Prosthetic Outcome Measures for Use With Upper Limb Amputees: A Systematic Review of the Peer-Reviewed Literature, 1970 to 2009

Virginia Wright, PT, PhD

ABSTRACT

The overall goal was to perform a systematic literature search and structured review of the state of outcome measurement in the field of upper limb prosthetics and to propose future directions for research in this area. The review is based on systematic searches of Medline, Cumulative Index to Nursing and Allied Health Literature, and RECAL electronic databases from 1970 to March 2009 and on subsequent review of the reference lists of the identified publications (citation tracking), as well as articles in the author's personal collection. Each article was independently reviewed by two readers. Publications were initially screened for relevance, and then fully reviewed to extract relevant data and review the methodological quality of the study. A quality rating form, based on guidelines provided by Terwee et al. (J Clin Epidemiol. 2006;60:34–42), was devised for evaluation of the clinimetric properties of the identified outcome measures. The search identified 660 peer-reviewed publications related to outcome measurement with upper limb amputees of which 25 met all of the author's inclusion criteria for full review. In the adult literature, seven outcome measures (4 amputees specific and 3 generic) were revealed. This compares with 25 measures identified in the lower limb adult amputee outcomes review by Condie et al. (J Prosthet Orthot. 2006;18:13–45). In the pediatric review, nine distinct outcome measures (5 amputees specific and 4 generic) were found. Two of the pediatric measures also have younger child versions. There was overlap of one measure, the Assessment of Capacity for Myoelectric Control (ACMC), between adult and pediatric studies. The use of standardized outcome measures with adult upper limb amputees is sparse in the published studies with this clinical population, and upper limb prosthetic-specific measures are few in number. Attention needs to be paid to all aspects of measurement development and validation across the International Classification of Functioning, Disability, and Health. The measures with greatest psychometric promise for use in upper limb prosthetics are the ACMC, the Upper Extremity Functional Status Module of the Orthotics and Prosthetics User Survey, the Disabilities of Arm, Shoulder, and Hand Questionnaire, and the Trinity Amputation and Prosthesis Experience Scales. Greater strides toward measure development and validation have been made with pediatric upper limb amputees. The emphasis that is currently needed is on the determination of the test-retest reliability and responsiveness of the most promising measures (ACMC, University of New Brunswick Test, the Assisting Hand Assessment, Prosthetic Upper Extremity Functional Index, and the ABILHAND-Kids) and discussion on how best to approach the measurement of participation and quality of life. (J Prosthet Orthot. 2009;21:3–5.)

The need for systematic and comprehensive approaches to outcome measurement in acute care and rehabilitation is currently the subject of much discussion and work internationally. There is an increasing recognition that rehabilitation practice is most comprehensive when it is grounded in the framework of the International Classification of Functioning, Disability, and Health (ICF), meaning that both interventions and outcome indicators need to take the ICF into account. In addition to consideration of outcomes related to the ICF components of "body structure and functions" (impairment as measured in areas such as range of motion, muscle strength, and movement functions), "activity" (carrying out tasks, including arm and hand use, within daily activities), and "participation" (involvement of the individual in a diversity of life situations), this also means thinking about the characteristics of the client (adult or child), their family, and the environment, as well as the overarching quality of life (QOL). The ICF framework also embraces the notion that interventions can be directed toward making changes in the client and/or to changing the people, systems, and environments, which are connected with the individual.

Ramstrand and Brodtkorb, in their article on evidence-based practice in prosthetics and orthotics, describe a cultural change that needs to take place such that research and clinical practice inform each other, and so that orthotists and prosthetists are more involved in the generation of knowledge and its transfer. This means incorporating evaluation of outcome into both the clinical and research arenas. A recent review of literature from 1980 to 2006 by Bidiss and Chau on roles of predisposing characteristics, need, and enabling resources on upper limb (UL) prosthesis use and abandonment underlines the importance of adoption of standardized outcome measures. Comprehensive measurement should facilitate the understanding of the impact of age of fitting, prosthetic technology, efficacy of training protocols, and health care services on prosthesis use and abandonment and should allow international comparisons of prosthetic outcomes.

The complexity of building and implementing systems of core outcome measures, which address the diversity and needs of clients who are receiving services, is very evident. In a 2004 publication, Davidson (an occupational therapy expert in the field of UL prosthetics) noted that there were no measures available to specifically evaluate the functional abilities of UL adult amputees. She used the pediatric Prosthetic Upper Extremity Functional Index (PUFI) as the example of what might be helpful as far as a self-report measure in adult prosthetics and noted that Hermansson et al. were validating an observational measure (the Assessment of Capacity for Myoelectric Control [ACMC]) for myoelectrics that was tailored to use in both adult and pediatric prosthetics. Davidson also stressed the need to also look at the overall disability for adults with UL amputation and proposed that the Disabilities Of Arm,
Shoulder, and Hand (DASH) questionnaire, UL measure of disability for individuals with UL musculoskeletal injury, might be suitable in focus to address this gap.

GOALS AND OBJECTIVES OF THE PROJECT
The overall goal of this project was to perform a systematic literature search and structured review of the state of outcome measurement in the field of UL prosthetics from 1970 to March 2009 inclusive, to establish the current state of measurement, and to propose future directions for the prosthetic field. The specific objectives were to:

i. Derive a comprehensive list of outcome measures that have been used in the English language peer reviewed publications in the area of adult and pediatric UL prosthetics;
ii. Identify the measurement focus and format of each of the measures listed;
iii. Determine the relative strengths and weaknesses of each of the reviewed measures;
iv. Appraise the level of evidence associated with each outcome measure according to a set of rating criteria developed for this review;
v. Identify gaps in measurement according to the ICF framework; and
vi. Make recommendations about the outcome measures that should be used for pediatric and adult UL prosthetic rehabilitation.

METHODS

SELECTION OF PUBLICATIONS
The review was based on systematic searches (as described below) of the Medline electronic database, the Cumulative Index to Nursing and Allied Health Literature (CINAHL) electronic database, and RECAL electronic database (RECAL Legacy database from the University of Strathclyde, cdlr.strath.ac.uk/recal/journaltitles.htm) from 1970 to March 2009 and on subsequent review of the reference lists of the identified publications (citation tracking) as well as articles in the author's personal collection. These search strategies parallel and expand on the ones used by Condie et al. in their systematic review of the lower limb prosthetic outcome measures literature. The search extended back to 1970 to ensure provision of a clear picture of the development of outcome measurement in the field.

Medline and CINAHL search strategies were as follows: (outcome assessment [health care] or questionnaires or quality of life or health status or treatment outcome or health status indicators or participation) and (amputation) and (upper extremity or arm or upper limb) and (prosthesis or prostheses or artificial limbs) and adult or adults.

For the child measures, we did the same search along with child search terms: ... and (pediatrics or pediatric or child or children).
In addition, we searched Medline and CINAHL for myoelectric and (amputees or prostheses or upper limb deficiency). This helped to identify several myoelectric articles that did not come up with the larger search approaches.

The RECAL database search strategies consisted of each of the following:
- Amputation and (upper limb or upper extremity) and instrument and evaluation;
- Amputation and outcome and instrument;
- Amputation and outcome and patient satisfaction;
- Amputation and outcome and (prediction or indication);
- Amputation and outcome and (predictor or indicator);
- Amputation and (upper limb or upper extremity) and outcome and rehabilitation;
- Amputation and health status and (measure or assess or examine or evaluation);
- Amputation and instrument; and Upper limb and (prosthetic or prostheses or prosthesis) and instrument.

SELECTION OF PUBLICATIONS
Publications were included in this review if:

1. They described one or more outcome measures that:
   a. Were developed for use with amputee patients/subjects or individuals with related UL musculoskeletal conditions, and
   b. Were intended to measure functional ability of the hand or entire UL, participation, or quality of life, and
   c. Were designed (or used) to evaluate or predict outcome.

2. The study population included UL amputees
   a. With unilateral or bilateral amputation,
   b. At any level of limb loss (i.e., from loss of a digit to shoulder disarticulation), and
   c. From any cause (i.e., congenital or acquired).
3. They included a sample of at least 20 participants who had an UL deficiency.

4. They were written in (or translated into) English.

Publications were excluded from this review if:

1. The study described an outcome measure specifically for use with UL neurological conditions,
2. They were published as a dissertation, thesis, book chapter, or conference proceedings,
3. The full publication was unavailable to the authors, or
4. The measure used was primarily for evaluation of wearing pattern, satisfaction with the prosthesis or evaluation of pain, or one that contained less than 30% of its items related to function, participation, or QOL.

These criteria are an adaptation of the ones used by Condie et al.\textsuperscript{11} as a way of maximizing comparability between their lower limb outcomes review and this UL amputee review. The area of participation outcomes was added to the search, as this is a component of the ICF that is receiving increased attention. Separate searches were made for the pediatric and adult literature, so as to ensure capture of all possible outcome articles. At the outset of our work, we anticipated a difference in the level of development and validation of prosthetic-specific outcomes in the adult and pediatric work, i.e., more in pediatrics,\textsuperscript{12} and, thus, planned from the outset to complete parallel reviews within these two subgroups.

Given the paucity of publications found on outcome measurement with UL amputees overall, we decided after our first round of reviews to be more liberal than Condie et al.\textsuperscript{11} as far as inclusion criteria. Thus, an article that had at least five participants with an UL amputation was eligible for review as long as it contained at least one outcome measure that had undergone extensive validation in another clinical population (revision of Condie et al.’s criterion viii). It was anticipated that these small number of studies would add to the body of knowledge about the measure. In the case of a measure that was not used in more than one amputee-related article, single publications were felt to at least show the potential (or lack of potential) of a particular measure for use with UL amputees.

After the first round of searching and the realization of the need for greater comprehensiveness in the outcome measure-related key words that had been used in the publications, the next step was to search lists of measures used to evaluate musculoskeletal-related dysfunction of the hand (fine motor) or UL.\textsuperscript{13–15} This was done as well from the references of several review articles of pediatric\textsuperscript{12,16} and adult outcomes measures\textsuperscript{12} in UL prosthetics. For the pediatric component, hand (fine motor) and UL measures that were summarized in the pediatrically-based “All About Outcomes Software”\textsuperscript{17} were also reviewed.

When a measure was identified, a specific search was undertaken through the Medline and CINAHL databases using the measure’s name, e.g., “Pediatric QOL Inventory (PedsQL™).” The titles of all articles that listed for the measure were screened, and the abstracts of any that connected with prosthetics or with related conditions such as radial deficiency, obstetrical brachial plexus palsy, and UL bone tumors were reviewed. The search was then narrowed by linking with the terms amputee, prosthetics and UL deficiency (separately and in combination) to see whether any additional articles with amputees were uncovered.

REVIEW METHODS
The title and abstract of each of the publications identified in the literature search were reviewed by the author of this article to provide an initial determination of eligibility. Full-text versions of all articles that appeared to meet the inclusion criteria were obtained. The author of this article and her research assistant then independently reviewed each article in terms of the eligibility criteria. An article had to fulfill all criteria before it was accepted for detailed review. Each article had to be accepted by both the reviewers to be included in the detailed psychometric review.

