A review of clinical upper limb assessments within the framework of the WHO ICF

Cheryl Metcalf¹ BA (Hons) MSc, Jo Adams² DipCOT MBAOT MSc PhD, Jane Burridge², Victoria Yule² BSc MMedSci PhD, and Paul Chappell¹ BSc PhD CEng MZET MIPEM MIEEE

¹School of Electronics & Computer Sciences, University of Southampton, UK, ²School of Health Professions & Rehabilitation Sciences, University of Southampton, UK

Abstract

This paper is intended to provide a practical overview for clinicians and researchers involved in assessing upper limb function. It considers 25 upper limb assessments used in musculoskeletal care and presents a simple, straightforward comparative review of each. The World Health Organization International Classification on Functioning, Disability and Health (WHO ICF) is used to provide a relative summary of purpose between each assessment. Measurement properties of each assessment are provided, considering the type of data generated, availability of reliability estimates and normative data for the assessment. Copyright © 2007 John Wiley & Sons, Ltd.

Key words: Upper limb, assessment, function, rehabilitation

Introduction

Useful and reliable assessment of the upper limb is important in both research and clinical practice. There are a large number of upper limb assessments available, and selecting the most appropriate outcome measure to communicate the effectiveness of clinical interventions can be challenging. Each outcome measure will have its own measurement properties and will be likely to have a primary assessment focus – for example, body functions and structures, activities or impact on social participation.

This short report reviews 25 upper limb functional assessments (ULAs) currently reported in research and clinical practice. These assessments are categorized into their prime assessment focus; body functions and structures, activities and/or
impact on social participation (World Health Organization, 2002). The paper presents a summary of the reliability properties of each of the 25 ULAs, in order to provide a practical overview for clinicians and researchers, to assist with the decision of which ULA might be the most appropriate to use for defined assessment purposes.

Measurement

Measurement is essential for today’s health services. Robust outcome measures are required to provide reliable information to endorse evidence-based practice, and communicate the impact of disease and the effectiveness of health care interventions (Haywood, 2006). The introduction of the internal market in health care provision has forced the National Health Service to evaluate health service outcomes (Department of Health, 2007) and to strive to establish services that are efficient, cost effective and effective (Bowling, 2000).

There is no one distinct variable that can define and encompass functional ability. Fundamentally, outcome measures will utilize a number of variables to act as indicators (McDowell and Newell, 1996), and these can be compiled to form a clinical assessment to provide a clinically meaningful deduction from the measurement. Outcome measures can be categorized into subjective or objective measures. Subjective (self-report) measures rely on the patient’s or clinician’s report, whereas objective measures use data that have been generated by clinicians using validated equipment and standardized measurement protocols.

Choosing the most appropriate outcome measure(s), and having a clear understanding of their strengths and limitations, is important in both clinical and research terms. Using appropriate, valid and reliable outcome measures can improve understanding of how disease progresses, the level of structural impairment and how this impacts on the individual in terms of function and participation (Long and Scott, 1994). However, for specific assessment of the upper limb, challenges lie in choosing measures that are stable, valid and appropriate. There remains limited published data on reliability levels for many musculoskeletal ULAs (Ellis et al., 1997; Fowler and Nicol, 2001; Nordenskiöld, 2001).

This short report provides a review of 25 ULAs presented within the framework of the WHO ICF (World Health Organization, 2002) and provides practical information for both clinicians and researchers interested in the range and focus of ULAs used in musculoskeletal care.

Methods

An online search of AMED, CINAHL, Embase, Ovid Medline, Ovid Old Medline and Science Direct databases was undertaken to identify ULAs reported in the
literature. The search criteria are included in Appendix A. In addition, other ULAs were recommended for inclusion by clinical colleagues, based on frequency of use in practice, and ULAs from previous reviews were also sourced (McPhee, 1987; Jones, 1989; Light et al., 1999; Finch et al., 2002). Using the results from the literature search, discussions and previous reviews, 25 ULAs were identified.

