

REVIEW ARTICLE

EVALUATION OF FUNCTIONAL OUTCOME MEASURES FOR THE HEMIPARETIC UPPER LIMB: A SYSTEMATIC REVIEW

Stephen Ashford, MSc, MCSP^{1,2}, Mike Slade, PhD³, Fabienne Malaprade, MCSP¹ and Lynne Turner-Stokes, DM, FRCP^{1,2}

From the ¹Regional Rehabilitation Unit, Northwick Park Hospital, ²Department of Palliative Care, Policy and Rehabilitation, School of Medicine, and ³Health Service and Population Research Department, Institute of Psychiatry, King's College London, London, UK

Objective: To identify valid and reliable outcome measures reflective of “real-life” active and passive function for application following focal rehabilitation interventions in the hemiparetic upper limb after stroke or brain injury.

Methods: A systematic review of the literature, incorporating a wide-based search including electronic databases, primary reports, abstracts and conference proceedings, was undertaken to identify measures, followed by a literature-based evaluation of the psychometric properties.

Results: Six measures met the review selection criteria, although 4 were different versions of the Motor Activity Log. None of the measures met all the psychometric evaluation criteria. The tools effectively formed a hierarchy with the ABILHAND and Motor Activity Log representing quite thoroughly evaluated tools of higher level “active function”. The Leeds Adult Spasticity Impact Scale addressed lower level tasks and passive function, but had little published psychometric evaluation.

Conclusion: As yet there is no single valid and reliable outcome measure available to capture the full range of “real-life” function in the hemiparetic upper limb. Validated tools are particularly required for passive and lower level function. The selection of measures for clinical evaluation will depend on the patient’s level of function and goals for treatment.

Key words: functional assessment, upper limb, outcome measure, rehabilitation.

J Rehabil Med 2008; 40: 787–795

Correspondence address: Stephen Ashford, Regional Rehabilitation Unit, Northwick Park Hospital, Watford Road, Harrow, Middlesex, HA1 3UJ, UK. E-mail: Stephen.Ashford@nwlh.nhs.uk

Submitted July 8, 2008; accepted August 28, 2008

INTRODUCTION

Hemiplegia is a common effect of brain injury or stroke, which may have significant impact on upper limb function. Whilst a proportion of patients will recover some degree of useful function in their upper limb following a stroke, for many the limb effectively becomes a passive object to be cared for either by the individual themselves or a carer.

Interventions for the hemiplegic upper limb may therefore be focused on a wide range of goals. At the higher level, interventions such as functional electrical stimulation (1) or constraint induced movement therapy (2, 3) may target “active” functional tasks, i.e. those involving voluntary activity of the affected upper limb. At a lower level, interventions such as spasticity management may be directed more towards goals in “passive” function, such as making it easier to get the arm through a sleeve or to maintain hygiene.

Outcome measurement is required to determine the effectiveness of rehabilitation interventions. Whether applied in clinical practice or for research, measures need to be valid, reliable and responsive to clinically relevant change. Global measures of function in daily activities, such as the Barthel Index (4), provide a general assessment of independence but are often unresponsive to focal interventions in the upper limb. Small changes, which may be extremely important to the patient and/or their carers, are easily lost amongst the larger number of unchanging items (5).

For these reasons, a number of focal motor function tests have been developed, for example the Wolf Motor Function Test (3, 6–8) and the Action Research Arm test (8). Conducted under close observation in the clinic, these may provide a more responsive and objective measure of motor activity. However, they do not necessarily reflect how the person actually functions in daily activities in their normal environment, and it is generally not practical to obtain this information through 24-hour observation in the home setting. Instead, this information on “real life” function may be gathered through direct enquiry from the patient and/or carer, for example using a task inventory administered by structured interview or a self-completion questionnaire (9).

The aim of this systematic review of the literature was to identify valid and reliable outcome measures that have been applied to assess changes following focal rehabilitation interventions in the hemiparetic upper limb in the context of stroke or brain injury, and are reflective of “real-life” function, for both active and passive tasks.

METHODS

The review was undertaken in 3 stages. In stage 1, a pool of possible measures was identified from a broad-based search. In stage 2, these

were narrowed down to those reflective of “real-life” performance (10). In stage 3, the published evidence was evaluated for psychometric properties for the selected measures. The Quality of Reporting of Meta-analyses (QUOROM) provides guidance on the most appropriate methods of presenting meta-analyses and review data and these principles were used in the presentation of data and results (11).

Stage 1: Identification of measures

Data sources. Electronic databases were searched, including: Medline, CINAHL, BIDS Science Citation Index, EMBASE, Specialised Register of Stroke Trials, National Health Service National Research, MRC Clinical Trials Directory, the Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects (DARE), Google, ProFusion and SIGLE (medical/rehabilitation grey literature). Other sources were: reference lists from papers identified, conference proceedings, books and book chapters and communication with lead authors of published studies and other researchers.

Search strategy. The search strategy followed a standardized format designed for Medline and adapted for the other databases used. The search strategy was constructed by the first author and confirmed with authors two and four prior to application. Following application of the search strategy by the first author, all authors were involved in reviewing the results. (The full search strategy is given in Appendix 1.)

We also hand-searched reference lists and contacted key authors to expand the breadth of the search or obtain further information about specific measures. Once the relevant studies for inclusion in the systematic review had been selected, the full publications were obtained and reviewed.

Criteria for selection and review are summarized in Table I. In stage 1, standardized outcome measures were identified for further consideration if they were: (i) applicable in the hemiparetic upper limb; (ii) included items measuring active or passive function.

Stage 2: Real-life relevance

In stage 2, selected measures were considered to have “real life” relevance if they reflected day-to-day performance in the person’s normal environment, as opposed to observation under test conditions.