DATA EXTRACTION
The six review tables (Tables 1–6) that were created followed the format and review category headings of Tables 2 to 7 in the lower limb review by Condie et al.\textsuperscript{11} As in Condie’s article, they cover six areas: general information about the study, practicality of completion, outcome measure reliability, validity, scaling, and potential for bias. Within the tables, three new evaluation categories were added for this review (i.e., an intrarater reliability column was added, because it is important to differentiate this from test-retest reliability; a Rasch measurement column was added; and a final column was added to the last table to present summary comments on notable strengths and limitations of each study). One category was reworded slightly (i.e., the study authors’ comments on suitability were expanded to also include the author's main conclusions), and the construct validity category was expanded to also include discriminant validity. Two review categories were removed (i.e., time from amputation did not have any bearing on the outcome review information and rarely was available, and comparison between responders and nonresponders was rarely used in Condie et al.’s review and did not seem to have relevance in this review).
We started with the same basic definitions and criteria for reliability, validity, and responsiveness categories as Condie et al.\textsuperscript{11} did and inserted some additional expectations to reflect recent changes in measurement practice. The checklist developed by Jerosch-Herold\textsuperscript{18} in 2005 for review of outcome measures and outcome measure review guidelines by CanChild Centre for Childhood Disability Research\textsuperscript{19} were used as the frame of reference/criteria for the review of study characteristics and findings. Details of study aim, study design, primary outcome measures, participants (age, diagnosis, and level of amputation), results, sample size, study limitations were among the characteristics that were independently summarized by the two reviewers directly into the study summary tables (Table 3). All discrepancies in the details extracted and review comments were discussed and resolved.

Entry of data into the review tables was done for each publication that was reviewed. In addition, each outcome measure that was used was given a separate data entry line, as done by Condie et al.\textsuperscript{11} This means that some studies appear several times in the review tables. This approach allowed easy separation of the results for each outcome measure. Individual consideration of measures was essential for the subsequent quality rating of the various measures. For study design and practicality review details, test information that was not provided in the article or was unclear was described as not stated, whereas measurement properties that were not studied in an article were described as not evaluated. A mention was made in the table if the authors indicated other work in which the particular measurement property had been studied.

**QUALITY RATING OF THE OUTCOME MEASURES**

After all of the publication reviews were completed, quality assessment ratings were done independently by the two reviewers for each of the identified measures (i.e., summarized across the data from all related studies). This was done according to the review form that was developed by this review (Table 7). It is important to make a few points on the systematic reviews with respect to outcome measurement. When one thinks of systematic reviews, this is typically in the context of clinical trials. There are a number of documented approaches to systematic reviews: American Academy of Cerebral Palsy and Developmental Medicine (AACPDM) criteria,\textsuperscript{20} the PEDro scale for rehabilitation research,\textsuperscript{21} and the American Academy of Orthotists and Prosthetists (AAOP) criteria.\textsuperscript{22}

The challenges of doing systematic reviews for outcome measures are evident in several recent publications. For example, an article in *Developmental Medicine and Child Neurology* (2008), a highly-ranked pediatric rehabilitation journal, described a systematic review of psychometric properties of neuromotor assessments for preterm infants.\textsuperscript{23} Although the search strategies were conducted in a systematic manner and each of the articles related to a particular measure was presented and evaluated with related articles in tables, there was no attempt at rating the quality of the psychometric evidence. A similar situation existed for a recent systematic review article on activity limitation measures.\textsuperscript{24}

Condie et al.,\textsuperscript{11} in their systematic review of lower limb prosthetic outcome measures, identified a number of difficulties in rating the quality of outcome studies and, therefore, chose not to attempt it. The authors noted that an outcomes rating form had been published by Jerosch-Herold\textsuperscript{18} just as they were finishing their work and suggested it might be worth using for subsequent reviews. This form takes the approach of rating each outcome article on its own rather than as a synthesized-evidence rating of a measure across validation articles.

There is, however, some promising new work in the development of scores for systematic rating of outcome measures.\textsuperscript{25} In their review of QOL measures, Tenwee et al.\textsuperscript{25} suggested that specific rating criteria are needed, analogous to those that are in use for evaluating the strength of evidence from clinical trials. They identified measurement attributes of primary interest when evaluating the quality of health status questionnaires: content validity, internal consistency, reliability, criterion validity, construct validity, responsiveness, floor and ceiling effects, interpretability, and respondent and administrative burden (clinical utility). They chose not to create a quality summary score by arguing that some aspects of the evaluation form should carry greater weight than others, e.g., content validity (measurement aim, concept, and item selection/reduction) is one of the most important attributes, whereas other attributes may vary in importance depending on the purpose of the measure. Instead, they operationalized definitions of their quality ratings (+, 0, and ?) and prepared a measurement chart that profiled each of the measurement attributes and each measure’s ability to meet the standard in that area. One publication, which used this outcome measure rating scheme,\textsuperscript{26} focused on a group of questionnaires for patients with osteoarthritis. Because Tenwee et al.\textsuperscript{25} did not consider evaluation of Rasch measurement processes in their form, specific guidelines for the review of Rasch measurement from a publication by Tenant and Conaghan\textsuperscript{27} were added to the quality review form for the review of this article.\textsuperscript{25}

Following the examples of Tenwee et al.\textsuperscript{25} and Veenhof et al.,\textsuperscript{26} a decision was made not to do ratings of each outcome-related prosthetic publication for this UL review, but rather to conduct a synthesis across articles for each outcome measure. Also, in line with the work by Tenwee et al.\textsuperscript{25} was the decision not to attempt to create a summary quality score because, as noted above, the various evaluation categories likely have different weighting as far as psychometric importance.\textsuperscript{25} Instead, the summary charts of ratings of excellent (+ +), good (+), fair (?), and poor (0) for the various measurement attributes of interest (Table 8 and Table 9) provide the reader with a visual picture of the areas in which validation work has been conducted in prosthetics (summarized across publications) and the nature of that work in terms of methodological strength. All ratings were done with respect to evidence of psychometric properties when used in UL prosthetics. Results of psychometric testing from other clinical populations were not noted in these tables. Absence of amputee-specific evidence about a particular psychometric property meant that the rating cell was left blank to highlight the gaps in psychometric evaluation.
RESULTS

GENERAL INFORMATION
The electronic searches revealed approximately 660 publications in the initial database searches. The author's personal file and citations from references of relevant articles and from citation link on Medline resulted in an additional 50 relevant articles for review. Review of titles and abstracts of all of the UL articles narrowed the list down to 74 potentially relevant articles, and following full review of these publications, a final group of 25 unique articles was reviewed in depth (Tables 3 to 6, 10 and 11 -see links above). There were 52 articles considered from the adult study search, and a final group of 12 met the criteria for full evaluation. In the pediatric search, 25 publications reached the short list, and of these, 16 satisfied the criteria for full evaluation. The pediatric list includes three combined adult/pediatric articles that were also included (counted) in the adult review. It also included a nonpeer reviewed publication on the University of New Brunswick (UNB) Test (the source article for all other authors/researchers who used it). The 49 articles excluded after first review are listed in Table 2 along with reasons for their exclusion.

The searches performed in the Medline, CINAHL, and RECAL databases, as outlined in the methods section, required multiple search strategies and innovative combinations of search words. The difficulties encountered in identifying prosthetic outcome articles from the databases may have been, in part, due to lack of use of outcome measure-related terms in the key words given in the publications.

DESCRIPTION OF MEASURES AND SUMMARY OF CHECKLIST RESULTS
The following sections present the results from the outcome measures reviewed, using the assessment details found in Tables 3 to 6, 10, and 11 (adult and pediatric measures - see links above). In the adult literature, seven outcome measures (4 amputees specific and 3 generic) were revealed. In the pediatric review, nine distinct outcome measures (5 amputees specific and 4 generic) were found. These were categorized according to four main areas of outcome measurement-specifically hand function, UL functional abilities, overall functional abilities/participation, and QOL. Measures pertaining to UL functional abilities were defined as those with a scoring focus that extends beyond the evaluation of specific methods of hand function (e.g., gripping, releasing, holding, and manipulating) and instead consider the use of the UL as an integrated unit for each activity. The measures of hand function tend to span the ICF components of body structure and functions and activity, whereas those related to UL function are situated within the ICF's activity component with some overlap into participation.

Within each outcome area, the results are categorized according to prosthetic-specific measures and generic (crossdiagnosis) measures. Each measure has its own section in which it is described, after which the results of use in prosthetics are discussed. Time to complete the measure, a key aspect of clinical utility, is reported whenever it was noted in one of the associated publications. All of the adult-study results are presented in section one of this report, followed by the pediatric results in section two. The reviews of the quality ratings of each of the outcome measures are presented in Table 8 (adult) and Table 9 (pediatric).

SECTION ONE: MEASURES USED WITH ADULTS WITH UPPER LIMB AMPUTATIONS

Measures of Hand Function

Amputee specific: Measures of prosthetic control
Assessment of Capacity for Myoelectric Control The ACMC is an observational measure designed in Sweden by Hermansson et al. and is intended specifically for the measurement of a child's or adult's capacity for control of a myoelectric prosthesis. The test is available in English, Swedish, and Dutch. The ACMC is a Rasch-built scale, and data are analyzed through a FACETS Rasch-measurement program.

Its 30 gripping, releasing, holding, and coordinating items are evaluated on a 4-point capability scale that measures the quality of performance. These actions are evaluated within the context of an UL functional activity that is considered by the patient as meaningful to them, e.g., cooking a meal, doing a craft, and playing a game. It is unclear from the publications whether the test can be scored live or has to be done from video. Although the test does not require specific supplies, completion of an intensive ACMC, 2.5–day, training workshop and test is required before the ACMC is used for clinical or research work. Information on the ACMC, training workshops, and scoring software for trained users can be accessed at the measure's website.

The published work with the ACMC consists of three validation studies by its developer and colleagues. One of the prime strengths of this work is the use of recognized Rasch methods to address scaling of the ACMC. This positions the ACMC well as far as its potential to measure change. Internal consistency and interrater reliability were studied comprehensively. The results indicate that reliability is highest for clinicians who have extensive myoelectric training experience. This high-skill demand is a key limitation of the test and may deter clinicians/clinics from its use. There is also preliminary evidence in this work from a subsample of seven patients that the ACMC has the ability to detect change in status, i.e., functions as an outcome measure.

Subsequent related measurement work by Lindner et al. confirmed the discriminant validity and unidimensionality of the ACMC and recommended that its rating scale structure could be further improved by collapse of several items and revision of the category-2 definition (capable on request). Although the authors suggested removing the rater's request part of this rating option to account for tester variation, it was not entirely clear to the unininitiated reader what this change entails. The impact of task difficulty on the
functioning of items was also noted as an area that requires further study. Work on construct validity is clearly needed, so that we have an idea how the ACMC results fit with other outcome areas. Furthermore, evaluations of responsive to change and validity that are done separately in pediatric and adult samples will be valuable because the research thus far has used a combined sample.