Each ULA was presented and compared using the framework of the WHO ICF (World Health Organization, 2002). Assessments were classified under the ICF categories of Body Functions & Structures, Activities and Participation (Figure 1).

Each ULA was then considered within the ICF framework to clarify which domain(s) were being measured by each ULA (Figure 2). The criteria for assigning each ULA to a particular category of the ICF were as follows. If the ULA:

- Generated clinician-assessed reports on physiological function (e.g. range of motion or grip strength)
  – Body Functions & Structures category
- Generated clinician-observed reports and/or patient self-reports of activities of daily living (ADL) ability (e.g. dressing or feeding ability)
  – Activities category
- Generated patient self-reports on quality of life and participation (e.g. ability to access the community, work or leisure pursuits)
  – Participation category
- Generated clinician-assessed reports on physiological function and also included some clinician-assessed reports on functional ADL assessment
  – Body Functions & Structures category leading into Activities category

FIGURE 1. ICF framework (World Health Organization, 2002)
• Generated data from clinician-observed reports on ADL assessment and generated some interval/ratio data from clinician-assessed reports on physiological function
  – Activities category leading into Body Functions & Structures category
• Generated data from clinician-assessed reports on ADL assessment and with some ordinal measures of quality of life from patient self-reports
  – Activities category leading into Participation category
• Generated data patient self-report on quality of life with some clinician-assessed reports on ADL assessment
  – Participation category leading into Activities category
• Generated predominantly ordinal data from patient self-reports on quality of life and with some interval/ratio data generated from clinician-assessed reports on physiological function
  – Body Functions & Structures category and Participation category

The characteristics of each ULA were then considered. The following criteria were used to provide sub-categories for each ICF-classified ULA; these sub-categories were:

• Generated self-report (subjective) data
• Generated clinician-assessed outcome using standardized protocols and equipment (objective) data
• Generated ordinal or interval/ratio data
• Were designed as a disease or pathology-specific assessment
• Had associated published test-retest and inter-rater reliability statistics
• Had associated data from unimpaired participants

Results

Figure 2 categorizes each ULA within the framework of the WHO ICF. The level to which a ULA crosses over into another ICF category is subjective, but it provides a basis for understanding at a glance the content of a particular ULA under consideration.

Twenty-five ULAs are presented and summarized (full details of all ULAs are presented in Appendix B, including the full titles of those referred to by acronyms in Table 1). Table 1 presents the breakdown of each ULA into the type of data recorded by the measure; whether the measure is considered to produce objective or subjective outcomes; if it has been designed specifically for musculoskeletal conditions; and whether there is any evidence of published reliability values and normative data for comparison purposes.
Together with Table 1, the options for assessing the structural impairment, functional ability and impact of upper limb injury or disease can be reviewed and considered, with a practical understanding of the content, reliability and assessment remit of 25 commonly used regional ULAs.

**Discussion**

Comparing Table 1 with Figure 2 illustrates that clinician-observed and patient self-report ULAs tend to fall into the **Activities** or **Participation** sections of the ICF breakdown and generate ordinal (i.e. non-parametric) data. These assessments usually rely on categories of subjective observations and employ a ranked scoring system, whereby the individual or observer/assessor assigns a number, and level of ability or structural impairment for the completion of any given task. Subjective methods of assessment such as this provide informed observations, or a useful means of understanding an individual’s perspective. Commentators (Ashcroft, 1996; Carr, 1996a, 1996b; Wolfe and Pincus, 1999; Guillemin, 2000; Haywood, 2006) have rightly criticized the over-dependence of health professional-defined outcomes in musculoskeletal health care and encouraged the use of patient self-report measures. However, self-report and clinician-observed report measures are
### TABLE 1. Twenty-five upper limb assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Data Type</th>
<th>Assessment Type</th>
<th>Pathology</th>
<th>Evidence of Reliability</th>
<th>Normative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Research Arm Test*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Arthritis Hand Function Test**</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Box and Block Test*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Clawson Test</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Cochin Rheumatoid Hand Disability Scale*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>NA</td>
</tr>
<tr>
<td>DASH*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Frenchay Arm Test</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Fugl-Meyer Test</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Grooved Pegboard Test*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Jepsen Test*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>MacBain Hand Function Assessment</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Michigan Hand Outcomes Questionnaire*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>NA</td>
</tr>
</tbody>
</table>
### TABLE 1. Continued