Stage 3: Evaluation of psychometric properties

The names of measures identified in stage 2 were used as terms for a further search of the electronic databases to obtain original and any subsequent publications concerning their development and psychometric evaluation. The databases searched were: Medline (1996 to 7 May 2008); CINAHL (1982 to 7 May 2008) and EMBASE (1974 to 7 May 2008).

On the basis of this published literature, the psychometric properties of each measure were evaluated against the following review criteria, by at least 2 reviewers.

- Practicality for use in everyday practice: time to complete, burden, readability.

Table I. Summary of review criteria

Stage 1 – Selection to identify standardized outcome measures

1. Relevance – applicable in the hemiparetic upper limb.
2. Include items measuring:
 - a. Active function
 AND/OR
 - b. Passive function

Stage 2 – Real-life relevance (as opposed to under test conditions).

3. Assessed in a manner reflective of “real life” function

Stage 3 – Literature-based evaluation of evidence for psychometric properties of measures

4. Practical to apply in everyday clinical practice
5. Valid and reliable for upper limb function evaluation
6. Responsive to change occurring as a result of intervention.

- Validity and reliability: content validity, internal consistency, construct validity, floor and ceiling effect, test-retest reliability, agreement.
- Responsiveness to change: demonstration of change following focal upper limb intervention, interpretability and minimal clinically important difference (MCID).

Descriptive information was extracted for each of the selected instruments, including: the items in the measure; the methods of administration; and the method of scoring applied to the measure. The quality criteria developed by Terwee et al. (12) and used by Bot et al. (13) for a “clinometric evaluation of shoulder disability questionnaires” from those produced by the Scientific Advisory Committee of the Medical Outcomes Trust (14) were used to evaluate the quality of these instrument properties. The full criteria used with minor alterations to those produced by Bot et al. (13) are given in Appendix 2 and the application in this instance is described in assessment methods below. Two reviewers, authors one and three, applied the criteria in evaluating each measure independently of each other. Findings were then compared and any discrepancies resolved through discussion. The option was available for a third reviewer to resolve any areas of none agreement following comparison, but this was not needed.

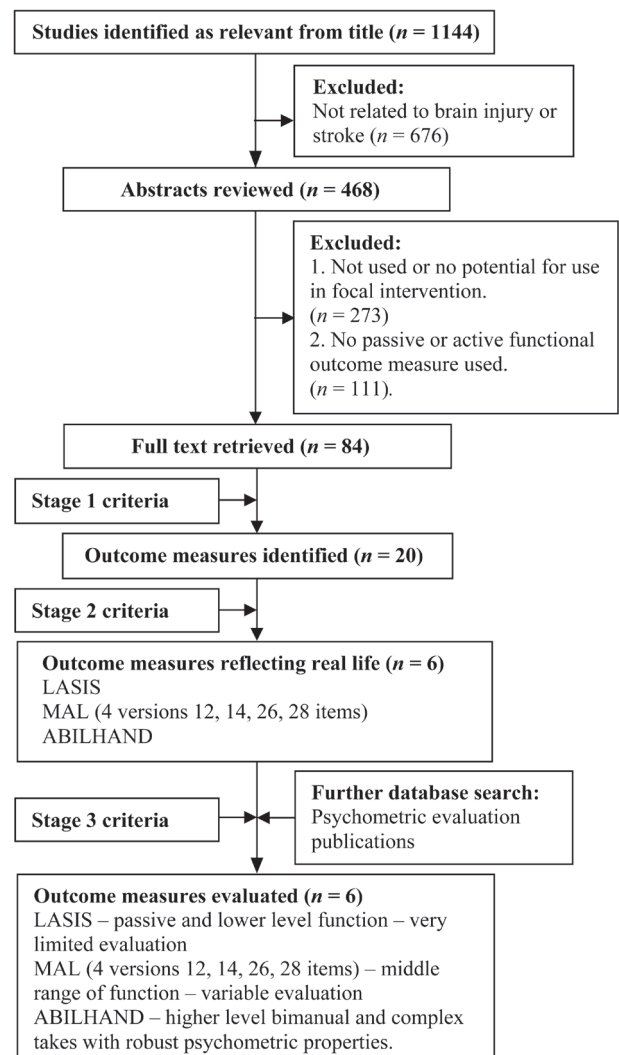


Fig. 1. Quality of Reporting of Meta-analyses (QUOROM) measure selection flow diagram. LASIS: Leeds Adult Spasticity Impact Scale; MAL: Motor Activity Log.

Procedure for evaluation of each measure

Practical burden. The method of scoring was used to rate administrative burden. The rating system was as follows: "easy" (+), when categorized items were simply summed; "moderate" (\pm), when an ordinal scale or a visual analogue scale was used to quantify individual items, and "difficult" (-) when a scale was applied in combination with a formula. Timing for completion of the measure was also rated as positive for measures completed within 10 min.

Validity. The instruments were evaluated for content and construct validity. A positive rating for content validity was given when there was evidence that patients, their carers or other experts had been consulted regarding the initial selection of items (e.g. through focus groups, surveys, etc.) or had provided evaluation/feedback as part of the development. A positive rating for construct validity was given if there was evidence that the measure was based on hypothetical constructs that had been tested and supported during its evaluation.

The positive rating for internal consistency was given if the factor structure of the measure had been tested through factor analysis, and/or when ratings for Chronbach's alpha were between 0.70 and 0.90 for each dimension or subscale, based on the recommendations by Bot et al. (13).

Floor and ceiling effects were considered present if more than 15% of respondents achieved the highest or lowest possible score, respectively. Floor effects were also considered present if the measure included only bilateral and/or complex tasks.

Reproducibility. Test-retest reliability was rated as positive if repeat testing of the same condition had yielded comparable results, e.g. an intraclass correlation coefficient of greater than 0.70 for total scores. In item-by-item analyses, agreement was also rated as positive if it had been evaluated and shown to be satisfactory using accepted statistical methods, such as the Kappa coefficient or standard error of measurement.