Southampton Hand Assessment Profile. The purpose of this observational assessment, developed in Britain by Light et al., is to determine the effectiveness of a terminal device with respect to unilateral prosthetic hand function. The Southampton Hand Assessment Profile (SHAP) contains 26 self-timed prehensile tasks of which 12 are abstract-unilateral tasks with form board objects, and 14 are activities of daily living (ADL) tasks. Although several of the ADL tasks are bilateral, the sound hand must be used as a stabilizer. Six prehensile patterns are represented in these tasks. All of the tasks involve very little arm movement, so that its prehensile ability is primarily assessed. The rationale for using a measure of time-to-complete as an indicator of ability is that if an individual uses an abnormal grip to perform the task, they are expected to take longer time to do it. The resulting Index of Functionality takes into account the subject’s prehensile pattern and time required. The SHAP requires about 20 minutes to complete. It comes as a portable kit with a standardized administration protocol and can be obtained from its developers. The authors indicate that it is a measure also suitable for use with individuals with different clinical groups such as arthritis, stroke, burn, and wrist/hand injury.

The SHAP developers reported on the assessment’s discriminant validity by comparing able-bodied adults and UL prosthetic users. The SHAP detected significantly impaired hand function among UL prosthetic wearers in comparison with norms (Index of Functionality scores). Although there was evidence of short-term test-retest reliability (minimum of 24 hours between three repeated tests) and interrater reliability in able-bodied young adults, test-retest work should be replicated with prosthetic users before clinicians can be confident about this property with their clients with UL amputations. No further interrater reliability work is indicated as the SHAP is a self-paced test, and the rater’s role is confined to instructions and recording the time. These factors would not be expected to be affected by the patient’s diagnosis. There are no published reports of use of the SHAP with any other patient groups beyond the discriminant-validity work noted above. There is one separate report of use in able-bodied individuals that reported the SHAP’s ability to distinguish between dominant and nondominant hand abilities.

Two unvalidated tests of prosthetic hand function (not included in the review tables). One test worth noting in this section is the Carroll Observational Test used by Graham et al. in their comparison of the hand function of adult belowelbow amputees (predominantly myoelectric users) with those of individuals who had arm replantation posttraumatic injury. The original Carroll test consists of observation of ease performance of 33 skills, representing 7 key hand function areas (grasp, grip, lateral prehension, pinch, placing, supination and pronation, and writing). Scoring is done on a 4-point degree of difficulty scale. Modifications were made by Graham et al. to eliminate the 12 individual digit-dependent prehensive tasks (e.g., grasp a marble with ring finger and thumb), so that this test would be suitable for evaluation of performance with a prostheses. In this modified version, it was able to detect differences in manual skills levels of individuals in the amputee group. However, there has been no specific validation work carried out on the modified Carroll test or is there evidence that it has been used again in amputee populations. Hence, it was not included in our full review.

Similarly, Lake adapted the pediatric version of the UNB Test by creating a new response scale on the efficiency of prosthesis use and adding new tasks for evaluation. Although the reliability of the modified test was not evaluated, Lake did demonstrate changes in performance of individuals (n = 5) using their UL before and after a set of functional-training session compared with that of a nontherapy control group (n = 5). No further work on validation of this modified UNB test for adults has been published.

Generic hand function measures

The infrequent mention of use of a generic hand function measures in the form of timed tests such as the Box and Block Test and the Jebsen Standardized Test of Hand Function was notable and unexpected. Although tests of manual dexterity (formal and informal) have been used in single- or multiple-case prosthetic research and development work, their use in larger group clinical studies was rare. The unpublished exception was a hand function outcome study by Goldfarb et al. that did not meet this report’s prosthetic user inclusion criteria. In this study of individuals with radius dysplasia, the Jebsen Test was used alongside the DASH outcome measure to evaluate hand function after centralization surgery. Because specific validation work has not been carried out in amputee populations with any of the manual dexterity tests, these generic measures are not presented in the review tables.

Measures of Upper Limb Functional Abilities

Amputee specific

Revised Upper Extremity Functional Status module of the Orthotics and Prosthetics User Survey. The development and initial validation of the Orthotics and Prosthetics User Survey (OPUS) were reported by Heinemann et al., its developers, in 2003. The original OPUS consisted of a lower limb functional status module, as well as satisfaction and healthrelated QOL (HRQOL) modules. Psychometric testing was done with prosthetic and orthotic users (combined sample of adults and children). It is not clear from this work that how many of these individuals had UL deficiencies, but the authors noted that the content focus at that time on lower limb functional status was due to evaluation limitations posed by the small population of persons with UL loss.

Subsequent unpublished work on the OPUS resulted in the creation and initial testing of a Upper Extremity Functional Status (UEFS)
module. This module is made up of questions pertaining to an individual's performance of 23 self-care- and instrumental UL-based daily living skills. The patients score their abilities on each item according to a 5-point degree of difficulty scale. More recent collaborative work between Heinemann and Burger et al. in Slovenia\(^\text{43}\) resulted in creation and initial validation of a Rasch-based, revised module of the UEFS. This work included the addition of a response option that determines whether the patient uses the prosthesis for the activity, and the development of a Slovenian-language version. The Rasch analysis identified four misfitting items, two of which were bilateral activities. These were removed, and the final item set for the modified UEFS consists of nine activities that are identified by the authors as purely unilateral (unaffected hand), seven that are typically bimanual but can be done in a unilateral manner, and seven activities that are truly bilateral. The original 5-point response scale was ultimately reduced by Burger et al. to a more reliable 4-point scale by collapsing the "very difficult" and "slightly difficult" categories into a single-response option. The satisfaction and QOL modules of the OPUS were not evaluated by this research team. The modified UEFS can be accessed through its developer, Burger.\(^\text{43}\)

**Generic**

Disabilities of the Arm, Shoulder and Hand Outcome Measure.\(^\text{7}\) The DASH is an UL-focused functional status questionnaire that was developed through Canadian-US collaboration. It was designed to evaluate symptoms and disability in individuals with UL musculoskeletal disorders (disease or injury). Unlike some of the other UL measures available, it was not meant for use with neurologic patients. Its intent is to consider the arm as a whole unit rather than assessing joint-specific issues.\(^\text{30}\) The DASH uses a self-report format that does not require interviewer guidance. It consists of 30 items in the categories of UL physical function, symptoms, and social/role functioning. There are two optional modules, sports/performing arts and work. All items are scored on a 5-point response scale with higher scores reflecting greater UL dysfunction. The response scale does not penalize respondents for using an assistive device to help perform the activity. A shortened 11-item version of the DASH, known as the Quick-DASH,\(^\text{44}\) also exists and has demonstrated good reliability and validity with adults who have had traumatic hand injuries. The DASH and Quick-DASH are available free of charge in both paper form and online version from the Institute of Work and Health and the American Academy of Orthopaedic Surgery.\(^\text{45}\) Users are required to sign an agreement that they will not modify the measure in any way.

The DASH is internationally recognized and has been extensively validated with various orthopedic patient groups since it was introduced in the mid-1990s.\(^\text{30,46,47}\) One nonprosthetic study with adults with radial deficiency that had been at least partly corrected by surgery\(^\text{41}\) pointed out the lack of relationship between hand function and the DASH. Specifically, DASH scores showed minimal overall functional disability of the UL, whereas the Jebsen Test showed marked restriction of the affected hand's function. This lack of penalty for functional compensation demonstrates the DASH's potential for UL assessment in prosthetics.

There are three published articles that focus on use of the DASH with amputees.\(^\text{5,48,49}\) In Davidson's work,\(^\text{5}\) there is evidence of the DASH's discriminant validity (e.g., differences in functional abilities between able-bodied adults and those with an UL amputation, and differences in functional abilities and self-esteem between amputees grouped according to level of amputation). As well, there was indication of ability to detect change in status, although no details were given in the responsiveness subsample as to the time frame for follow-up or nature of interventions in the follow-up period. These limitations make the magnitude of the change scores difficult to interpret. In research by Lifchez et al.\(^\text{49}\) with individuals with multiple hand digit amputations and digital prostheses, DASH scores were significantly higher when rated for the prosthetic-on condition than when scored for the prosthetic-off condition. The research by Wada et al.\(^\text{48}\) indirectly gave information on the DASH through its use as a functional measure in a construct validity evaluation of a tumor-classification measure (Enneking score). There was an indication of a strong relationship between the DASH and Enneking score. The authors noted that this was consistent with the relationships seen between the DASH and arm/elbow indices in other populations. In all of these articles, the test-retest reliability of the DASH was assumed rather than tested, based on its reported strength in other clinical populations. There is no information on the effect of the reduced item pool on the psychometric properties of the Quick-DASH in amputees or any indication that this shorter version has been used in UL prosthetic evaluations.

**Measures of Overall Functional Abilities/Participation**

Amputee specific

None identified in the adult UL prosthetic literature.

Generic

None identified in the adult UL prosthetic literature.

**Measures of Health-related Quality of Life**

Amputee specific

**Trinity Amputation and Prosthesis Experience Scales.**\(^\text{50}\) The Trinity Amputation and Prosthesis Experience Scales (TAPES) is a self-report, HRQOL questionnaire, which was developed by a prosthetic research team in Ireland. It was designed expressly for use by adults with upper or lower limb amputations. This multidimensional assessment assesses psychosocial processes that are linked with adapting to a prosthesis, to the activity restrictions associated with wearing a prosthesis, and to satisfaction with the prosthesis. There are 9 subscales in this 54-item questionnaire: psychosocial adjustment (general adjustment, social adjustment, and adjustment to limitation), activity restriction, (functional, social, and athletic) and satisfaction (weight of prosthesis, functional, and esthetic...
Published work on the TAPES' validation with upper and lower amputees has been limited to that done by members of its development team. With lower limb amputees, the internal consistency and factor analysis have been well-established with some preliminary work on construct and predictive validity. In the single publication to date with the TAPES and UL amputees, Desmond and MacLachlan (members of the TAPES development team), noted that the study provides preliminary evidence of the factorial structure and internal consistency of what the authors referred to as the TAPES-upper. They indicated the need to develop an item pool for bimanual functional items to enhance the activity-restriction scale. The authors also noted the need to do predictive validity studies to show the ability of the scale to serve as a measure of "adaptation to UL amputation." As with the TAPES for lower limb amputees, there is a need for studies of test-retest reliability and construct validity.

Generic

Nottingham Health Profile. This self-report survey for adults (ages 16 years and above) was designed by Hunt et al. in the 1980s in Great Britain, and it measures perceived health status (mental, social, and functional) in population surveys and HRQOL outcomes in clinical and research contexts. The Nottingham health profile (NHP) consists of two parts that take 10 to 15 minutes to complete either by mail survey or interview and is available in numerous languages. The first part consists of 6 subscales and 38 items and covers the areas of emotional reactions, physical mobility, pain, sleep, social isolation, and energy level. Part 2 consists of seven questions about the effect of health problems on various aspects of life (e.g., work, personal relationships, and holidays). Each item is linked with a yes/no response. Scores for each subscale are calculated on a 0 to 100 scale of worst to best QOL, respectively. Part 1 can be used without Part 254 and appears to be the section most commonly employed. Information about ordering the NHP can be obtained through its developer.

