Chest* 1989;95:344–7.
48. Research Foundation—State University of New York. *Guide for use of the Uniform Data Set for Medical Rehabilitation including the Functional Independence Measure for Children (WeeFIM)*. Buffalo, NY: State University of New York, 1991.
49. Russell D, Rosenbaum P, Gowland C, Hardy S, Lane M, Plews N, McGavin H, Cadman D, Jarvis S. *Gross Motor Function Measure manual*. Hamilton, Ontario, Canada: Gross Motor Measures Group, 1990.
50. Russell DJ, Rosenbaum PL, Cadman DT, Gowland C, Jarvis S. The Gross Motor Function Measure: a means to evaluate the effects of physical therapy. *Dev Med Child Neurol* 1989; 31:341–52.
51. Sackett DL, Haynes RB, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine*. Toronto: Little, Brown, 1985.
52. Schipper H, Levitt M. Measuring quality of life: risks and benefits. *Cancer Treat Rep* 1985;69:1115–23.
53. Singh G. *Copy of Childhood Health Assessment Questionnaire and guidelines*. Palo Alto: Stanford University School of Medicine, 1992.
54. Singh G, Brown B, Arhreya B, Goldsmith D, Rettig P, Block D, Fries J. Functional status in juvenile rheumatoid arthritis: sensitivity to change of the childhood health assessment questionnaire [Abstract]. *Arthritis Rheum* 1990;33S:S15.
55. Sparrow S, Balla D, Cicchetti D. *Vineland Adaptive Behavior Scales, interview edition: survey form manual*. Circle Pines, MN: American Guidance Service, 1984.
56. Spitzer WO. State of science 1986: quality of life and functional status as target variables for research. *J Chron Dis* 1987;40:465–71.
57. Steel KO, Glover JE, Spasoff RA. The motor control assessment: an instrument to measure motor control in physically disabled children. *Arch Phys Med Rehabil* 1991;72:549–53.
58. Stower S, Huber CJ. Developmental and screening tests. In: King-Thomas L, Hacker BJ, ed. *A therapist's guide to pediatric assessment*. Boston: Little, Brown, 1987:43–142.

59. Streiner DL, Norman GR. Biases in responding. In: Streiner DL, Norman GR, ed. *Health measurement scales: a practical guide to their development and use.* New York: Oxford University Press, 1989:54–66.
60. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* New York: Oxford University Press, 1989.
61. Stuberg WA, White PJ, Miedaner JA, Dehne PR. Item reliability of the Milani-Comparetti Motor Development Screening Test. *Phys Ther* 1989;69:328–35.
62. Van Wendt L, Svanberg K, Clausen M, Janlert U, Back I, Therell M. Screening for delayed motor development in preschool age children using Statt-Moyes-Henderson's Test of Motor Impairment. *Physiother Can* 1985;37:350–3.
63. Ware JE. Standards for validating health measures: definition and content. *J Chron Dis* 1987;40:473–80.
64. Wesson DE, Williams JI, Spence LJ, Filler RM, Armstrong PF, Pearl RH. Functional outcome in pediatric trauma. *J Trauma* 1989;2929:589–92.
65. World Health Organization. *International classification of impairments, disabilities, and handicaps.* Geneva: World Health Organization, 1980.
66. Wright FV, Law M, Goldsmith C, Dent P. Development of a self-report functional status index for children and teens with juvenile arthritis (JA). *Physiother Can* 1991;43:11 (conference insert).
67. Wright FV, Longo-Kimber J, Law M, Goldsmith C, Dent P. A functional status index for juvenile arthritis (JA). *Physiother Can* 1992;44:6 (conference insert).
68. Wright JG, Feinstein AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *J Clin Epidemiol* 1992;45:1201–18.
69. Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone Joint Surg [Br]* 1992;74:287–91.