Responsiveness. Responsiveness was rated as positive if the measure had demonstrated significant change in response to intervention, in the context of an appropriate study design (see full criteria Appendix 2).

Interpretability. Interpretability is the degree to which qualitative meaning can be assigned to quantitative scores (15). Positive ratings were given if: (i) at least 2 types of information were given to aid in understanding of the scores, such as means and standard deviations of the score totals before and after treatment; or (ii) information was given

in relation to other clinical variables that might be expected to change; or (iii) information was given on the minimum change in score that might be clinically meaningful (the minimal clinically important difference).

RESULTS

Stage 1: Identification of measures

A summary of the stages of review, according to QUORUM, is given in Fig. 1. The search yielded 1144 studies, including primary reports, abstracts and conference proceedings. Of these, 84 studies were identified following initial review of the abstracts as including measures of functional outcome following focal upper limb intervention, yielding a total of 20 outcome measures after stage 1 (Fig. 1). The properties and initial evaluation of these 20 measures are shown in Table II.

Stage 2: Real-life relevance

Six measures were identified that met both stage 1 and stage 2 criteria (i.e. were relevant to real life functional performance). These were the Leeds Adult Spasticity Impact Scale (LASIS), ABILHAND and Motor Activity Log (MAL), which had 4 different versions with 14, 26, 28 and 12 items, respectively (see Table II). All 4 versions of the MAL had undergone some elements of psychometric evaluation and were therefore all included for further analysis. The scaling methods, number of items and methods of administration for these measures are shown in Table III.

Stage 3: Evaluation

The detailed evaluation of the properties of the selected measures is presented in Table IV. The quality criteria developed by Bot et al. (13) were used to evaluate the quality of each instrument's properties, summarizing each variable as adequate (+), doubtful (\pm), or poor quality (-) or as unknown (?) if insufficient information was available.

Table II. Identified outcome measures

Number	Outcome measures	Reflective of real-life	Apply hemiparetic upper limb	Active function elements	Passive function elements	Evidence of formal psychometric testing in neurological rehabilitation
1	Leeds Adult Spasticity Impact Scale (LASIS) (22–29)	√	√	√	√	
2	Disability Assessment Scale (DAS) (30–34)		√	√	√	√
3	Motor Activity Log (3, 7, 8, 16, 35, 36)	√	√	√		√
4	Reduced Upper Extremity Motor Activity Log (37)	√	√	√		√
5	Motor Activity Log 26 item (Dutch – Translated) (16)	√	√	√		√
6	Motor Activity Log 28 Item (17)	√	√	√		√
7	ABILHAND (19, 20)	√	√	√		√
8	Wolf Motor Function Test (3, 6–8)		√	√		√
9	Box and Block Test (38–41)		√	√		√
10	Action Research Arm Test (8, 42–45)		√	√		√
11	Frenchay Arm Test (45–47)		√	√		√
12	Rivermead Motor Assessment Arm Scale (44, 45)		√	√		√
13	Nine-Hole Peg Test (7, 45, 48, 49)		√	√		√
14	Upper Extremity Function Test (37)		√	√		√
15	Motor Club Assessment (45)		√	√		√
16	Jebson Hand Function Test (41, 45, 50, 51)		√	√		√
17	Fugl-Meyer Upper Limb Test (42, 45, 52)		√	√		√
18	Purdue Peg Board Test (45, 53, 54)		√	√		√
19	Arm Motor Ability Test (55)		√	√		√
20	Chedoke Arm and Hand Activity Inventory (56)		√	√		√

√: attribute is present.

Table III. Selected measures of function

Outcome measure	Method and procedure of scoring	Context for development
Leeds Adult Spasticity Impact Scale (LASIS) (22–29)	Items: 12 Scoring: Patients or carers, over the past 7 days. Items rated between 0–4. Scores summed and divided by the number of questions answered. Administration: Semi-structured interview.	Spasticity intervention
Motor Activity Log (MAL-14) (3, 7, 8, 16, 35, 36)	Items: 14. Scoring: By patients, over the past 7 days. Scores range from 0 (never use the affected limb for this activity) to 5 (always use the affected arm for this activity). Subjects are rated on the amount they use their paretic arm (“amount scale”) and on the quality of their movement during the functional activities (“how well scale”). Administration: Semi-structured interview.	CIMT
26-item MAL (MAL-26) (16)	Items: 14 original items, 11 additional items and 1 optional item chosen by the patient. Scoring: By patients, over the past 7 days as per the MAL. Administration: Semi-structured interview.	CIMT
28-item MAL (MAL-28) (17)	Items: 28 Scoring: By patients, over the past 7 days or past 3 days. Administration: Semi-structured interview.	CIMT
Reduced Upper Extremity Motor Activity Log (MAL-12) (37)	Items: 12 Scoring: By patients, over the past 7 days as per the MAL. Administration: Semi-structured interview.	CIMT
ABILHAND (19, 20)	Items: 23 Scoring: Patients asked to estimate their ease or difficulty of performing each task (without help) only on tasks they have performed. Score categories were 0 (impossible), 1 (any difficulty) and 2 (easy). Administration: Semi-structured interview.	Chronic Stroke rehabilitation

CIMT: Constraint Induced Movement Therapy.

Table V shows the item content of each of the 6 identified measures, which could be broadly placed in a hierarchy of increasing difficulty. At the lowest level, the LASIS, primarily included passive function items. In the middle of the range the MAL contained items of active function, increasing in the following order: MAL-14, MAL-26, MAL-28, MAL-12. At the uppermost level, the ABILHAND contained complex items often requiring bilateral hand use, and at this level, the order of difficulty has been confirmed by Rasch analysis (16, 17).