As a generic measure, the NHP has been used and validated across an extensive variety of acute and chronic adult health conditions (as listed by Finch et al. 15). Test-retest reliability was established in a group of adults with major limb amputations, with intraclass correlation coefficients (ICCs) of 0.60 to 0.83 (confidence intervals 0.56 to 0.87) across the various subscales. It was not possible to tease out the reliability for UL amputees versus those with lower limb amputation. Indeed, reliability for the group as a whole may be higher than it is within upper and lower limb subgroups, because there may have been a wider spread in scores across the whole sample. In the discriminant-validity evaluation by Demet et al., there was evidence of higher HRQOL for UL amputees than for lower limb amputees in all categories except the NHP's social isolation subscale. From a content validity perspective, Demet et al. noted that the NHP does not seem to be as well suited to persons with UL amputation as it is for those with lower limb deficiency given its greater emphasis on lower limb functional abilities. No other validity work has been published for the NHP with amputees or has any responsiveness work been reported for this population.

Short Form 36 Health Survey. This internationally-known questionnaire that was designed in the US is an indicator of perceived health status for individuals, age 14 years and above, who have acute or chronic adult health conditions. The Short Form (SF)-36 is scored according to eight subscales that cover the areas of physical functioning, role limitations due to physical health problems, pain, general health perceptions, vitality, social function, role limitations due to emotional problems and mental health, emotional reactions, physical mobility, pain, sleep, social isolation, and energy level. It includes 36 items, each composed of a 5- or 6-level response scale. Physical and mental health component scores can be calculated either manually or via scoring software that permits handling of missing responses. The SF-36 is available in more than 50 languages, can be administered either by an interviewer or self-administered, and requires 5 to 10 minutes to complete. A summary of the SF-36 provides an update of current and future directions for the measure and its derivatives. The Medical Outcomes Trust (MOT), Health Assessment Lab (HAL), and QualityMetric Incorporated, co-copyright holders of all SF-36®, SF-12® and SF-8™ Health Surveys, have merged their licensing and user registration programs. Licensing information can be accessed at the SF-36 website.

SF-36 validation work has been completed with various clinical populations internationally. There do not appear to be SF-36 publications specific to amputees other than longitudinal work with lower limb traumatic amputees, and its use as a validity standard in the article by Wada et al. that is evaluated in this report. Wada et al. assessed the SF-36 through its use as a functional measure in a construct validity evaluation of a tumor classification measure (Enneking score). There was no indication of any relationship between the SF-36 and the Enneking score in the UL amputees. The authors noted that only 3 of 10 of the SF-36's physical function items are directed toward the UL.

SECTION TWO: MEASURES USED WITH CHILDREN WITH UPPER LIMB AMPUTATIONS

Hand Function

Amputee specific: Measures of prosthetic control

Assessment of Capacity For Myoelectric Control. The reader is referred to the description in the adult section earlier. There were no separate analyses or conclusions from Hermansson's or Lindner's work for the pediatric component of the sample and no other published studies of use of the ACMC specifically with pediatric UL amputees.
Unilateral Below Elbow Test. This observational test was developed by Bagley et al.59 for use within a Shriners Children's Hospital UL amputee study that compared the functional abilities and QOL of wearers and nonwearers of prosthetic devices.60 The Unilateral Below Elbow Test (UBET) was designed to fill a gap in measurement of hand function capability of children and youth (ages 2 to 21 years) who have an amputation and do not wear a prosthesis. It consists of nine bimanual tasks that are specific to one of four developmentally-based age groups. If the child is a prosthetic wearer, he performs the tasks in both the prosthesis on and off conditions, whereas non-wearers perform the tasks without a prosthesis. Ratings are done on two subscales: Completion of Tasks (5-point degree of difficulty-based scale) and Method of Use (4-point nominal scale). Testing of the nine tasks takes about 20 minutes.59 Administration guidelines can be obtained from the first author of the validation publication.59

The UBET's use in published UL prosthetic articles has been in the original validation study by Bagley et al.59 and in subsequent studies by James et al.60 and Buffart et al.61,62 There was good to excellent interobserver reliability for Completion of Task scores and fair to excellent interobserver and intraobserver reliability for Method of Use with prosthesis-on/prosthesis-off.59 Measurement-error estimates from test-retest evaluation may be too large to allow detection of small change.61,62 Although the test was deemed to be quick and easy to perform, there was limited support for convergent and construct validity in children with UL deficiency.61 There was better validity performance in children with radial deficiency.62 A ceiling effect was evident in both of the Buffart et al. studies. From the outcomes perspective, responsiveness to change requires additional investigation, because the evaluation of change judgments by Buffart et al.61,62 were made on the basis of measurement-error estimates alone.

University of New Brunswick Test of prosthetic function.63 This assessment was the first of the formal observational tests of function, which was developed specifically for children with a unilateral UL amputation and subjected to preliminary validation work. It was developed by an engineer and an occupational therapist at the Bioengineering Institute of the UNB, Canada, with international occupational therapist collaboration. The UNB Test measures the method of performance and spontaneity of prosthetic performance in children (ages 2 to 13 years). It has four age-based modules (covering 3 year age intervals) in which developmentally appropriate bimanual tasks were chosen. Method and spontaneity are each rated on 5-point scales. In its scoring, it is assumed that performance using the prosthesis actively is superior to using it as a stabilizer, an assumption that is questionable with some of its tasks.64 The UNB Test is well-known in clinical circles internationally. Indeed it was designed for use within the prosthetic clinic. The UNB Test manual is available online as a pdf file,65 and the UNB Test kit can be built by the clinician according to the instructions in the manual. It is scored from live performance and usually takes 20 to 30 minutes to complete.63,66 There is no specific training program for those administering the test.

The UNB Test manual contains some information on the validation work done by its development group, but these results were not published in a peer reviewed journal. This work provided an initial indication of interrater reliability, although use of Pearson correlation coefficients for this work reduces its interpretability. The UNB Test's use in published UL prosthetic articles has been limited to the studies by Wright et al.,66 Burger et al.,67 and Ballance et al.68 In the work by Wright et al.66 and Burger et al.,67 the UNB Test was used as an observational performance standard in construct validity evaluations of self-report functional questionnaires. The associations with reported performance were moderate to strong overall,66,67 and the test also showed reasonable age-based discriminant validity. There was also evidence of moderate to strong relationships between the UNB's subscales (skill and spontaneity) in all three articles.66–68 There is concern, particularly with older children, about the presence of a ceiling effect.67 The test's responsiveness has not been investigated, and the absence of measurement-error estimates from the reliability work means that it is not possible to estimate minimum detectable change.

Generic/other diagnoses

Assisting Hand Assessment.69,70 The Assisting Hand Assessment (AHA) is an observational assessment that was designed in Sweden by Krumlinde et al.66,68,69. It measures the effectiveness with which children with unilateral impairment use the affected hand along with the noninvolved hand in bimanual play activities. The intention is to see how the involved hand functions as an "assisting hand." The purpose of the assessment is to elicit spontaneous and natural performance of grasp, release, and manipulation skills during 12 to 14 fun-play activities. The AHA was designed for children with hemiplegic cerebral palsy (CP) or obstetric brachial plexus palsy (OBPP) from age 18 months to 12 years (school-aged module available now). The AHA is a Rasch-built measure, which has 22 actions that are scored on a 4-point effectiveness of performance-rating scale. It takes about 15 minutes to perform this play-based assessment.71 The scoring is done from a video of the assessment. There is an intensive 2.5-day training program that is required before a therapist uses the AHA. Information on training and the manual and materials kit is available at the developers' website.71

Most of the published work focuses on children with CP or OBPP by members of the AHA's development team69,70,72 in which the interrater and intrarater reliability, and construct validity and sensitivity to change have been demonstrated using a Rasch-measurement model for analysis. Recent use in a clinical trial with children with CP has shown the AHA's potential to detect change in a child's ability postintervention.73 There are two publications that focus on the use of the AHA with UL amputees.61,62 From these two studies, there is strong evidence of the AHA's test-retest reliability and construct validity, and ability to detect change in both clinical groups. Children with unilateral UL amputation scored comparatively low on the AHA because of its focus on quality of
performance rather than difficulty. In contrast, a ceiling effect was noted on the AHA for children with radial deficiency. Based on their study results, the AHA is one of the two observational measures of hand function that Buffart et al.\textsuperscript{61,62} recommended for use in pediatric UL deficiency.

**Measures of Upper Limb Functional Abilities**

**Amputee specific**

**Child Amputee Prosthetics Project-Functional Status Inventory.**\textsuperscript{74} The Child Amputee Prosthetics Project-Functional Status Inventory (CAPP-FSI) is part of a pediatric amputee measurement set developed in California (University of California at Los Angeles). One of the developers (J.V.) was also the creator of the PedsQL.\textsuperscript{75} The advantage of having a suite of measures such as the CAPP, which covers the pediatric age span, is that the child can transition from one measure to the next as they develop. Follow-up scores are on the same metric and, thus, interpretable when looking at change in status during extended periods of time.

The CAPP-FSI was the first to be published as a functional status parent-report questionnaire for pediatric prosthetics. It was designed for use with parents of children ages 8 to 17 years (i.e., it does not have a child-report component) and is made of upper (34 bilateral items) and lower limb (6 items) functional activities. Each item is rated on "does the activity" and "uses the prosthesis" scales, which employ a 5-point "frequency of time" rating scheme. Because the scales were validated separately, it is acceptable to choose either the upper or lower limb sections, depending on the diagnosis of the child. The measure has English and Spanish versions. It is not clear as to how to acquire the CAPP-FSI, how long it takes to do, and whether it is still being used.

Despite considerable time since its development, published reports on the CAPP have been limited to the initial validation work (internal consistency, content validity, and construct validity evaluation) by its developers\textsuperscript{74} and a construct-validity study by Burger et al.\textsuperscript{67} using just the UL component of the CAPP-FSI. It was not used in Buffart et al.'s measurement-comparison studies.\textsuperscript{61,62} This is perhaps because, as Buffart et al.\textsuperscript{16} note in their outcomes review, the CAPP-FSI neither rates how a child does the activity nor does it consider the difficulty that a child has with prosthetic use. Pruitt et al.\textsuperscript{74} concluded that the CAPP-FSI has high internal consistency and is able to discriminate between children with upper and lower limb deficiency. There is also initial evidence of construct validity.\textsuperscript{67} Burger et al.\textsuperscript{67} noted that although the CAPP-FSI is useful or quick evaluation, observation of hand skills on a validated prosthetic measure such as the UNB Test should also be done.

**Child Amputee Prosthetics Project-Functional Status Inventory Preschool.**\textsuperscript{76} This measure is a derivative of the CAPP-FSI, and its only difference is in the number and content of items (31 UL and 6 lower limb items). It is intended for parents of children ages 4 to 7 years. Published reports on the CAPP-FSI preschool (CAPP-FSIP) have been limited to the initial validation work by its developers\textsuperscript{74} and a construct validity study by Burger et al.\textsuperscript{67} with the UL component of the CAPP-FSIP. It was validated by its developers in exactly the same manner as the CAPP-FSI, with similar conclusions. Burger et al.\textsuperscript{67} noted that although the CAPP-FSI is useful for quick evaluation, observation of hand skills on a validated prosthetic measure such as the UNB Test should also be done.

**Child Amputee Prosthetics Project-Prosthesis Satisfaction Inventory Toddler (CAPP-FSIT).**\textsuperscript{77} This measure is a derivative of the CAPP-FSI, and its only difference is in the number and content of items (31 UL and 6 lower limb items). It is intended for parents of children ages 1 to 3 years. Published reports on the CAPP-FSI toddler (CAPP-FSIT) have been limited to the initial validation work by its developers.\textsuperscript{77} It was validated by its developers in exactly the same manner as the CAPP-FSI, with similar conclusions. No other validation work has been published with the Toddler version.