Review of Upper Limb Assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Data Type</th>
<th>Assessment Type</th>
<th>Pathology</th>
<th>Evidence of Reliability</th>
<th>Normative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ordinal</td>
<td>Interval/Ratio</td>
<td>Both</td>
<td>Objective</td>
<td>Subjective</td>
</tr>
<tr>
<td>Nine Hole Peg Test*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Neill Hand Function Assessment**</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Capacities</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation of Hand Skill</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purdue Pegboard*</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SACRAH*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>SODA*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SADLE</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith Hand Function</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sollerman Hand Function Test*</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHAP*</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength-Dexterity Test</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Extremity Function Test</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker Test</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td><strong>16</strong></td>
<td><strong>3</strong></td>
<td><strong>23</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

§ = kappa and ‡ intraclass correlation coefficient; * = available from the internet/download or can be purchased, and ** = need to be made up in department

NA = not applicable
not without their limitations. Many will produce ranked ordinal data, and, as such, confidence intervals cannot then be used to extrapolate research findings (that use such measures) to a wider population. Dixon (1996) criticizes patient-defined outcomes in rheumatology as ‘soft’ and too subjective to be used in robust research. Van Den Ende et al. (1995) have also demonstrated further issues for consideration when, in the case of individuals with rheumatoid arthritis, there may be discordance between self-report and clinician-rated measures. Snaith et al. (1976) state that such measures are liable to be influenced by the individual’s ability to read/write and their wish to present themselves in a positive light. Studies by Ward (1994) and Spiegel et al. (1988) recount that self-report measures of pain and global status are confounded by the individual’s mood. Thus, on interpreting self-report measures, one needs to be mindful that changes in self-report measures might not be reflected by similar changes in clinician-rated measures. Self-report measures have been seen to be simple, inexpensive ways of obtaining good clinical data (Wolfe and Cathey, 1991), and in many instances they are easy to score and interpret for the clinician, adding valuable information to the clinical picture (Jacobs et al., 1992). As health is a social as well as a biological construct, individuals’ own views of their health state may be the most important to consider (Blaxter, 1990).

Finally, the assumption that self-report measures may be ‘soft’, and that using clinician-assessed variables are more reliable, is misguided. Both types of measurement suffer from the same threats to reliability and there are few guarantees to assure that any individual, whether a professional or a patient, will be consistent and reliable in their findings from day to day. If an outcome measure is badly designed, with weak psychometric properties, neither patients nor health care professionals will able to record reliable data.

By contrast, most clinician-assessed ULAs use measurements that are examined in terms of mathematical or physical variables, such as range of movement or velocity. These are argued to be more objective methods of assessment, and should provide an unbiased and comparable form of collecting information. Yet they may not truly reflect what matters to patients’ daily lives (Hewlett et al., 2001). For clinical researchers interested in generating wider generalizations from their research studies, this may be particularly relevant, as associated confidence intervals may be calculated, providing wider estimations than \( p \) values and significance levels alone (Gardner and Altman, 2000); something that cannot be achieved with ordinal data. Clinician-assessed ULAs tend to fall between the Body Functions & Structures section and the Activities section, and will produce internal/ratio data.

There are notable exceptions to these assignments, namely the clinician-observed ULAs in the Body Functions & Structures section. In the Frenchay Arm Test (De Souza et al., 1980), although each task measures an element of body function or structure, an assessor scores each test according to the particular criteria of...
the assessment. Some overlaps also occur when ULAs contain both subjective (self-report and clinician-observed report) and objective (clinician-assessed) elements. For example, the Arthritis Hand Function Test (Backman et al., 1991) could be assigned to both Body Functions & Structures and Activities, whereas the Score for Assessment and Quantification of Chronic Rheumatic Affections of the Hands (Leeb et al., 2003) could be assigned to both Body Functions & Structures and Participation.