Administrative burden and time for completion

The administrative burden was adequate for all measures apart from the LASIS, which required the calculation of the measure total; however, this calculation in practice was not complex. The calculation of the LASIS involved totalling the item scores and then dividing the total by the number of items answered. This results in total scores between 0 and 4 representing disability or carer burden, however this could

be based on 1 item answered or on all 12, which may not be representative of actual disability or function in the arm and may have implications for the validity of the measure.

All of the measures were designed for administration by a clinician through structured interview with the patient and/or carer, which generally requires a significant allocation of clinician time (18). The LASIS, all versions of the MAL (except the MAL-12), and the ABILHAND were thought to involve a time for completion of greater than 10 min.

Validity

Internal consistency was demonstrated in 3 measures; 2 versions of the MAL (14 and 28) and ABILHAND. Construct validity had also been addressed in the same 2 versions of the MAL (14 and 28) and ABILHAND.

Information on floor and ceiling effects was difficult to identify or not formally addressed in the majority of measures. However, given that the tools have a hierarchical relationship in their item

Table IV. Psychometric evaluation from the literature of the selected measures

Measure	Time	Administrative burden	Content validity	Internal consistency	Construct validity	Floor/ceiling effect	Reliability	Agreement	Responsiveness	Interpretability	MCID
LASIS	+	±	?	?	?	?	?	?	?	?	?
MAL-14	–	+	?	+	±	±	–	–	–	+	–
MAL-26	–	+	?	?	?	?	?	?	?	?	?
MAL-28	–	+	?	+	±	±	±	±	?	?	?
MAL-12	+	+	?	?	?	?	?	?	?	?	?
ABILHAND	–	+	+	+	+	–	+	+	+	+	+

MCID: minimal clinically important difference; method or result was rated as: + adequate; ± doubtful; – poor; ? no data available; LASIS: Leeds Adult Spasticity Impact Scale; MAL: Motor Activity Log.

content, it may be expected that the LASIS would have ceiling effects in a higher function group, and similarly the MAL and ABILHAND would have floor effects for detecting changes in lower level and passive functional tasks.

Reproducibility

Test–retest reliability evaluation was documented in 3 measures: the ABILHAND and MAL versions 14 and 28 (19, 20). Adequate methods have been used in the ABILHAND, but were less convincingly applied in the MAL-14 and MAL-28.

Responsiveness

Responsiveness was demonstrated in the ABILHAND and was also assessed in the MAL-14 (16, 19, 20). However, the change in the MAL-14 did not correspond to change identified by other measures, and responsiveness was therefore rated as inadequate in this evaluation (16). Responsiveness in both measures was evaluated in post-stroke hemiplegic patients who had good return of arm movement and related function.

Interpretability

Interpretation of specific scores with respect to qualitative meaning had been evaluated only in the MAL-14 and ABILHAND. The ABILHAND had been evaluated using Rasch analysis and demonstrated a clear gradation of increasing ability of different items within the scale (19, 20). It was therefore given a positive rating. The MAL, however, did not show an adequate relationship between overall scores or achievement of individual items and qualitative meaning, and the MCID was not clear. It was therefore given a negative rating.

DISCUSSION

This systematic review identified 6 measures (including the 4 versions of the MAL), which had been used in the published literature to evaluate function reflective of real life or actual performance. The 6 measures appeared to fall broadly into a hierarchy of increasing difficulty, with the LASIS addressing passive function and low-level active function, such as using the affected hand to hold and stabilize objects. The MAL and ABILHAND were more comprehensive measures for active function, but with more complex activities, representing a wide range of activities, including unilateral and bimanual function.

In terms of their psychometric properties the LASIS, MAL-12 and MAL-26 have received only scant evaluation and met only one of the

Table V. Items included in each measure

Functional items	LASIS	MAL-14	MAL-26	MAL-28	MAL-12	ABILHAND
<i>Passive Function Items</i>						
Cleaning the palm affected hand	1					
Cutting fingernails affected hand	2		25*			4*
Cleaning the affected elbow	3					
Cleaning the affected armpit	4					
Cleaning the unaffected elbow	5					
Putting arm through coat sleeve	6	1*	1*			
Difficulty putting on a glove	7					
Difficulty rolling over in bed	8					
Doing physiotherapy exercises to arm	9					
<i>Active Function Items</i>						
Difficulty balancing standing	10					
Difficulty balancing walking	11					
Hold object steady, use other hand (jar ^a)	12					10 ^a
Steady myself while standing		2	2			
Carry an object from place to place		3	3	23	12	
Pick up fork of spoon, use for eating		4	4	24	10	
Comb hair		5	5	25		
Pick up cup by handle		6	6	26	11	
Hand craft/card playing		7	7			
Hold a book for reading		8	8			
Use towel to dry face or other body part		9	9			
Pick up a glass		10	10	20	5	
Pick up toothbrush and brush teeth		11	11	21	6	
Shaving/make-up		12	12			
Use a key to open a door		13	13		7	
Letter writing/typing		14	14		8	
Poor coffee/tea			15			
Peel fruit or potatoes			16			3
Dial number on the phone			17			
Open/close a window			18			
Open an envelope			19			
Take money out of a wallet or purse			20			
Undo buttons on clothing			21			
Buttons on clothing (shirt ^a , trousers ^b)			22	27 ^a		13 ^a 17 ^b
Undo a zip			23			
Do up a zip (jacket ^a , trousers ^b)			24			11 ^a 21 ^b
Other optional activity			26			
Turn on a light with a light switch				1		
Open a drawer				2		
Remove item of clothing from drawer				3		
Pick up phone				4	1	
Wipe kitchen counter				5		
Get out of car				6		
Open refrigerator				7		
Open a door by turning a door knob				8	2	
Use a TV remote control				9		
Wash your hands				10		
Turn water on/off with faucet				11	4	
Dry your hands				12		
Put on your socks				13		
Take off your socks				14		
Put on your shoes				15		
Take off your shoes				16		
Get up from chair with arm rests				17		
Pull chair away from table before sitting				18		
Pull chair toward table after sitting				19		
Use a key to unlock a door				22		
Eat half a sandwich or finger food				28	3	
Use removable computer storage					9	
Hammer a nail						1
Thread a needle						2
Wrap gifts						5
File nails						6