**Prosthetic Upper Extremity Functional Index.**\textsuperscript{64} This functional status questionnaire was developed in Canada by a pediatric prosthetic research team. It was designed specifically for evaluation of children and teens who have a unilateral UL amputation and have a prosthetic device. It evaluates a child's ability to perform bimanual activities with and without prosthesis and also looks at the perceived usefulness of the prosthesis. The PUFI consists of the young-child version (ages 3 to 6 years) with 26 items and an older-child version (ages 7 to 18 years) with 38 items. The young-child version is parent-report format, whereas the older-child version has both parent-report and child-report formats. The PUFI consists of four separate response scales (method of performance 6-point scale; ability to perform with prosthesis 4-point difficulty scale; usefulness of prosthesis 3-point scale; and ability to perform without prosthesis 4-point scale). Although the PUFI originated as a paper-report questionnaire, it was redesigned as a direct access software program to facilitate administration, scoring, and data collection and to make it child-friendly and parent-friendly in format. The PUFI takes 20 to 30 minutes to complete,\textsuperscript{64,66} and can be done by the parent or child after a standardized introduction to it by the clinician. The PUFI is available in English, Dutch, Swedish, Slovenian, and French and can be obtained from its first author free of charge in conjunction with joining the PUFI database network.

The PUFI has been used in the initial validation studies by its developers,\textsuperscript{64,66} in work by Buffart et al.,\textsuperscript{61,62} and by James et al.\textsuperscript{60} The initial validation work by Wright et al.\textsuperscript{64,66} was with the paper version, whereas all subsequent studies have been done using the software version. This makes the reliability work by Buffart et al. particularly important, because an assumption had been made previously by the PUFI developers that the reliability of the paper form and electronic version was comparable. The PUFI showed good test-retest reliability for older-child respondents and fair to good reliability for parent respondents for its various sections, with lowest reliability for the parent-report older-child version.\textsuperscript{64} As hypothesized, there were fair to moderate correlations between the
UNB Test and PUFI\textsuperscript{66} in a group of young and older child prosthesis wearers. Buffart et al.\textsuperscript{61} noted that their work provided additional support for test-retest reliability and construct validity of the PUFI for children with UL reduction deficiency and initial evidence of the measure’s value in children with radial deficiency. They provide the first evidence of the PUFI’s potential for measuring change as well, although large-scale prospective work is required in this area. Based on these study results, the PUFI is one of the two measures of hand function that Buffart et al.\textsuperscript{61} recommend for use in pediatric UL deficiency.

\textbf{Generic/other diagnoses}

\textbf{ABILHAND-Kids}\textsuperscript{78,79} The ABILHAND-Kids functional status questionnaire was developed in France and assesses a child’s ability to perform everyday manual activities. It is a parent-report paper questionnaire that assesses the child’s difficulty in performance as perceived by the child’s parents. In terms of item content, it was designed to be suitable for use with children 6 years of age and older. Its 21 activities are mostly bimanual tasks. Each task is rated on a 3-point degree of difficulty scale. The ABILHAND-Kids was developed using the Rasch-measurement model, so that it is a linear measure with a unidimensional scale. The scale was calibrated in children with CP aged 6 to 15 years.\textsuperscript{79} The authors note that it takes about 10 minutes to complete. The ABILHAND-Kids is available in French, English, and Dutch, and the assessment and scoring program can be accessed free of charge online through the ABILHAND website after registration.\textsuperscript{80}

Published studies are limited to the validation work in pediatric CP\textsuperscript{79} in which test-retest reliability was established and to the two studies by Buffart et al.\textsuperscript{61,62} described in this report. Because the Rasch scaling was done in CP, there is no guarantee that items are arranged in the same hierarchical order with amputees. Thus, Buffart et al.\textsuperscript{61,62} appropriately relied on raw scores rather than logits when they used the ABILHAND-Kids with children with radial deficiency or unilateral UL amputation. Test-retest reliability was excellent in both groups. A ceiling effect was noted for both groups of children. There was evidence of ability to detect change in both clinical groups, and construct-validity comparisons revealed fair to moderate relationships with other UL functional tests.

\textbf{Measures of Overall Functional Abilities/Participation}

\textbf{Amputee specific}

None identified in the pediatric UL prosthetic literature.

\textbf{Generic}

\textbf{Pediatric Orthopedic Data Collection Outcomes Instrument}.\textsuperscript{81,82} The Pediatric Orthopedic Data Collection Outcomes Instrument (PODCI) is a generic questionnaire that was designed for use for school-aged children and youth with musculoskeletal conditions as a measure of ability to participate in normal daily activities and sports activities, as well as a brief measure of pain and overall health. There is a parent-report questionnaire (for use with parents of children up to 11 years of age), a parent-report adolescent version, and an adolescent-report version. The PODCI has four main functional scales (basic mobility and transfers [11 items], sport and physical functioning [12 items], UL and physical function [8 items], and pain/comfort [3 items]) and additional items that assess pain/comfort, treatment expectations, happiness, and satisfaction with health. Only seven of the functional items are bimanual activities. Functional ratings are done on 4-point to 6-point degree of difficulty response scales. It takes about 20 minutes to complete the PODCI.\textsuperscript{83,84} The child and parent questionnaires and scoring templates are available as a PDF file and can be accessed without charge through the American Academy of Orthopaedic Surgeons’ website.\textsuperscript{84}

Validation work (namely discriminant and construct validity evaluations) has been done with children/youth with CP (often in the context of orthopedic surgery), children undergoing scoliosis/spine surgery, and those with fractures and other musculoskeletal conditions.\textsuperscript{81,82} Orthopedic studies that are closely linked to UL amputees are those with children with OBPP\textsuperscript{85,86} and an outcomes review by Pakulis et al.\textsuperscript{87} in which PODCI, pending population-specific validation, was recommended as potentially suitable measure for patient with bone tumors. Most of this work has been done by investigators who are part of the same research organization affiliated with the Shriners Hospitals in the US.

The studies by Lerman et al.\textsuperscript{88} and James et al.\textsuperscript{60} cited in this report are the only ones that pertain to its use with pediatric amputees. In the evaluation of prosthetic and nonprosthetic users by James et al.,\textsuperscript{60} there was no evidence of a relationship between the PODCI and the functional and QOL measures used. The PODCI fared better in the study by Lerman et al.\textsuperscript{88} in which it differentiated between able-bodied children and UL amputees for all of its scales for parent respondents, and for UL and sports scales for adolescent respondents. There was a moderate association between UL function and amputation level for adolescent responses. A ceiling effect was strongly suggested in both studies for all of the PODCI’s subscales. There are no publications on the PODCI’s test-retest reliability or responsiveness with amputees.

\textbf{Measures of Health-related Quality of Life}

\textbf{Amputee specific}

None were identified in the pediatric UL prosthetic literature.

\textbf{Generic}

\textbf{Pediatric QOL Inventory}.\textsuperscript{89} The purpose of this internationally-known questionnaire is to evaluate HRQOL of children. The 26-item
generic version of the Pediatric QOL Inventory (PedsQL) consists of four multidimensional scales (physical, emotional, school, and social function). The generic version has age-appropriate versions and allows child-report for those who are aged 5 years and above. It is available in multiple languages, and there is now an internet version. It has been used as a population health measure as well as for numerous diagnostic groups (e.g., asthma, diabetes, CP, cancer, heart disease, arthritis, attention deficit disorder, obesity, spina bifida, fractures, and brain injury) and in numerous cross-cultural validation projects. There are conditionspecific modules for a number of disorders including asthma, arthritis, cardiac, and diabetes. The questionnaire takes about 5 minutes to complete. The PedsQL suite of measures can be obtained online. There is a licensing fee that is based on the type of use (research or clinical) and the funding in place.

There is an extensive reference list for published studies at the PedsQL website, strong evidence of its reliability and validity, and early evidence of responsiveness, although the latter does not seem to be a key topic of study for the PedsQL. One notable gap in the evaluation of the PedsQL is with children, who have upper or lower limb amputations, either as an isolated group or as part of a study of children with disabilities (i.e., not included in the PedsQL work by Varni et al. of 10 disease clusters). The single published work in prosthetics by James et al. gave a brief look at the adequacy of the PedsQL for children with UL amputations. Scores overall were high (suggesting a ceiling effect), with the PedsQL differentiating between wearers and nonwearers only for school scale of psychosocial health domain (wearers scored higher, p < 0.01). There is no other information to date about the psychometric properties of this measure with pediatric amputees (upper or lower limb).

DISCUSSION

In the adult literature, seven outcome measures (4 amputees specific and 3 generic) were revealed: the SHAP and the ACMC for hand function, the DASH and UEFS (from the OPUS) for UL-focused functional abilities, and the NHP, SF-36, and TAPES for QOL. This compares with 25 measures (amputee specific and generic) in the lower limb amputee outcomes review by Condie et al. The two measures that were common in the adult UL and lower limb reviews were the TAPES and the SF-36.

In the pediatric review, nine outcome measures (5 amputees specific and 4 generic) were evident: the ACMC, UBET, UNB Test, and AHA for hand function, the ABILHAND-Kids, CAPP-FSI, and PUFI for UL-focused functional abilities, the PODCI for participation, and the PedsQL for QOL. Two of the pediatric measures, the CAPP-FSI and PUFI have younger child versions. There was overlap of one measure, the ACMC, between adult and pediatric studies.

In the research and development phase in the adult literature in particular, there has been a clear focus on biomechanical or impairment-based evaluations of innovative prosthetic hands, but with a few exceptions, there is little accompanying evidence of the validation status of the hand function tools used. For the most part, the prehension or timed tests used were constructed for the purpose of the study. Although this assessment emphasis on hand skills makes sense given the complexity and uniqueness of many of the devices being evaluated and the need to identify which features work and which ones need refinement, it misses the important link to higher level UL function. Lack of use of validated instruments also limits the interpretability of the results.

It became increasingly clear during this review that the typical focus of functional outcome evaluation in adult prosthetics is on the use of satisfaction or functional questionnaires/surveys that were designed for the purpose of the study, subjective reports, and/or wear time estimates as the signal of success, rather than the application of validated functional or QOL measures. This is shown clearly in the list of excluded articles in Table 2. Although in pediatrics, this notion that success relates to wear time was also evident, development and validation of functional status measures have clearly been a major focus of the pediatric prosthetic research teams since the mid-1990s.

One of the key questions pondered was why there are so few adult prosthetic-specific measures in comparison with the number of pediatric ones that have been designed or adapted for prosthetics. This paucity perhaps fits with the identified issue of lack of use of functional outcome measures in adult hand rehabilitation in general, a problem that seems to be based both on the absence of “good” measures that fit the goals of rehabilitation and on lack of clinician time to carry out detailed hand assessments even when good measures are available. There were also comments by Heinemann et al. in their discussion of development of the OPUS functional status measure that they focused on the lower limb for this adult measure because of the limited number of individuals who have UL amputations. In contrast, it is clear that there is a strong international pediatric rehabilitation community (e.g., through groups such as Association of Children's Prosthetic-Orthotic Clinics [ACPOC] and International Society of Prosthetics and Orthotics [ISPO]) that provides support and encouragement for the development and validation of outcome measures. There have been many challenges by funders in pediatrics to find evidence for the early fitting of prosthetic devices to young children and a number of the prosthetic-specific functional measures have directly arisen out of this need.