APPENDIX

Scale name	Purpose of scale	Domains, format, & scoring	Population	Method of administration	Standardization of measure	Taxonomic class (as per Table 1)
		GENERIC MEASURES				
Functional Independence Measure for Children (WeeFIM) (14,41,48) original ref. cited as (15)	Burden of care; discriminative, evaluative	Degree of assistance required (provided by a caregiver or assistive device) 7-point ordinal scale 6 subdomains: self-care, sphincter control, mobility, locomotion, communication, social cognition 18 questions derived from previous scales	Adult scale modified for children 0.5–7 yr olds Generic population	Trained clinician observation (different sections to be done by specialist clinicians)	No patient data reported. Developers state that face validity and reliability were established in >50 facilities but no reference cited Stated to measure performance, but administration requires clinician observation of capability According to Gowland (14): adequate interrater and excellent intrarater reliability; validity information not reported Manual contains sufficient information for use, but no standardization data (48)	Clinician observation of capability [1.2]
Motor Control Assessment (MCA) (57)	Motor control skills (not functional ability); evaluative	113 items	2- to 5-yr olds Mild to severe physical disability (n = 161, primarily neurologically impaired)	Clinician observation 30–60 min	Validity: correlation with Physical Abilities score = 0.9 Reliability ICCs: intrarater = 0.99, for interrater = 0.97	Clinician observation of capability [1.2]
Tufts Assessment of Motor Performance (TAMP) (12)	Physical function and motor performance; evaluative	3 domains: mobility, ADLs, and physical aspects of communication 32 items, divided into 113 skills Scored on four dimensions: Assistance (5-point ordinal scale), Approach (2 points), Pattern (2 points), and Proficiency (3 points)	6 yrs upward including adults (reliability study n = 20 adults and 20 children) (12) (item grouping study n = 206 subjects 6–86 yrs of age (19) Neurological and musculoskeletal disability	Clinician observation 1 hr, standard equipment	Intrarater reliability using a videotaped assessment exceeded 0.85 (ICC) for all domain/dimension combinations Factor analysis of data on 206 subjects used to determine empirically item groupings: dynamic balance, fasteners, ambulation, manipulation, mat mobility, typing, grasp/release (19)	Clinician observation of capability [1.2]

(Continued)

APPENDIX (*Continued*)

Scale name	Purpose of scale	Domains, format, & scoring	Population	Method of administration	Standardization of measure	Taxonomic class (as per Table 1)
Klein-Bell ADL scale (14,33, 36,37); original adult scale cited as (29,30)	ADL function; evaluative, discriminative	6 domains: dressing, bathing/hygiene, elimination, functional mobility, eating, emergency communication 170 skills items Scores: able, unable, N/A Includes age norms beside each question More upper extremity function items than lower	All ages Test population 10 CP and 10 normals	Clinician observation ~1 h to administer all items	Validity: discriminated between normal and CP subjects $p < 0.001$ Reliability for 5 children interrater ICC = 0.99, test–retest ICC = 0.98 Responsiveness: greater change in normals than CP ($p = 0.08$) and agreement with parental ratings of change gave corrected $\kappa = 0.77$ According to Gowland (14): excellent content validity and reliability, adequate construct validity and responsiveness These conclusions were supported by Law (36)	Clinician observation of capability [1.2]
Barthel Index (14,36). Original ref. cited as (39)	Activities of daily living; discriminative, predictive, evaluative	ADLs Ordinal scale	Applied to adult and adolescent chronically disabled patients	Expert clinician observation 1 h to complete	According to Law (36): Excellent content validity, construct validity, inter- and intrarater reliability Good responsiveness Poor manual and internal consistency According to Gowland (14): Adequate content validity, criterion validity, inter- and intrarater reliability Poor manual No normative data	Clinician observation of capability [1.2]
Karnofsky Scale (23,25,45)	Global rating of physical capacity; evaluative, predictive	Based primarily on mobility level Scoring: 0–100 in 10-unit Guttman intervals	Undefined cancer population (generally poor description of samples)	Physician report 2 min	Weak evidence for validity demonstrated by comparing measure with other clinical criteria Reliability: achieved 29 and 35% agreement between raters Previous reviewers concluded that the scale is not appropriate for children, particularly preschoolers, and that it is unable to predict recurrence in brain tumor pediatric patients (45)	Physician report of capability [1.2] Note: because of the question wording it is possible that performance was reported by some.
Vineland Adaptive Behavior Scales (14). Original ref. cited as (55)	Developmental Assessment Tool Included as example of developmental scale classification	4 domains: communication, daily living skills, socialization, and motor skills (impairment)	0–18 yrs Normative data based on a large sample of disabled children	Trained clinician interview of parent 20–90 min	Reliability and validity reported by Gowland to be excellent	Interview of parent measure of performance [2.11]
Quality of Well-being (47). Original ref. cited as (24)	Quality of life; discriminative, evaluative	3 domains: mobility (5-level ordinal scale), social activity (5-level ordinal scale), physical activity (4-level ordinal scale) Scores weighted according to population preferences Similar to Rand Health Insurance Scale	Adult tool applied to children 25 boys and 19 girls with CF Ages 7–36, mean 16.5 ± 6.9 yrs	Interview administered to parents or patient depending on age	Moderate construct validity assessed by comparing QWB to PFTs and exercise tolerance: QWB/FEV_1 $r = 0.6$ QWB/$FEF_{25-75\%}$ $r = 0.5$ QWB/PEFR $r = 0.4$ QWB/Vo_2 max $r = 0.6$	Interview-administered measure of performance [2.11]

