

REVIEW ARTICLE

SYSTEMATIC REVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR FUNCTIONAL PERFORMANCE IN THE LOWER LIMB

Stephen Ashford, PhD, MSc, MCSP^{1,2}, Susan Brown, MCSP¹ and Lynne Turner-Stokes, DM, FRCP^{1,2}

From the ¹Regional Rehabilitation Unit, Northwick Park Hospital and ²Department of Palliative Care, Policy and Rehabilitation, School of Medicine, King's College London, London, UK

Objective: To identify psychometrically evaluated patient-reported outcome measures reflective of 'real-life' function (active and passive) for application following focal rehabilitation interventions in the lower limb after stroke or brain injury.

Data sources: A literature search conducted in MEDLINE, CINAHL, Embase, Web of Science, PubMed, National Health Service National Research Register, MRC Clinical Trials directory, Database of Abstracts of Reviews of Effects (DARE), Google Scholar and the Cochrane Database of Systematic Reviews.

Study selection: Interventional study designs using patient-reported outcome measures for outcome evaluation meeting the review objective.

Data extraction: Evaluation of the psychometric properties used the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) process, by two independent reviewers reaching consensus, with adjudication by a third reviewer.

Data synthesis: One-hundred and thirteen studies were identified following initial review of the abstracts, yielding 12 outcome measures. Eight measures were identified, which were relevant to real life functional performance. These were the Brain Injury Community Rehabilitation Outcome, Climbing Stairs Questionnaire, Human Activity Profile, Lower Extremity Functional Scale, Nottingham Extended ADL Index, Rivermead Mobility Index (RMI), Sickness Impact Profile, Stroke Impact Scale.

Conclusions: All measures addressed active function, with none evaluating passive function. The RMI met most psychometric criteria, but may have a ceiling effect for high functioning patients.

Key words: systematic review; outcome assessment; psychometrics; lower limb; function; stroke; brain injuries.

J Rehabil Med 2015; 47: 9–17

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

Accepted Jun 23, 2014; Epub ahead of print Oct 13, 2014

INTRODUCTION

The rehabilitation process for patients who have suffered a neurological event such as stroke or brain injury should be focused on the needs of the individual and be person-centred (1, 2). The views of patients within this process of rehabilitation are particularly important (3, 4). Patient-Reported Outcome Measures (PROMs) are standardised, validated questionnaires that are completed by patients to measure their perceptions of their own functional status and wellbeing (5). PROMs have been identified as a method of involving patients in evaluating their ability and outcome following intervention (5, 6). PROMs also have the benefit of evaluating what patients do in their daily lives rather than only in the clinic setting and are therefore particularly relevant to the evaluation of person-centred outcomes (5, 7).

Brain injury or stroke often significantly impact motor function in the leg (8–11). Whilst many patients will recover to some degree of useful function in their lower limb, for a minority the limb effectively becomes a passive object to be cared for, either by the individual themselves or by a carer and may interfere with function (12–14).

Interventions for the lower limb may therefore be focused on a wide range of goals. At the higher level, interventions such as functional electrical stimulation (FES) or gait re-training may target recovery of mobility, walking and in some cases running. At a lower level, interventions such as spasticity management may be directed more towards goals in 'passive' function, such as making it easier to maintain perineal hygiene (14).

Outcome measurement is required to determine the effectiveness of rehabilitation interventions (15–17). Whether applied in clinical practice or for research, measures need to be valid, reliable and responsive to clinically relevant change (18, 19). Global measures of function in daily activities, such as the Barthel Index (14, 16), provide a general assessment of independence but are often unresponsive to focal interventions. Small changes which may be extremely important to the patient and/or their carers are easily lost amongst the larger number of unchanging items (20–22).

For these reasons, a number of motor function tests of gait, walking and balance have been developed, for example the

Berg balance test, 10-m walk test and the timed-up-and-go (TUG) test (23, 24). Conducted under close observation in the clinic, these may provide a more responsive and objective measure of motor activity (23, 24). However, they do not necessarily reflect how the person actually functions in terms of mobility in their normal environment, and it is generally not practical to obtain this information through 24-h observation in the home setting (2, 7). Instead, this information on 'real life' function may be gathered through direct enquiry (self-report) from the patient and/or carer, for example using a task inventory administered by structured interview, self-completion questionnaire or PROM (25).

The aim of this systematic review was to identify valid and reliable patient (and/or carer) reported outcome measures that have been applied to assess changes following focal rehabilitation interventions in the lower limb in the context of stroke or brain injury, and are reflective of 'real-life' function (activity according to the International Classification of functioning, Disability and Health - ICF) for both active and passive function tasks (26).

Box. Active and passive function

Active function: Where a functional task is performed by active movement of the individual's affected limb e.g. to stand, walk or actively participating in lower limb dressing.