For the adult literature especially, potentially erroneous assumptions seem to have been made that the outcome measures that are imported from other areas of rehabilitation maintain their psychometric properties when transferred to use with UL amputees. The underlying belief was less evident in pediatrics, such that the majority of measures used have been designed and validated for with prosthetic groups. That being said, the use of generic functional measures such as the PODCI and PedsQL has not been accompanied with disclaimers regarding the unknown state of validation with pediatric UL amputees.

As Condie et al. noted with the lower limb literature, the measurement articles tended to be dense in content and detailed in
statistical concepts (especially the Rasch analysis articles) and did not have much information on clinical applicability. This means that they may not influence clinical practice as much as they should. There were a number of articles, particularly in the adult literature, in which validation results were buried within the article, which makes them difficult to locate.

SUMMARY OF THE PROPERTIES OF OUTCOMES MEASURES REVIEWED

There is clearly no single accepted UL measure in either pediatric or adult prosthetic work. In the pediatric literature, there is a sense of the value in using both an observational hand function scale and a self/parent-report questionnaire of function to try to get a broad picture of abilities and to help to solve any issues that the child might have in actual operation of the prosthetic limb. Readers are referred to the outcome measures rating table (Table 8 and Table 9) and the final column in Table 11 for the background for the quality evaluation and points made below.

Measures of Hand Function

Adults

Three measures that were evaluated in the articles reviewed fit into this category. The ACMC shows great promise from the psychometric standpoint, but is limited to use at present with individuals who wear a myoelectric prosthesis. Furthermore, given the complexity of the assessment, it requires that the assessor have strong experience in working in myoelectric training and, also, have taken and passed the ACMC course and criterion test. These requirements may mean that its clinical uptake internationally is limited.

The SHAP, although designed for use with prosthetic users, has received little testing as of yet with UL amputees, and until it does, it is not a candidate for a strong recommendation for an outcomes core set. Its assessment focus however does make it a highly relevant measure, particularly for use in the research and development phase. The UNB Test, a similar type of measure, has not been formally adapted for use with adults, but as shown in the work by Lake, it has potential as a clinically-applicable test that may be worth adapting and validating, given its clear focus on prosthetic use regardless of type of prosthesis.

Pediatrics

The prosthetic-specific measures that were evaluated are the ACMC, the UNB Test, and the UBEI, whereas the nonprosthetic measure was the AHA. As noted in the adult section above, the ACMC shows great promise from the psychometric standpoint, but is limited to use with individuals who wear a myoelectric prosthesis. Its training requirements may mean that its clinical application is limited. Responsiveness to change is unknown.

The UNB Test suffers mostly from lack of validation efforts. It has achieved quite high clinical acceptance suggesting its clinical utility, but was developed in the 1980s before extensive psychometric evaluation was considered a priority. Rather than addressing this needed reliability work for the UNB Test, other researchers seem to believe in its psychometric strength and use it as a validation standard. The results of its use in this way in the studies reviewed in this article have revealed that its construct validity is supported. The unknown factors relate to a ceiling effect that has been suggested and a related issue in terms of its ability to evaluate change.

From a content perspective, the UBEI fills a gap in measurement of the functional capability of children and youth, who have an amputation and do not wear a prosthesis. However, the UBEI demonstrated weakness in terms of its construct validity. In addition, validation-study methodology issues limit the strength of its conclusions on its reliability. The test does not appear to be widely available. Responsiveness to change is yet unknown.

The AHA was used for the first time with amputees and showed promise as far as test-retest reliability and minimum detectable change. The AHA correlated moderately with the PUFI, which means that they provide complementary and additional information about the child’s hand function. Children scored low on the AHA compared to the PUFI, perhaps because of its focus on quality of performance rather than on degree of difficulty to perform. Buffart et al. recommended the AHA for use with pediatric amputees alongside the PUFI. It will be important to tease out whether there is additional value to using the AHA along with a prosthetic-specific observational hand function measure such as the UNB Test.

Measures of Upper Limb Functional Abilities

Adults

The UEFS module from the OPUS was the single prosthetic-specific measure that was available in this category. Although on the surface, the revision by Burger et al. appears to have considerable promise given the positive results of the Rasch analysis work, there are two key limitations. The first is the inclusion of multiple one-handed activities in the task list. Indeed, fewer than half of the 19 activities in the revised version were classified as bimanual activities. These unilateral tasks will not give the clinician information on the way that the affected hand is used in daily activities, and hence it is unlikely that this measure will be as capable of detecting change related to prosthetic device modifications or improved skill as a measure that is comprised of purely bilateral tasks. Second, Burger et al. note that the UEFS is relatively basic in the daily living skills that it evaluates, and they suggest the need to include more difficult skills such as those found in measures like the ABILHAND, PUFI, or UNB. The issue of a ceiling effect is a drawback when one thinks of evaluating outcome in younger individuals, in particular, following traumatic hand injuries, because they likely have high functional demands/expectations.

The DASH is the sole, nonprosthetic functional measure that was studied with amputees. In initial validation work with amputees, it
showed evidence of construct validity, and as a result, it was recommended by Davidson as a suitable measure for amputees.\(^5\) It is important to note that the DASH has two optional modules (sports and performing arts) that might help to identify issues with more active individuals. It is one of the most extensively validated and accepted UL functional status questionnaires internationally across a diversity of hand/arm dysfunction clinical populations. Its responsiveness to change is not yet tested with UL amputees.

**Pediatrics**

The amputee-specific measures that were studied are the CAPP-FSI and the PUFI, and the nonprosthetic measure in this area is the ABILHAND-Kids. Validation of the CAPP-FSI series of measures has been limited to work done by its developers. Other than one study by Burger et al.\(^57\) in 2004, there has been no published work on the CAPP scales since the late 1990s. Although the early results were promising in terms of formal establishment of validity with sensible validity constructs, the work on reliability has been limited, and there is no information of responsiveness.

Work with the PUFI has been more widespread than the CAPP-FSI, with the most recent publications in 2007 by Buffart et al.\(^61\) The advantage of this later work is that more recent reliability concepts have been applied to the measure, both confirming the construct validity results presented by the PUFI’s developers and adding to reliability information by providing estimates of measurement error and detectable change. Construct validity has also been evaluated in several studies with similar results across the various investigations. The greatest criticism of the PUFI is the number of items/response columns that the respondent (parent or child) has to complete, reducing its clinical acceptability. Its responsiveness to change is as yet untested with amputees.

The ABILHAND-Kids was used for the first time with amputees by Buffart et al.\(^61,62\) and showed strong promise as far as test-retest reliability and minimum detectable change. The question remains as to whether or not this parent-report questionnaire provides any additional information/outcome measurement advantage over one of the prosthetic-specific functional questionnaires such as the PUFI.

**Measures of Participation**

**Adults**

There were no measures studied in this category, although the DASH has several items that address participation.

**Pediatrics**

The one measure that was reviewed was the PODCI, a generic parent-report questionnaire. The information that came from its use in two related studies cannot be considered sufficient to support or refute its use with child amputees.\(^61,62\) Given the strength of the PODCI with other pediatric orthopedic groups, it might be a good measure to evaluate in terms of the range of scores that is obtained with pediatric amputees. If the score range indicates that the measure is suitable in terms of content (i.e., does not demonstrate a strong ceiling effect), test-retest reliability could then be assessed with pediatric amputees.

**Measures of Quality of Life**

**Adults**

Two generic scales (SF-36\(^57\) and NHP\(^53\)) were evaluated, and one amputee-specific scale (the TAPES) was reviewed. Both of the generic scales are well-known and have been tested in adult populations and found to be reliable and valid. The evaluations done in the amputee articles can only be considered preliminary, but demonstrate that the NHP has potential as a measure of QOL with amputees. However, the SF-36 seems to suffer from a ceiling effect in this population.

The TAPES\(^50\) is a soundly-built measure that has undergone rigorous evaluation by its developers with both upper and lower limb amputees to establish its internal consistency and factor structure. Work is needed to determine its test-retest reliability and responsiveness to change in UL amputees. Concerns were expressed by authors that participants in their factor-analysis study were mostly older veterans with long-standing amputations and as such might not be a representative of individuals with a recent amputation or of younger individuals. This could be a limitation in respect to its content validity.

Although there is a module of the OPUS\(^42\) that pertains to HRQOL, the lack of information on the number of subjects with UL loss (or any breakdown of results for the UL subgroup) in Heinemann et al.’s\(^42\) work on its validation, and the lack of evaluation of this module in Burger et al.’s study\(^49\) precludes comment on its measurement adequacy. Specific validation work with UL amputees is indicated for the QOL module before it can be recommended for use.

**Pediatrics**

The PedsQL\(^89\) is the only QOL measure that was used in the pediatric articles, and the information that came from its use in a single study cannot be considered sufficient to support or refute its use with child amputees. Given the strength of this test in multiple other branches of pediatrics, it might be a good measure to evaluate both in terms of the range of scores that is obtained with pediatric amputees (i.e., is there a ceiling effect) and in its test-retest reliability if the score range indicates that the measure is suitable in terms of content.

**CONSIDERATION OF NONPROSTHETIC MEASURES OF HAND FUNCTION**

Given the serious outcome measure gaps noted from review of many of the published studies (particularly with adults), it is important...
to think outside the box when deriving a list of recommended measures. Are there nonprosthetic measures that might be appropriate for use in prosthetics, or does time and funding need to go into the development and testing of new prosthetic specific measures? If we include any of the generic measures, they must be chosen carefully to ensure suitability in terms of content and focus. As noted by Light et al. \(^{105}\) in their development of the SHAP, "The assessment of prosthesis users warrants special criteria. Prosthesis users require coordinated movement of the UL and therefore do not exhibit the separable functions of hand shaping and arm movement seen in natural upper-limb subjects . . . . Unilateral prosthesis wearers rarely use the device for reaching and grasping of objects, and it mainly fulfills a stabilizing role for the natural hand in bimanual activities." However, Buffart et al. \(^{116}\) noted that generic hand tests and functional arm assessments might be suitable for use if they concentrate on bimanual activities and contain an assessment of quality of movement (e.g., ratings of degree of difficulty and speed or accuracy of performance).

**Adults**

For a comprehensive list of published UL tests that are available for use with adults, the reader is referred to the article by Metcalf et al. \(^{14}\) that summarizes these assessments according to the ICF framework. The ones that fall within the ICF's body structure and functions category and cover underlying components of hand function (e.g., prehensile grasp and dexterity) may hold considerable promise for evaluation of prosthetic devices in the research and development phase. Although an argument can be made for considering hand and UL measures from the orthopedic realm (e.g., injury and arthritis), the same does not hold true for measures from neurology (e.g., stroke). Clients poststroke typically have cognitive issues as well, so are potentially more compromised in function, and because they are typically seniors, the functional expectations postrehabilitation are not highly demanding (e.g., Arm Motor Ability Test \(^{115}\)). These measures also tend to focus on unilateral tasks allowing performance solely by the unaffected side in the case of UL amputees (e.g., Wolf Motor Function Test \(^{114}\)). In the orthopedic literature, there are numerous well-known, validated tests of dexterity that are available (e.g., Jebsen Standardized Test of Hand Function \(^{37}\)) as listed in Metcalf et al. \(^{14}\). However, many of these do not look at broad-based hand skills and, like the stroke scales, are typically unilateral in nature (e.g., the Sequential Occupational Dexterity Assessment [SODA] test for adults with arthritis \(^{115}\)).