One of the main criticisms of outcome measures that classify categories of function, assessed by clinicians’ observed reports, is that the categories may not have specific relevance to the individuals’ lifestyles or daily routines (Wylde et al., 2006). Relying solely on such methods of assessment can be open to assumption and misunderstanding of the personal impact of functional (dis)ability. It has been seen that musculoskeletal patients’ perspectives of disease and functional ability can differ from health professionals’ assessments at different stages of the disease (Kievit et al., 2005), and that the perception of activity and function may be substantially different between health professionals and their patients (Nothnagl et al., 2005). Conversely, objective methods of assessment concentrate solely on the feedback from a device, and therefore can be susceptible to errors of inaccuracy and inadequate calibration due to human error. Objective methods of assessment also provide no place for the individual’s perspective, relying only on the data received. More recently, outcome measures have been developed with the input of relevant patient groups, but the relevance of assessment tasks to patients has not always been considered in the development of measures (Carr, 1996b).

Limitations

This review is not without limitations. Formal appraisal of the literature using quality-scoring tools was not carried out; instead, more practical aspects of the assessments have been considered. This was intended to enhance the usability of the report for clinicians, but, in so doing, readers need to be aware that this short review does not necessarily follow the formal guidelines for systematic reviews. Moreover, the levels of validity of the ULAs are not compared or presented within this paper, and future consideration should be given to this area.

Conclusion

A review of the literature and online database search was undertaken, identifying 25 ULAs. The review highlights the content and specific characteristics of each ULA, and emphasizes which ULAs have associated information that testify to its
reliability. Each ULA has been categorized within the context of the WHO ICF framework, the results of which identify the type of data generated by the ULA, as well as whether the assessment is subjective or objective in nature. This review should serve as a reference for clinicians and researchers to locate an appropriate ULA, or a battery of tests that fulfil the needs for measuring the impact of disease, evaluating treatment effectiveness and communicating this to both patients and wider audiences. Historically, the emphasis when selecting an appropriate ULA has been on assessing limitation and impairment of function; a more positive approach in the future, focusing on ability and participation, would ultimately benefit the patients and maximize their potential for functional recovery.

References


Correspondence should be sent to Cheryl Metcalf, School of Electronics & Computer Sciences, Room 1001, Building 86, University of Southampton, Southampton, SO17 1BJ, UK. Tel: 02380 599204; Fax: 02380 592901. E-mail: cdm03r@ecs.soton.ac.uk


Copyright © 2007 John Wiley & Sons, Ltd
Appendix A

The following key words were included in the search in February 2007:

1. Scale
2. Measurement
3. Assessment
4. Test
5. Outcome
6. Function
7. Hand
8. Upper extremity
9. Upper limb
10. 7 AND 8 AND 9
11. 6 AND 10
12. 1 AND 11 OR 2 AND 11 OR 3 AND 11 OR 4 AND 11 OR 5 AND 11
13. 11 AND 12

The search was restricted to English language and human. Duplicates were then removed from 13.

Appendix B

**Action Research Arm Test**

**Arthritis Hand Function Test**

**Box and Block Test**

**Clawson Test**
Cochin Rheumatoid Hand Disability Scale

Disabilities of the Arm, Shoulder and Hand (DASH)

Frenchay Arm Test

Fugl-Meyer Test

Grooved Pegboard Test

Jebsen Test

MacBain Hand Function Assessment

Michigan Hand Outcomes Questionnaire

Nine Hole Peg Test

O’Neill Hand Function Assessment

Physical Capacities Evaluation of Hand Skill
Purdue Pegboard

Score for Assessment and Quantification of Chronic Rheumatic Affections of the Hands (SACRAH)

Sequential Occupational Dexterity Assessment (SODA)

Simulated Activities of Daily Living Examination (SADLE)

Smith Hand Function Evaluation

Sollerman Hand Function Test

Southampton Hand Assessment Procedure (SHAP)

Strength-Dexterity Test

Upper Extremity Function Test

Walker Test