Passive function: Where a task is carried out on the affected limb by the individual using the unaffected upper limb or by a carer e.g. cleaning the perineal area or putting on trousers or positioning the limb.

METHOD

The systematic review was performed by the authors in 3 stages, as described below, according to the methodology described by Ashford et al. (2). The review methodology is published and registered on the PROSPERO registry (CRD42013005046), Centre for Reviews and Dissemination, University of York. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement provides guidance on the most appropriate methods of presenting systematic review data and these principles were used, in the presentation of results (27).

Measure selection criteria were: 1) Application of the PROM in acquired brain injury (including stroke and traumatic brain injury). The stage 1 systematic search is used to identify measures applied in this area of practice. Studies are therefore not excluded on the basis of methodological design. 2) Measures are retained at stage 3 provided they address psychometric properties included in the COSMIN criteria (18). Psychometric evaluation is undertaken on all the retained measures.

The exclusion criteria were: 1) Studies not evaluating any aspect of lower limb function. 2) Outcome measurement tools which are not PROMs. 3) Psychometric evaluation cannot be identified on the measures from the literature.

Stage 1. Data sources

In stage 1, a pool of possible measures was identified from a broad-based search of intervention studies. The following data sources were searched including Ovid MEDLINE, Ovid MEDLINE In-process, CINAHL, Embase, Web of Science, PubMed, National Health Service National Research Register, MRC Clinical Trials directory, Database of Abstracts of Reviews of Effects (DARE), Google scholar and the Cochrane Database of Systematic Reviews. The keywords used were leg, lower limb, hip, knee, ankle, foot, stroke, post stroke, cerebro-

vascular accident, CVA, cerebral haemorrhage, brain haemorrhage, haemorrhage, haematoma, hematoma, brain injuries, outcome measurement, outcome assessment, function (ing/al), activity, walking, transfers and mobility.

Stage 2. Study selection

The title was reviewed to identify potentially relevant studies. The abstract was then reviewed if the title was found to be relevant. When the abstract indicated relevance, the full text paper was retrieved and a final decision made about inclusion of the study. Initial selection was undertaken by the author and was then evaluated by a second reviewer; any areas of disagreement were discussed. Publications selected were restricted to those in the English language.

Measures were then excluded if they did not use a method of assessment reflective of 'real-life' function to measure day-to-day performance. In Stage 2, selected measures were considered to have 'real-life' relevance if they assessed day-to-day performance in the person's normal environment, as opposed to performance when observed under test conditions (such as a standardised test in a clinic setting).

Stage 3. Data extraction and synthesis

In stage 3 a second systematic search was conducted, to enable evaluation of the published evidence for the psychometric properties of the selected measures, in addition to searching the reference lists of publications selected at stage 2. The CONsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist was applied in the study evaluation for assessing the methodological quality of the measurement instruments at stage 3 (18).

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. 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. Additional search terms used were: psychometric evaluation, testing, validity, reliability, application and clinical application. Authors of outcome measures were contacted for further details when required, in addition to searching the reference lists of the psychometric publications identified.

Based on this published literature, the psychometric properties of each measure were evaluated against the following review criteria based on the COSMIN process: Practicality for use in everyday practice: time to complete, burden, readability. 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 lower limb intervention, interpretability and minimal important change (MIC).

Descriptive information was tabulated for each of the selected measures including; the items in the measure, the methods of administration and the method of scoring applied. Two reviewers independently evaluated each measure using these criteria. Findings were then compared and any discrepancies resolved through discussion. The option was available for a third reviewer to resolve any areas of disagreement following comparison, but was not used.

Procedure used to evaluate each measure

The quality criteria developed by Bot et al. (19) were then used to operationalize the evaluation of the quality of each instrument's properties, summarising each variable as adequate (+), doubtful (\pm), or poor quality (-), or as unknown (?) if insufficient information was available.

Administrative burden

Administrative burden was assessed using the same scoring method, modified as follows: Easy (+), when dichotomous items were simply summed; Moderate (\pm), when an ordinal or visual analogue scale was used to quantify individual items then summed, and Difficult (-) when a summary score was applied in combination with a formula. Timing

for completion of the measure was also rated as positive for measures completed within 10 min. Outcome measures should be both practical to use in routine practice and retain psychometric properties, thus ensuring the utility of the data produced (28, 29).

Validity

The instruments were evaluated for content and construct validity on the scale used for all psychometric properties. Content validity, evaluates that the instrument covers all the relevant concepts or domains (30). A positive rating for content validity was given when there was evidence that either patients, carers or other experts had been consulted regarding the initial selection of items (e.g. through focus groups or surveys) or had provided evaluation or feedback as part of the development. Construct validity originates from the idea that the new measure evaluates the construct it has been designed to measure (31). A positive rating for construct validity was given if there was evidence that the measure was based on hypothetical constructs, which had been tested and supported during its evaluation.