(*Continued*)

APPENDIX (*Continued*)

Scale name	Purpose of scale	Domains, format, & scoring	Population	Method of administration	Standardization of measure	Taxonomic class (as per Table 1)
Canadian Occupational Performance Measure (COMP) (14,34,35)	Evaluative Subjects generate their own items Useful for comparison within individual patients rather than between patients	Domains: self-care, productivity and leisure Dimensions: importance of activities, level of performance & satisfaction with performance Scoring: 10-category ordinal scales	Not age specific Developed for adults and applied to children	Clinician administered Items spontaneously elicited from each patient Interview of parent or child Considers environmental demands	No evidence of validity or reliability included (14)	Interview measure of capability and performance [2.11 & 2.21]
Pediatric Evaluation of Disability Inventory (PEDI) development edition (10,17,18)	Physical Function & Independence Measure; evaluative Expert reviewers preferred to class the tool as discriminative rather than evaluative (17)	Domains: self-care, mobility, social function Scoring dimensions: functional capability, caregiver assistance, environmental modifications Scored able/unable for 197 functional skill items, 6 ordinal responses for 20 caregiver assistance items, 4 types of environmental modifications for 20 items	Chronically ill and disabled children 0.5–7 yrs old	Parent report 20 min to 1 h to complete	Concurrent validity: moderately high correlations with Battelle Developmental Inventory Screening Test for self-care and mobility domains but not social function domains (10) Significant differences between normals and disabled samples (10) Content validity and reliability studies are currently underway, but unpublished normative data (sample of 412) not yet published (10) Manual (18) includes a detailed report of standardization, which is adequate in all areas, discussion of scaling methods, describes calculation of standardized scores	Parent report of performance [2.12]
Play Performance Scale (31,32)	Play; evaluative	Concepts based on Karnofsky's Scale Based on active play, quiet play, degree of physical limitation, degree of independence Scored 0–100 in 10-point increments	1–16 yr olds Brain tumors (n = 98 oncology patients, 8 ± 4.71 yrs of age; 29 siblings, 8.76 ± 4.42 yrs; 40 normals, 8.59 ± 4.98 yrs) (32)	Parent report <5 min to complete	Interrater reliability (mother vs. father) r = 0.71, n = 41 Construct validity: detected significant difference between patients and siblings, and significantly related to global measure of performance from nurses and researchers (r = 0.75 and r = 0.92, respectively) (32)	Parent report measure of performance [2.12]
Rand Health Insurance Study Scale (HIS) (6,7)	Physical activity; discriminative	4 domains: mobility, physical activity, role activity, and self-care Children's tool similar to AIMS (7)	Ages 0–13 yrs (n = 2,152 children in 6 U.S. cities) (7) Ages 4–16 yrs (n = 3,294 children in Ontario) (4) Healthy populations (n = 156 pediatric trauma survivors age 8.7 ± 4.4) (64)	Researcher administered Parent report	Construct validity: comparison of HIS classification of able/disabled with 11 other scales showed significant differences for all 11 scales; however, actual differences were small and the sample large Found a 57/1,000 prevalence of disability (4) Wesson et al. found the HIS not to be able to discriminate severity in a trauma population (64) Administration manual (6)	Parent report measure of performance [2.12]