Table V. *contd*

Functional items	LASIS	MAL-14	MAL-26	MAL-28	MAL-12	ABIL-HAND
Cut meat						7
Peel onions						8
Shell hazel nuts						9
Open pack of chips						12
Sharpen pencil						14
Spread butter						15
Fasten "snap"						16
Cap of a bottle						18
Open mail						19
Squeeze toothpaste						20
Unwrap chocolate						22
Wash hands						23

Items in the table are given the number at which they appear in order in the measure. Items in LASIS included under passive function all asked respondents "how difficult" a task was to undertake related to care of the limb by the patient him or herself or a carer. *Items in the passive function section included in MAL-14, MAL-26 or ABILHAND could be done either passively or with more active involvement by the individual, with the focus being on active involvement in these measures. ^a and ^b refer to specific subjects used for the functional items in the specified measure. LASIS: Leeds Adult Spasticity Impact Scale; MAL: Motor Activity Log.

stage 3 review criteria each. The MAL-14 and MAL-28 have been more extensively validated, but although they each met 2 criteria, their performance was doubtful on the remainder. Only the ABILHAND has been thoroughly evaluated and was shown to meet 9 of the 11 criteria, although it failed on time for completion and floor effects in a more dependent group.

The implication of these findings for clinicians is that there are several tools available and the choice of measurement tool will depend on the patient's current level of function and the anticipated goals for treatment.

- The LASIS is likely to be useful for individuals who have little or no active movement or function, but nevertheless have care and maintenance issues related to the hand and upper limb.
- The MAL-14 contains more unilateral and simple items, which may be useful for detecting change in individuals who have some, but limited, arm function.
- The MAL-26 also includes these 14 items but adds a further 12, including some tasks (such as peeling potatoes or taking money out of a purse) that require 2 hands.
- The MAL-28 includes 7 items from the MAL-14/26, but adds a further 21 functional tasks, some of which challenge reach and strength (such as putting on shoes and socks or pulling a chair towards a table after sitting), while the MAL-12 represents a short version that spans the entire range of MAL items.
- The ABILHAND has 6 items in common with these scales, but adds a further 16, all of which are more complex bilateral tasks. It is therefore likely to be useful for patients functioning at a higher level (see Table III for details of the measures and Table V for the included items).

Limitations of the review

This review has presented a number of challenges and limitations.

Identification of measures. Our starting point was the scientific literature, and it is possible that we have missed tools that are

used in clinical practice, but have not been applied in research. However, as we wished to identify measures for which there was some evidence of psychometric evaluation, we considered it appropriate to base our initial search in the research literature.

The possibility of missing studies. Our secondary search for literature regarding psychometric evaluation included identification of references from the original publications and a search of the cited literature based on the name(s) of the measures (the LASIS has had more than one name in the course of its evolution). Again, it is possible that this narrower search may have missed some of the grey literature. However, it was anticipated that these other publications would generally be of lower quality, and would not add significantly to the body of evidence that was found.

Evaluation of psychometric properties. The use of formal evaluation criteria for psychometric properties supported a detailed assessment of the published psychometric properties of the respective measures. The criteria published by Terwee et al. in 2007 (12) were based on earlier work by the same group (Bot et al. in 2004 (13)). The criteria were not developed for the context of hemiplegia, although the original review by Bot et al. (13) was a systematic review of shoulder disability questionnaires for application following musculoskeletal injury. It is interesting to note that our review did not identify any of the same measures evaluated by Bot et al. (13), and this is not surprising due to the different patient populations considered. For example, Bot et al. (13) identified the Disabilities of the Arm, Shoulder and Hand (DASH) (21) questionnaire, which best met their search criteria and had undergone the most extensive psychometric evaluation. The DASH is a measure of everyday active function, administered by self-completion questionnaire. This approach would have significant advantages in reducing the clinical time required to administer the tool. However, it is designed to assess higher level function and, like the ABILHAND, is likely to show significant floor effects in a neurologically impaired population. At the other end of the scale, none of Bot's measures contained any passive function items. Passive function applies particularly in the context of neurological damage, but could also have relevance in very severe musculoskeletal conditions, such as deforming arthritis. Once again this emphasizes the wide range of functional activities of the upper limb.

Summary and future research

For this evaluation of the published literature, it appears that there is a reasonable selection of validated tools available for the evaluation of "real life" active function in the hemiparetic upper limb. However, as yet there are none that provide a comprehensive assessment of active and passive function. Depending on the difficulty of the goals for treatment, clinicians may select from the 6

measures presented in this review, but should be aware of the limitations in psychometric evaluation for some of these measures.

The ABILHAND appears to be a robust tool for higher levels of function, and the range of different versions of the MAL allow for a more or less detailed assessment of abilities in the middle range. However, there is a dearth of validated tools to assess passive and lower level function. Moreover, all of the measures identified in this review are administered by structured interview and have implications for clinician's time if routinely assessed as part of clinical practice. The development of self-completed questionnaires has the potential to improve the practicality of application, although some patients with neurological disability may find this difficult, especially if they have significant cognitive or communicative problems. Further exploration and development of measures in a variety of different formats is now required.