A systematic review of hand function outcome measures \(^{116}\) as they apply to individuals who have had a hand injury (not amputees) was published in *Archives of Physical Medicine and Rehabilitation* early in 2009, and the author of this current article became aware of it several days after the State of Science Conference meeting during a final search of the hand outcomes literature. This rigorous review by van de Ven-Stevens et al. \(^{116}\) is important for two reasons. First, it stresses the lack of methodologically strong evidence pertaining to the validation status of the majority of generic measures of hand function when applied to evaluation of outcome following hand injury. The lack of knowledge related to responsiveness to change of these generic measures is particularly relevant. Second, it provides a direct point of comparison for the validation of measures in hand injury and prosthetics, because the review process that was developed by Terwee et al. and used in our review was also used by van de Ven-Stevens et al. \(^{116}\).

Measure from the hand injury outcomes review \(^{116}\) that showed strong psychometric properties (i.e., the DASH, Purdue Pegboard Test, Box and Block Test, Jebsen Test of Hand Function, and Canadian Occupational Performance Measure [COPM]) may be good candidates for consideration in UL prosthetics.

When thinking about borrowing tests from clinical areas outside of prosthetics, clinicians and researchers need to ensure that the focus, scaling, and difficulty level of the tasks, are suitable for prosthetic device users. It is essential to also acknowledge that critical psychometric properties of the test such as measurement error may be different when used with amputees than those established with other clinical populations, \(^{116}\) meaning that validation work needs to be conducted before confident use as an outcome measure.

**Pediatrics**

When looking beyond prosthetics to find measures of hand function, there is a link with measures of bilateral hand use in the context of CP. Clearly, motor control aspects of hand and arm use in children with CP are different from prosthetics, and so CP-based measures such as the Quality of Upper Extremity Skills Test \(^{117}\) that look specifically at quality of unilateral hand function (quality of performance in terms of finger action and dexterity of the natural hand) are not appropriate. Also, inappropriate are those measures that assess development as a whole and contain only a few hand function items (e.g., Peabody Developmental Motor Scales or Movement Assessment Battery for Children). \(^{16}\) However, the recently-developed AHA observational assessment (which was designed for individuals who have a UL disability, but used mostly in CP) is suitable, because it measures the way that the involved hand functions as an assisting hand in bilateral activities. Its application to pediatric prosthetics was described in this report.

**GENERIC MEASURES OF UPPER LIMB–FOCUSED FUNCTIONAL ABILITIES**

**Adults**

There are strong arguments that support the use of a total-arm questionnaire when looking at UL disability, stressing that the UL acts as a single functional unit and, hence, should be assessed as such rather than by site-specific measures such as hand or shoulder measures. \(^{118}\) There are several well-known generic measures that focus on hand or UL function in adults that are in the self-report questionnaire category. One that was initially considered is the Michigan Hand Outcomes Questionnaire. \(^{119}\) However, it is not ideal as it contains a majority of questions that look at unilateral function and hand symptoms. One relevant measure is the DASH, \(^{7}\) and indeed it has been used in studies with UL amputees (as reviewed in this article and also discussed in the participation section below). \(^{5,48,120}\)
The other potentially useful generic-arm function measure is the ABILHAND, a self-report questionnaire of manual ability for adults with UL impairments (unilateral as in stroke or unilateral/bilateral as in arthritis). It measures a person's ability to manage daily activities that require the use of the ULs regardless of approach taken. Its 22 items are rated on a 3-point degree of difficulty scale (i.e., easy, difficult, and impossible). It can be accessed through the measure's website. Publications on use of the adult version of the ABILHAND focus on use in the areas of stroke or arthritis. Rasch analysis was done separately for the arthritis group, recognizing the potential differences in the scaling in comparison with the stroke group. It has yet to be applied with adults who have UL amputation, but has been successfully used in pediatric prosthetics (as described in this report).

**Pediatrics**

In pediatrics, there is a lack of arm-function questionnaires such as those available for adults. One exception is the ABILHAND-Kids questionnaire, which was adapted from the adult ABILHAND to measure the ease with which children with CP used their hands in everyday bimanual activities. It transfers well to children with UL amputations as the item content focuses on performance of high-level manual activities that any child would do. It has been used successfully in pediatric prosthetics as reviewed earlier in this report.

**GENERIC MEASURES OF OVERALL FUNCTIONAL ABILITIES**

**Adults**

As noted in lower limb prosthetic outcomes review, there are several well-known functional questionnaires that could be considered for use with adult amputee. These measures included the Functional Independence Measure® (FIM Instrument), the Barthel Index, as well as the Frenchay Activities Index, that was developed for use with adults post stroke. In the application of these measures with adults with lower limb deficiency, the review by Condie et al. described issues with each, e.g., ceiling effects, length, and content. Although the FIM Instrument initially seemed to have potential with UL amputees, there was only one UL study in which it was used. This study was ultimately excluded because only three of the patients were UL amputees. Other than the self-care scale, the majority of items in the FIM's minimum data set pertain to the lower body or cognition. Hence, from a content perspective, the FIM does not seem well-suited to use with UL amputees. In addition, the FIM's response options are such that an individual who uses any type of device cannot receive full marks in the independence ratings even if they are fully independent in performing the skills in question.

One generic, Rasch-scaled observational test that allows an evaluation ease of performance of personal and instrumental ADL is the Assessment of Motor and Process Skills (AMPS). The task list can be UL-focused if desired, and items can be added to suit the client population. The client and clinician choose a relevant task on which the client would like to be assessed (similar to the ACMC). This could be an activity such as cooking tasks, sweeping outside, and vacuuming the car. Although the AMPS does not seem to have been used with amputees, it is worth investigating. It has been validated in many other clinical adult groups with physical impairments. Its greatest drawbacks are the formal training required of the assessor and time requirements of the assessment.

**Pediatrics**

When looking at the overall functional status of the child and, thus, including activities that involve the UL (unilateral or bilateral), one can then branch to well-known parent-report or child-report activity questionnaires such as the Pediatric Evaluation of Disability Inventory, the Activities Scale for Kids (ASK), the WeeFIM Instrument, and the PODCI (activity and participation combined). The only one of these diagnosis-free measures that has had any formal study in pediatric amputees (upper or lower limb) is the PODCI (reviewed in this report). If one wants to know about how successful a child is at integrating prosthetic skill, hand use, and alternate methods of manual performance into a wide range of everyday tasks, it may be valuable to add on a generic functional measure or a participation measure (see below) rather than relying on prosthetic-specific UL functional measures alone to give the broad picture. However, without a direct comparison with condition-specific measures, it is not possible to know whether these measures provide any additional outcome information about a child's function over and above what we learn from use of a prosthetic-specific parent-report or self-report questionnaire such as the PUF or CAPP-FSI.

As described above in the adults section, the AMPS may hold potential for evaluation of ease of performance in personal and instrumental ADL. The AMPS has also been validated in pediatrics, but not with amputees.

**GENERIC MEASURES OF PARTICIPATION**

**Adults**

The Community Integration Questionnaire for adults, one of the best-known rehabilitation measures of participation, is a possibility for use. Thus far, it has been studied only in neurological populations. The Life Habits (Life-H) scale (the predecessor to the Life-H for children) might be suitable for use with adults, although validation research with the Life-H has been limited to use with adults with spinal cord injury or with neuromuscular disorders.

In a discussion article on participation and activity in adults (inpatient and community), Jette et al. noted blurring in the distinction between participation and activity and suggested that the two concepts are not distinct, but rather different aspects of a single construct.
between ICF activity and participation concepts and suggested that it may be important to rethink the ICF definitions of activity and participation. They referenced work by Nordenfelt, who noted that it might be better to think about actions and then define these with respect to simplicity and complexity. This may mean that in the interest of measurement efficiency, it is not necessary to go beyond existing broad-based measures such as the DASH, because it would be possible then just consider the activities evaluated with respect to their complexity.

**Pediatrics**

Pediatric clinicians and researchers are also conducting similar debates on the definition and measurement of activity and participation. Could participation be better defined by considering the extent of involvement in life situations that matter most to the individual? Composite activities (some of which would be with the prosthesis) might then be among the building blocks for this involvement. Measurement can consider components that are either subjective (considerations of satisfaction with involvement and the sense of belonging) or objective (what the individual is involved in doing out a list of activities that are considered to be a part of daily life). Because participation is clearly not something that is impairment/condition based, it is probably advisable to follow a more generic pediatric path and not create a prosthetic-specific participation tool.

If choosing an objective approach to participation measurement, one could go with a measure like the Life-Habits scale, which is broadly community-based. With this measure, it is important to be aware that inappropriate assumptions will be made in the assistance subscale with this platform when equating the use of devices as an inferior means of participation. It is also essential to be careful of the purpose of measuring participation: is it being done as a descriptive approach (e.g., to know a child’s status for purposes of planning interventions and physical/environmental supports [e.g., discriminative questionnaires such as the Child Assessment of Participation and Enjoyment (CAPE) might work well for this]), or is it to look at change over time? Currently, there is not a recommended way to look at the subjective aspect of participation. Perhaps the subjective aspect of participation is captured better within life satisfaction/ QOL scales.

As for adults with UL amputations, it will be essential to complete validation work with any pediatric generic measure that is chosen. It also will be important to find out with children whether there is a ceiling effect with these measures, because there is literature that suggests that children with congenital amputations tend to compensate well for their physical limitations.

**GENERIC MEASURES OF HEALTH-RELATED QUALITY OF LIFE**

**Adults**

In the adult UL prosthetic literature, QOL has been measured using two generic scales (SF-36 and NHP). Two other QOL measures with adults that might be considered are the Sickness Impact Profile (SIP) and the European QOL Scale (known as the EQ-5D), which were discussed in the adult lower limb amputee outcomes review. The greater use of QOL measures with adults than in pediatrics is perhaps because many of the adults have gone through trauma associated with UL amputation. Consequently, a negative change in QOL is anticipated and goals of intervention are often directed toward QOL improvements.

**Pediatrics**

Methods for QOL assessment were less evident in the pediatric prosthetic literature. Evaluation thus far has been limited to use of the PedsQL. After a secondary search for other measures that are typically used pediatrics (such as the Child Health Questionnaire, KIDS-SCREEN, and the DISABKIDS, no articles using such measures with child amputees (upper or lower limb) were found. The lack of use of QOL measures in pediatrics is puzzling, because it is not consistent with other diagnostic groups. Is it reasonable to think that the children with UL amputation, most of who had congenital amputation and are well-adapted physically, have very little in the way of QOL issues? It is also not clear whether an amputee-specific QOL measure is required, or whether a generic tool such as the PedsQL will suffice.