(Continued)

APPENDIX (*Continued*)

Scale name	Purpose of scale	Domains, format, & scoring	Population	Method of administration	Standardization of measure	Taxonomic class (as per Table 1)
Childhood Health Assessment Questionnaire (CHAQ) (53,54)	Functional status; evaluate functional status	8 domains: dressing & grooming, arising, eating, walking, hygiene, reach, grip, and activities 4-point ordinal scale for each item Unusual method of aggregation	JRA (n = 62) 1–19 yr olds Derived from an adult tool	Parent or patient self-administered 10 min	No documentation of validity or reliability in children Responsiveness: CHAQ was a significant predictor of parents' global rating of change $p < 0.02$ Report that "parents are reliable proxy reporters of their children's functional status" but no data or reference are provided to support this	Parent or self-report of capability [2.22]
		DISEASE SPECIFIC—NEUROLOGICAL[a]				
Gross Motor Function Measure (GMFM) (49,50)	Gross motor abilities; evaluative, discriminative	85–88 items Gross motor skills Assesses the quantity of skill capability not quality 4-point ordinal scale per item Items equally weighted within dimension Domains: lying, sitting, crawling & kneeling, standing, and walking-running-jumping Scoring clearly described with examples	Cerebral palsy (n = 111) and acute head injuries (n = 25) (control subjects were 34 normals <5 yrs old)	Trained clinician observation Highly structured assessment Time estimated at >1 h	Validity: moderate correlations with clinicians' judgments Reliability: intrarater correlations >0.92, interrater correlations >0.87 Responsiveness: significant difference in scores of subjects who had changed and no significant difference in the scores of patients whose parents rated them as unchanged Manual (49) includes a detailed report of standardization, which is adequate in all areas	Clinician observation of capability [1.2]
Gross Motor Performance Measure (GMPM) (2,3)	Qualitative aspects of gross motor function; evaluative	Assesses the quality of capability Domains: alignment, coordination, dissociated movement, stability, weight shift Impairment	Cerebral palsy	Trained clinician observation	Content validity assessed as good by expert rating: mean rating = 4.0/5.0 Usefulness as an evaluative tool: mean rating = 4.2 Further work in progress Note: developers define performance as what the child does when observed in the clinic, which equates to capability by our terminology	Clinician observation of capability [1.2]
Seated Postural Control Measure (SPCM) (11)	Seating function	Postural alignment = 22 items Functional movements = 12 items	1–19 yrs of age Neurologically impaired (n = 45)	Clinician observation 20 min 3- or 4-level ordinal scale for each item	Report face validity but no details of methodology Reliability: interrater κ statistics; 0.45 for alignment, and 0.85 for function Test–retest reported as r = 0.35 for alignment and 0.29 for function (poor)	Clinician observation of capability [1.2]
Children's Adaptive Behavior Checklist (CABS) (14) original ref. cited as (26)	Adaptive behavior; discriminative, evaluative	5 subdomains: language, independent function, family role performance, economic vocational activity, and socialization Ordinal scale	5–11 yr olds Developmentally disabled	Trained psycho-educational evaluator Interview of child 45 min	According to Gowland (14): Excellent inter- and adequate intrarater reliability Adequate content, construct, and criterion validity	Interview-administered measure (not clear whether performance or capability) [2.11 or 2.21]