REFERENCES

- Cisari C, Carda S. Functional electrical stimulation of the upper limb in poststroke adult rehabilitation. *Eura Medicophys* 2002; 38: 195–202.
- Page SJ, Elovic E, Levine P, Sisto SA. Modified constraint-induced therapy and botulinum toxin A: a promising combination. *Am J Phys Med Rehabil* 2003; 82: 76–80.
- Dettmers C, Teske U, Hamzei F, Uswatte G, Taub E, Weiller C. Distributed form of constraint-induced movement therapy improves functional outcome and quality of life after stroke. *Arch Phys Med Rehabil* 2005; 86: 204–209.
- Wade DT, Collin C. The Barthel ADL index: a standard measure of physical disability? *Int Disability Stud* 1988; 10: 64–67.
- Ashford S, Turner-Stokes L. Goal attainment for spasticity management using botulinum toxin. *Physiother Res Int* 2006; 11: 24–34.
- Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke* 2001; 32: 1635–1639.
- Ring H, Rosenthal N. Distributed form of constraint-induced movement therapy improves functional outcome and quality of life after stroke. *Arch Phys Med Rehabil* 2005; 86: 204–209.
- Page S, Levine P. Forced use after TBI: promoting plasticity and function through practice. *Brain Inj* 2003; 17: 675–684.
- Jones L, editor. *Jebson test of hand function* (British Version). London: National Hospital for Neurology and Neurosurgery; 1990.
- Sheehan GL. Botulinum treatment of spasticity: why is it difficult to show a functional benefit? *Curr Opin Neurol* 2001; 14: 771–776.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analysis of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999; 354: 1896–1900.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker JH, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.
- Bot SD, Terwee CB, van der Windt DA, Bouter Lex M, deVet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004; 63: 335–341.
- Scientific Advisory Committee of the Medical Outcomes Trust. *Assessing health status and quality-of-life instruments: attributes and review criteria*. *Qual Life Res* 2002; 11: 193–205.
- Nunnally J, Bernstein ICH, editor. *Psychometric theory*. 3rd edn. New York: McGraw-Hill; 1994.
- van der Lee JH, Beckerman H, Knol DL, de Vet HCW, Bouter LM. Clinimetric properties of the motor activity log for the assessment of arm use in hemiparetic patients. *Stroke* 2004; 35: 1–5.
- Uswatte G, Taub E, Morris D, Light K, Thompson P. The Motor Activity Long-28 Assessing daily use of the hemiparetic arm after stroke. *Neurology* 2006; 67: 1189–1194.
- Brashear A, Zafonte R, Corcoran M, Galvez-Jimenez N, Gracies JM, Gordon MF, et al. Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb post-stroke spasticity. *Arch Phys Med Rehabil* 2002; 83: 1349–1354.
- Penta M, Thonnard J-L, Tesio L. ABILHAND: a Rasch-built measure of manual ability. *Arch Phys Med Rehabil* 1998; 79: 1038–1042.
- Penta M, Tesio L, Arnould C, Zancan A, Thonnard J-L. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients – Rasch-based validation and relationship to upper limb impairment. *Stroke* 2001; 32: 1627–1634.
- Hudak P, Amadio P, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). *Am J Ind Med* 1996; 29: 602–608.
- Bakheit AMO, Thilman AF, Ward AB, Poewe W, Wissel J, Muller M, et al. A randomized, double-blind, placebo-controlled, dose-ranging study to compare the efficacy and safety of three doses of botulinum toxin type A (Dysport) with placebo in upper limb spasticity after stroke. *Stroke* 2000; 31: 2402–2406.
- Bakheit AM, Pittcock S, Moore AP, Wurker M, Otto S, Erbguth F, et al. A randomized, double-blind, placebo-controlled study of the efficacy and safety of botulinum toxin type A in upper limb spasticity in patients with stroke. *Eur J Neurol* 2001; 8: 559–565.
- Bakheit AM, Sawyer J. The effects of botulinum toxin treatment on associated reactions of the upper limb on hemiplegic gait – a pilot study. *Disabil Rehabil* 2002; 24: 519–522.
- Bakheit AMO, Fedorova NV, Skoromets AA, Timerbaeva SL, Bhakta BB, Coxon L. The beneficial antispasticity effect of botulinum toxin type A is maintained after repeated treatment cycles. *J Neurol Neurosurg Psychiatry* 2004; 75: 1558–1561.
- Bhakta BB, Cozens JA, Chamberlain MA, Bamford JM. A randomised double blind placebo controlled trial of botulinum toxin treatment on the disabling effects of severe arm spasticity in stroke. *Cerebrovasc Dis* 1999; 9 Suppl 1: 124.
- Bhakta BBC, Cozens JA, Chamberlain MA. The impact of botulinum toxin type-A (Dysport) treatment on the disabling effects of severe upper limb spasticity following stroke: a randomized, double-blind, placebo-controlled trial. *Toxins* 1999; 99.
- Bhakta BB, Cozens JA, Chamberlain MA, Bamford JM. Impact of botulinum toxin type A on disability and carer burden due to arm spasticity after stroke: a randomised double blind placebo controlled trial. *J Neurol Neurosurg Psychiatry* 2000; 69: 217–221 [Erratum in *J Neurol Neurosurg Psychiatry* 2001; 70: 821].
- Bhakta BB, Cozens JA, Chamberlain MA, Bamford JM. Randomized double-blind placebo-controlled trial of botulinum toxin treatment on the disabling effects of severe arm spasticity in stroke. *Clin Rehabil* 2000; 14: 213.
- Brashear AM, McAfee AL, Kuhn ER, Ambrosius WT. An open-label trial of botulinum toxin type B in upper limb spasticity. *Arch Phys Med Rehabil* 2001; 82: 1340.
- Brashear A, Zafonte R, Corcoran M, Galvez-Jimenez N, Gracies JM, Gordon MF, et al. Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity. *Arch Phys Med Rehabil* 2002; 83: 1349–1354.
- Brashear A, Gordon MF, Elovic E, Kassicheh VD, Marciniak C, Do M, et al. Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke (comment). *N Engl J Med* 2002; 347: 395–400.
- Brashear AM, McAfee AL, Kuhn ER, Ambrosius WT. Treatment with botulinum toxin type B for upper-limb spasticity. *Arch Phys Med Rehabil* 2003; 84: 103–107.
- Brashear A, McAfee AL, Kuhn ER, Fyffe J. Botulinum toxin type B in upper-limb poststroke spasticity: A double-blind, placebo-controlled trial. *Arch Phys Med Rehabil* 2004; 85: 705–709.
- Yelnik AP. Spasticité du membre supérieur après AVC, traitements pharmacologiques. *Ann Readapt Med Phys* 2004; 47: 559–575.
- van Kuijk AA, Geurts AC, Bevaart BJ, van Limbeek J. Treatment of upper extremity spasticity in stroke patients by focal neuronal or neuromuscular blockade: a systematic review of the literature. *J Rehabil Med* 2002; 34: 51–61.
- Popovic MB, Popovic DB, Sinkjaer T, Stefanovic A, Schwirtlich L. Clinical evaluation of functional electrical therapy in acute hemiplegic subjects. *J Rehabil Res Dev* 2003; 40: 443–453.