**MEASUREMENT OF INDIVIDUALIZED GOALS**

The measurement of individualized goals has become increasingly popular over the last 5 years, because it fits well with current client-centered approaches. Two of the internationally-known approaches to measurement of goal accomplishment, Goal Attainment Scaling (GAS) and the COPM, have shown strong psychometric properties when used with rehabilitation clients (adult and pediatric). Although GAS has been studied in lower prosthetics, neither it nor the COPM was mentioned in the UL prosthetic articles. This is curious, because individualized goals are inherently a part of the prosthetic interventions provided, particularly given the efforts that are made to tailor a device to the individual's needs.

Excellent examples of individualized goals come from Walker et al., in which they based the success of recreational terminal device use on whether or not the child still used it at the follow-up period. However, there were no indications in the article as to how well the devices actually worked. Examples of activities chosen are weight lifting (n = 5 youth), Hi Fly fielder (n = 5), Violin bow adapter (n = 2), golf grip (n = 1), slap shot hockey TD (n = 1), and trumpet slide (n = 1). It might have been possible for clients to rate their performance and satisfaction on the COPM as follows: preacquisition, after 2 to 3 months, and then again perhaps after 6 months.

**METHODOLOGY LIMITATIONS**

Many of the same issues that Condie et al., noted in their lower limb outcome study also appeared in the UL literature.
RECOMMENDATIONS FOR A CORE SET OF MEASURES

In forming a core set of outcome measures for a particular clinical group, it is essential to first consider which areas of the ICF are of priority, and then see what the various available measures will provide. What are the outcome questions that are being asked? Should hand function (i.e., ICF impairment and activity) specifically (as in the SHAP) be evaluated and/or is it important to assess the UL as a whole as the DASH does (i.e., ICF activity and participation)? Is it important to differentiate and consider outcomes that relate to hand/arm as well as those that are more encompassing such as participation and QOL, i.e., how much of the ICF framework do we want to try to cover? How does measurement of individualized goals fit in? How do all of the impairment measures that clinicians use with a myriad of different assessment tools link with the core set? How can clinician-respondent and client-respondent burden be minimized with respect to outcome measurement when using a set of measures?

One important thing to consider is that changes in one area of the ICF are not necessarily correlated with or predictive of changes in another. For example, if one wants to know about the extent to which participation outcomes have changed, one should not assume that large changes in underlying skills will lead to large changes in an individual's participation. If both areas are important to the outcomes questions that are asked, it is important to measure each one.

As discussed in a previous publication, one approach to development of a core set might be to use two or more prosthetic-specific hand and UL functional measures along with a participation questionnaire and a measure of QOL. Selection of a measure should not be done simply because it is the only one available; it must actually have sufficiently strong psychometric properties and adequate content fit to support its use. For example, if one thinks of a core set for evaluation of adults, the ACMC and the TAPES are the two measures that have been designed for use with adult amputees and have undergone some degree of validation. For myoelectric users, the ACMC should serve the purpose of helping us to understand skill with the prosthesis (i.e., hand function). However, it does not yet have applicability to individuals who use a body-powered prosthesis. If one wants to know about their hand skills, it might be necessary to go to the SHAP, which requires further validation in prosthetics, or consider a redevelopment and validation of the UNB Test.

If one wants an overall picture of arm function, consideration could be given to the UEFS module of OPUS. One key question though is how to measure the high level demands of adult amputees engaged in competitive sports or similar activities. Are there any scales that look at the functional needs of these individuals both in terms of work activities that they want/need to be able to perform and the sports and hobbies that they want to resume? Is it worth constructing and validating a high-level function arm function questionnaire using templates of existing measure such as the ABILHAND or PUFI as structural starting points and adapting the items accordingly? The generic DASH might be applicable, because it has demonstrated content relevance for high-level activities and some psychometric promise with amputees. Indeed, the development of the sports and hobbies module of the DASH is a good example of this extended approach to measurement to provide a better fit of a tool with the individual client. Perhaps an individualized measure, such as the COPM or GAS, could be used to tap into individual functional priority areas that are the explicit goals of prosthetic prescription and rehabilitation programs. For QOL evaluation, the TAPES could be used in the adult core set to cover the prosthetic experience part of QOL. Consideration would need to be given as to whether a more broadly based QOL measure such as the NHP might also be of added value.

When building a core outcome set for children, there are more prosthetic-specific outcome measures to choose from and, also, more potentially suitable generic outcome measures available in the areas of overall functional abilities and participation. To learn about hand skills, the well-established UNB Test seems to have the most promise for prosthetic wearers overall, although content updating, extension of the test items through the teenage years, test-retest reliability, and responsiveness to change work are still needed.
ensure the test's value as an outcome measure. With myoelectric users, as with the adult core set, use of the ACMC could be considered. Further thought needs to be given to the AHA to decide whether, as a nonprosthetic measure, it gives enough outcome information to the clinical team. According to the extent of validation work done, the PUF1 seems to have the most psychometric support for use as a measure of handfocused and arm-focused functional skills, and allows a comparative look at prosthetic/nonprosthetic performance. Its chief limitations are its length and the lack of information at present on its ability to detect change. As with the adult core set, an individualized measure, such as the COPM or GAS, could be used to tap into individual functional priority areas that are the explicit goals of prosthetic prescription and rehabilitation programs. There is too little information at present on measurement of participation or QOL to make any comment on the measures that should be used. This is an area that needs careful attention in the future.

It is important to recognize that although it is ideal to maintain as much consistency in the core set of measures as possible to support the longitudinal follow-up over clients over a number of years, the introduction of newly-validated (emerging) measures that either fill a measurement gap or that have stronger psychometric properties than existing measures is also an important consideration. A diagram of the evolution of a core set of measures over a 5-year period is provided in Figure 1 to give an idea of the flexibility that should be built into the process.

**CONCLUSIONS**

The use of standardized outcome measures with adult UL amputees is sparse in the published studies of this clinical population, and validation work with the measures that have been used is in its early stages across all components of the ICF. The measures with greatest psychometric promise for use in UL prosthetics are the ACMC, UEFS module of the OPUS, DASH, and TAPES. Consideration needs to be given to adaptation and validation of existing instruments from other areas of hand/UL evaluation or to development of new prostheticspecific measures in areas in which there are measurement gaps. Given the extensive work, time, and skill involved in the development and validation of new measures, this should only be undertaken by measurement experts when there is clear consensus that current measures are not capable of filling the gap.

Greater strides toward measure development and validation have already been made with pediatric UL amputees, and the emphasis that is currently needed is on the determination of the test-retest reliability and responsiveness of the most promising measures (ACMC, UNB Test, AHA, PUF1, and ABILHAND-Kids), as well as discussion on how best to approach the measurement of participation and QOL.

It is advisable to involve the international prosthetic community in the formation of these core outcome sets, perhaps through use of an informal or formal (Delphi consensus voting) approach with groups of prosthetic experts judging various outcome scenarios and possible measurement approaches. A first stage of this process might be to have clients inform the group about the outcome areas that are of greatest priority and also give feedback on the available measures in terms of ease of use, content, etc. Finally, when thinking about the membership of an expert panel, especially for input on the participation and QOL components of the core set, it is very important to involve individuals from all branches of the rehabilitation team, i.e., go beyond occupational therapists and prosthetists and researchers and also include experts from areas such as social work, psychology, nursing, creative arts as well as policy makers.

One of the most important considerations in devising the core outcome sets is the continuing work on the design and building of more functional and innovative prosthetic hands. This is very much in evidence when one refers to the articles by Lake and Dodson and
Kyberd et al. Rapid advancements are being made by researchers (i.e., the Defense Advanced Research Projects Agency initiatives) and prosthetic manufacturers. These new technologies will likely change the focus of how we evaluate specific hand movements and may give more possibility to what individuals will be able to do in their lives. Regardless of the nature of the technological developments, there will still be a strong argument for measuring high-level functional skills with our UL measures/participation tools as well as use broad-based QOL/life satisfaction measures.

CURRENT WORK WITH EXISTING MEASURES

As a final step, it was important to find out what is happening with prosthetic-specific measures as far as current research and future development plans. Validation work continues on the TAPES by its developers in Ireland including a study on optimizing prescription of UL prosthetics (written communication, P Gallagher, Dublin Psychoprosthetics Group, Department of Psychology, Trinity College, Dublin, Ireland, June 20, 2008).

According to the UBET’s developers (written communication, A Bagley, May 11, 2008, and L Wagner May 16, 2008, Shriners Hospital for Children, Sacramento), the UBET is being used clinically at the Shriners Hospitals and was used in research by a prosthetic group from the Netherlands as presented at MEC conference 2008. Plans are underway for review and validation work with UNB Test (oral communication, W. Hill and P. Kyberd, University of New Brunswick Myoelectric Program, Fredericton, Canada March 29, 2009), and it continues to be used clinically at various pediatric facilities including Bloorview Kids Rehab and the University of New Brunswick myoelectric program. Although the CAPPFSI was used by the Shriners Hospitals for Children in their unilateral congenital below elbow deficiency study, it is not used clinically within the Shriners Hospitals at this time or is there indication that other facilities are using it (written communication, J Shida, Child Amputee Prosthetics Project, Shriners Hospital for Children, Los Angeles, April 28, 2008).

Development work also continues with the PUFi by its development team (the leader of this team is the author of this article). An international database exists, managed by the PUFi team at Bloorview Research Institute, Toronto, Canada. Work is nearing completion on a nonprosthetic user module (the UFI) for the software, and progress is being made on the development on an adult module as well as a high-functioning module for young adults/military personnel who have unilateral UL amputations. There has also been further validity work using the PUFi by Dutch prosthetic researchers as reported at the MEC 2008 conference in Fredericton, New Brunswick, in August 2008.

The current ACMC work is led by its developer, Liselotte Hermansson, who indicated that Slovenian and Canadian-French versions are under development. Several ACMC training courses were held in 2008, including one at the MEC 2008 conference in Fredericton, NB in August 2008. The Swedish group is conducting validation work on a Swedish version of the OPUS for use with UL amputees and incorporating the changes suggested by Burger for its UEFs module. The Swedish research group is also working on development and validation of the AHA prosthetic version for children and adults (oral communication, L. Hermansson, Limb Deficiency and Arm Prosthesis centre, Örebro University Hospital, Örebro, Sweden, March 29, 2009).

ACKNOWLEDGMENTS

The author thank Brian Hafner, PhD (Prosthetics Research Study, Seattle, WA) and Sharon Hubbard, MS (Prosthetics Research Study, Seattle, WA) for the insights that she provided into pediatric prosthetic practice. Brian Hafner, PhD (Prosthetics Research Study, Seattle, WA) and Sharon Hubbard, MS (Prosthetics Research Study, Seattle, WA) for their valuable thoughts into adult prosthetic practice and measurement gaps; Angela McDonald, BSc, who on development and validation of the AHA prosthetic version for children and adults (oral communication, L. Hermansson, Limb Deficiency and Arm Prosthesis centre, Örebro University Hospital, Örebro, Sweden, March 29, 2009).

Correspondence to: Virginia Wright, PhD, Bloorview Research Institute, Toronto, Canada; e-mail:

VIRGINIA WRIGHT, BSC (PT), MSC, PHD, is affiliated with the Bloorview Research Institute and Department of Physical Therapy, University of Toronto, Toronto, Canada.

References:


*Source: Journal of Prosthetics and Orthotics 2009; Vol 21, Num 4S, p P3*