(*Continued*)

APPENDIX (*Continued*)

Scale name	Purpose of scale	Domains, format, & scoring	Population	Method of administration	Standardization of measure	Taxonomic class (as per Table 1)
		DISEASE SPECIFIC—ARTHRITIS[b]				
Juvenile Arthritis Functional Assessment Scale (JAFAS) (38)	Speed of physical function; discriminative Note: development selected only those items that could differentiate between cases and controls	10 items from previous scales Scored based on time to complete activities compared to a criterion (time for control subjects + 2 S.D.) 0 = time < criterion 1 = time > criterion 2 = unable to do	JRA (n = 71) Ages 7–18 Polyarticular (35%), pauciarticular (57%), systemic (28%) 63 normal control subjects ages 7–17	Clinician observation 10 min to administer	Convergent validity: with number of involved joints r = 0.4, p = 0.003, with Steinbrocker class r = 0.59, p = 0.0001, with disease activity r = 0.32, p = 0.01 High internal consistency: Cronbach's α = 0.85 Reliability not reported Found 5 items to vary directly with age, but effect on overall score was not statistically significant	Clinician observation of capability [1.2]
Juvenile Arthritis Functional Assessment Report (21) For children: JAFAR-C For parents: JAFAR-P	Independent performance of activities; discriminative	23 items Scored based on the frequency with which the child was able to independently perform the activity during the previous week 0 = all of the time 1 = sometimes 2 = almost never	Same as above with the addition of parents (JAFAR-P not administered for controls)	JAFAR-C interviewer-administered to children JAFAR-P self-administered to parents	Construct validity: differences between patients and controls (p < 0.001) No differences between patients' and parents' reports (p = 0.54) No significant age correlation Correlation of JAFAR With JAFAS, child = 0.69, parent = 0.69 With Steinbrocker function class, child = 0.49, parent = 0.60 With disease activity, child = −0.43, parent = −0.42 With number of joints involved, child = 0.44, parent = 0.49 Cronbach's α, child = 0.85, parent = 0.93	JAFAS-C self-report of performance [2.12] JAFAS-P parent report of performance [2.12] Note: because of the question wording it is possible that capability was reported by some.
Juvenile Arthritis Self-report Index (JASI) (66,67)	JRA physical function; evaluative, discriminative	Domains: self-care, domestic, mobility, school, extracurricular Scored on 7-point Likert for each item	Intended for juvenile rheumatoid arthritis patients Ages 8–18 (n = 30)	Self-report 272 items (current version has been reduced to 100 items)	Face validity assessed by 17 clinicians Test–retest reliability ICC = 0.99 (67) Validity assessed relative to joint count and grip strength with correlations of 0.51 and 0.64, respectively (67)	Self-report of capability [2.21]
Arthritis Impact Measurement Scales (AIMS) (5)	Physical limitations; discriminative	2 dimensions: physical disability (modified), pain 9 items (from the original 45 items) 2–6 response options per item	Adult tool, modified for children with juvenile arthritis (n = 77) Ages 2–17 yrs, mean = 9.3 yrs (70% girls) (97% white)	Clinician administered Interview of parents in a clinic setting	Coulton et al. state attainment of convergent validity; however, this conclusion is not well supported by data Correlations with diagnostic category = 0.24–0.26, and joint count = 0.31–0.35 Discriminated between active and inactive disease groups at p < 0.01	Interview measure of capability [2.21]

[a] See also WeeFIM, MCA, PEDI, TAMP, and COPM.
[b] See also CHAQ and COPM.