38. Alon G, Sunnerhagen KS, Geurts ACH, Ohry A. A home-based, self-administered stimulation program to improve selected hand functions of chronic stroke. *NeuroRehabil* 2003; 18: 215–225.
39. Desrosiers J, Bravo G, Hebert R, Dutil E, Mercier L. Validation of the box and block test as a measure of dexterity of elderly people: reliability, validity and norms studies. *Arch Phys Med Rehabil* 1994; 75: 751–755.
40. Higgins J, Mayo NE, Desrosiers J, Salbach NM, Ahmed S. Upper-limb function and recovery in the acute phase poststroke. *J Rehabil Res Dev* 2005; 42: 65–76.
41. Lannin NA, Herbert RD. A home-based, self-administered stimulation program to improve selected hand functions of chronic stroke. *NeuroRehabil* 2003; 18: 215–225.
42. Berglund K, Fugl-Meyer AR. Upper extremity function in hemiplegia. A cross-validation study of two assessment methods. *Scand J Rehabil Med* 1986; 18: 155–157.
43. Broeks JG, Lankhorst GJ, Rumping K, Prevo AJ. The long-term outcome of arm function after stroke: results of a follow-up study. *Disability* 1999; 21: 357–364.
44. Parry RH, Lincoln NB, Vass CD. Effect of severity of arm impairment on response to additional physiotherapy early after stroke. *Clin Rehabil* 1999; 13: 187–198.
45. Wade DT. Measurement in neurological rehabilitation. Oxford: Oxford University Press; 1992.
46. Alon G, Dar A, Katz-Behiri D, Weingarden H, Nathan R. Efficacy of a hybrid upper limb neuromuscular electrical stimulation system in lessening selected impairments and dysfunctions consequent to cerebral damage. *J Neurol Rehabil* 1998; 12: 73–79.
47. Lagalla G, Danni M, Reiter F, Ceravolo MG, Provinciali L. Post-stroke spasticity management with repeated botulinum toxin injections in the upper limb. *Am J Phys Med Rehabil* 2000; 79: 377–384.
48. Rodgers H, editor. What is the clinical effect and cost effectiveness of treating upper limb spasticity due to stroke with botulinum toxin? London: National Research Register; 2008.
49. Lindberg P, Schmitz C, Forsberg H, Engardt M, Borg J. Effects of passive-active movement training on upper limb motor function and cortical activation in chronic patients with stroke: a pilot study. *J Rehabil Med* 2004; 36: 117–123.
50. Richardson D, Edwards S, Sheean GL, Greenwood RJ, Thompson AJ. The effect of botulinum toxin on hand function after incomplete spinal cord injury at the level of C5/6: a case report. *Clin Rehabil* 1997; 11: 288–292.
51. Jones L, Lewis Y, Harrison J, Wiles CM. The effectiveness of occupational therapy and physiotherapy in multiple sclerosis patients with ataxia of the upper limb and trunk. *Clin Rehabil* 1996; 10: 277–282.
52. Berglund K, Fugl-Meyer AR. Upper extremity function in hemiplegia. A cross-validation study of two assessment methods. *Scand J Rehabil Med* 1986; 18: 155–157.
53. Desrosiers J, Hebert R, Bravo G, Dutil E. The Purdue pegboard test: normative data for people aged 60 and over. *Disabil Rehabil* 1995; 17: 217–224.
54. Hurvitz EA, Conti GE, Brown SH. Changes in movement characteristics of the spastic upper extremity after botulinum toxin injection. *Arch Phys Med Rehabil* 2003; 84: 444–454.
55. Kopp B, Kunkel A, Flor H, Platz T, Rose U, Mauritz K, et al. The arm motor ability test: Reliability, validity and sensitivity to change of an instrument for assessing disabilities in activities of daily living. *Arch Phys Med Rehabil* 1997; 78: 615–620.
56. Barreca SR, Stratford P, Lambert CL, Masters LM, Streiner D. Test-retest reliability, validity and sensitivity of the Chedoke arm and hand activity inventory: a new measure of upper limb function for survivors of stroke. *Arch Phys Med Rehabil* 2005; 86: 1616–1622.
57. Dickersin K, Scherer R, Lefebvre C. Systematic reviews: identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286–1291.

APPENDIX 1. Systematic review search strategy

The following data sources were searched:

1. Medline search based on the strategy outlined by Dickersin et al. (57). (1996 to 7 May 2008).
2. CINAHL (1982 to 7 May 2008).
3. BIDS Science Citation Index (until 7 May 2008).
4. Embase (1974 to 7 May 2008).
5. Relevant trials were identified in the Specialised Register of Stroke trials (to 7 May 2008).
6. National Health Service National Research Register, MRC Clinical Trials directory, Database of Abstracts of Reviews of Effects (DARE), Google, ProFusion and SIGLE (medical/rehabilitation grey literature) (to 7 May 2008).
7. The Cochrane Database of Systematic Reviews (to 7 May 2008).
8. Reference lists from papers identified above.
9. Conference proceedings, books and book chapters.
10. Communication with lead authors of published studies and other researchers.

The following search strategy was applied:

The search strategy for data sources 1 and 2

1. (hemiplegia or hemiplegic or hemiparesis or hemiparetic)
2. AND (arm or upper limb or hand or shoulder)
3. AND (stroke or post stroke or CVA)
4. OR (brain haemorrhage or haemorrhage or haematoma or hematoma)
5. OR (brain injury)
6. OR (brain tumour or tumor)
7. OR (brain infection or encephalitis or abscess)

Was used to identify the clinical group.

8. AND (Outcome measurement (MESH))
9. OR Outcome assessment
10. AND (function*)
11. OR activity).

Was used to identify the outcome measurement sub-group.

Recommended by the Cochrane Collaboration (Dickersin et al. (57)).

The search strategy for data sources 3, 4, 5 and 6 was modified to that given above due to the more limited search capacity available in those search tools. Single search term searches were undertaken and then combined to allow full searching in those data sources. Search terms were altered when required to comply with “key” terms used in other databases as appropriate.

The search strategy for data source 6 involved a further modified and simplified search strategy from that used in 1 and 2.

Reference lists and textbooks were searched by hand (8 and 9). Textbooks were either identified in the wider search (as indicated above) or through searching the catalogues of a number of medical libraries or through discussion with “expert” clinicians in rehabilitation medicine or physiotherapy. Conference proceedings were identified from the search strategies applied in 1 and 2 searching medical libraries and discussion with expert clinicians or researchers as well as searching key textbooks.

Where appropriate as indicated by the literature or through discussion with other clinicians or researchers, published authors and researchers were contacted about the scales that they have used or developed (11). Contact was usually by e-mail or telephone and was followed up a maximum of twice if an initial response was not obtained.

Medline, CINAHL and the reference lists of identified publications containing relevant outcome measures were then searched to identify further literature on the development of these outcome measures and their psychometric properties. This involved a further review of the literature including publications in peer-reviewed journals, books and compendiums. Authors of outcome measures were contacted for further details when required in some instances.

APPENDIX 2. Psychometric evaluation criteria

Clinometric property, definition and criteria used to rate the psychometric quality adopted with minor alteration from Bot et al. (13)

Content validity. The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire.

- 1) Patients were involved during item selection and/or item reduction.
- 2) Patients were consulted for reading and comprehension.

Rating:

- + patients and (investigator or expert) involved
- ± patients only
- no patient involvement
- ? no information found on content validity

Internal consistency. The extent to which items in a (sub)scale are intercorrelated; a measure of the homogeneity of a (sub)scale

- 1) Factor analysis was applied in order to provide empirical support for the dimensionality of the questionnaire.
- 2) Cronbach's alpha between 0.70 and 0.90 for every dimension/subscale

Rating:

- + adequate design & method; factor analysis; alpha 0.70–0.90
- ± doubtful method used
- inadequate internal consistency
- ? no information found on internal consistency

Construct validity. The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.

- 1) Hypotheses were formulated.
- 2) Results were acceptable in accordance with the hypotheses.
- 3) An adequate measure was used.

Rating:

- + adequate design, method, and result
- ± doubtful method used
- inadequate construct validity
- ? no information found on construct validity

Floor and ceiling effects. The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved

- 1) Descriptive statistics of the distribution of scores were presented.
- 2) 15% of respondents achieved the highest or lowest possible score.

Rating:

- + no floor/ceiling effects
- more than 15% in extremities
- ? no information found on floor and ceiling effects

Test-retest reliability. The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred

- 1) Calculation of an intraclass correlation coefficient (ICC); ICC > 0.70.
- 2) Time interval and confidence intervals were presented.

Rating:

- + adequate design, method, and ICC > 0.70
- ± doubtful method was used
- inadequate reliability
- ? no information found on test-retest reliability

Agreement. The ability to produce exactly the same scores with repeated measurements

- 1) For evaluative questionnaires reliability agreement should be assessed.

- 2) Limits of agreement, Kappa or standard error of measurement (SEM) presented.

Rating:

- + adequate design, method and result
- ± doubtful method used
- inadequate agreement
- ? no information found on agreement

Responsiveness. The ability to detect important change over time in the concept being measured

- 1) For evaluative questionnaires responsiveness should be assessed.
- 2) Hypotheses were formulated and results were in agreement.
- 3) An adequate measure was used (effect size (ES), standardized response mean (SRM), comparison with external standard).

Rating:

- + adequate design, method and result
- ± doubtful method used
- inadequate responsiveness
- ? no information found on responsiveness

Interpretability. The degree to which one can assign qualitative meaning to quantitative scores

Authors provided information on the interpretation of scores:

- 1) Presentation of means and standard deviation of scores before and after treatment.
- 2) Comparative data on the distribution of scores in relevant subgroups.
- 3) Information on the relationship of scores to well-known functional measures or clinical diagnosis.
- 4) Information on the association between changes in score and patients' global ratings of the magnitude of change they have experienced.

Rating:

- + 2 or more of the above types of information was presented
- ± doubtful method used or doubtful description
- ? no information found on interpretation

Minimal clinically important difference (MCID). The smallest difference in score in the domain of interest which patients perceive as beneficial and would mandate a change in patient's management. Information is provided about what (difference in) score would be clinically meaningful.

Rating:

- + MCID presented
- no MCID presented

Time to administer. Time needed to complete the questionnaire

Rating:

- + less than 10 min
- more than 10 min
- ? no information found on time to complete the questionnaire

Administration burden. Ease of the method used to calculate the questionnaire's score

Rating:

- + easy: summing up of the items
- ± moderate: visual analogue scale (VAS) or simple formula
- difficult: VAS in combination with formula, or complex formula
- ? no information found on rating method