Internal consistency

Internal consistency refers to the interrelatedness of a set of items (32), and is often attributed to homogeneity of the items. A positive rating for internal consistency was given if the factor structure of the measure had been tested through factor analysis, or where ratings for Cronbach's alpha were between 0.70 and 0.95 for each dimension or subscale.

Floor and ceiling effects

Floor and ceiling effects were considered present if more than 15% of respondents achieved the highest or lowest possible score, respectively.

Reproducibility (reliability)

Reliability is concerned with detecting the amount of error occurring during application of a measurement instrument (30). Test-retest reliability was rated as positive if repeat testing of the same condition had yielded comparable results, e.g. an intraclass correlation coefficient (ICC) 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 is defined as 'the ability of an instrument to measure a meaningful or clinically important change' (33), when change occurs and record 'no-change' when the condition is stable. Responsiveness was rated as positive if the measure had demonstrated significant change in response to intervention, in the context of an appropriate study design.

Interpretability

Interpretability is the degree to which qualitative meaning can be assigned to quantitative scores (34). Positive ratings were given if at least two types of information were given to aid in understanding of the scores. Information considered included, means and standard deviations of the score totals before and after treatment, information in relation to other clinical variables, which might be expected to change, or information on the minimum change in score that might be clinically meaningful using the MIC (or MCID, Minimal Clinically Important Difference).

RESULTS

Stage 1. Data sources

A summary of the stages of review, according to the PRISMA measure selection flow diagram, is given in Fig. 1. The search

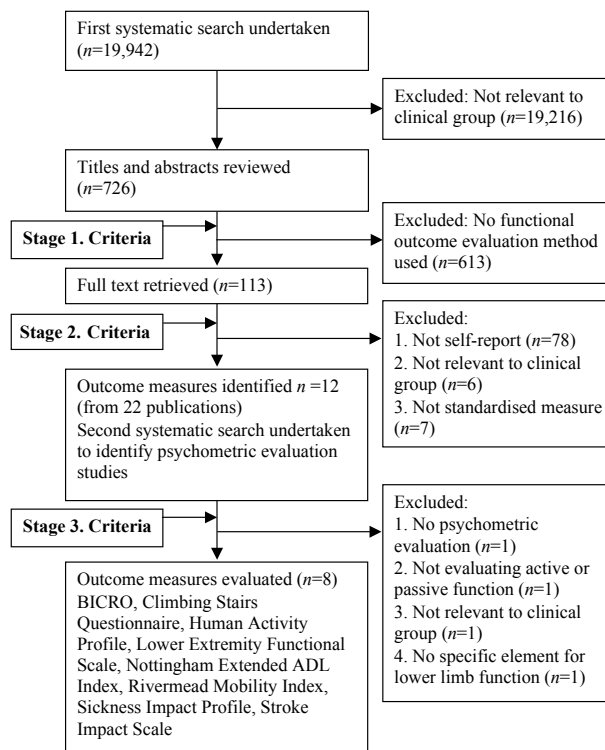


Fig. 1. PRISMA Measure selection flow diagram.

yielded 19,942 studies, including primary reports, abstracts and conference proceedings. One hundred and thirteen studies were identified following initial review of the abstracts as including measures of functional outcome following focal lower limb intervention, yielding a total of 12 outcome measures after stage 1 (see Fig. 1).

Stage 2. Study selection

Eight measures were identified, which met both stage 1 and stage 2 criteria (i.e. were relevant to real life functional performance). These were the Brain Injury Community Rehabilitation Outcome (BICRO) (35); Climbing Stairs Questionnaire (36, 37), Human Activity Profile (38–50), Lower Extremity Functional Scale (51–55), Nottingham Extended ADL Index (56–64), Rivermead Mobility Index (8, 65–78), Sickness Impact Profile (9, 79–93), Stroke Impact Scale (Table I) (3, 4, 10, 11, 94–104). The scaling methods, number of items and methods of administration for these measures are also shown in Table I.

Stage 3. Extraction and synthesis

The detailed evaluation of the properties of the selected measures is presented in Table II. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments (18), was applied in study evaluation.

Table III shows the categorisation of content of each of the 8 identified measures. The item categories form a conceptual hierarchy of increasing difficulty.

Table I. Selected measures of function

Outcome measure	Method and procedure of scoring	Context for development
Brain Injury Community Rehabilitation Outcome Scales (BICRO) (35)	Items: 39 items; 8 domains (Personal care-6 items; Mobility-6 items; Self-organisation-6 items; Contact with partner/children-2 items; Contact with parents/siblings-3 items; Socialising-6 items; Productive employment-4 items; Psychological well-being-6 items) Scoring: 0–5 scale with variation in the descriptors according to the domain Administration: Patient self-report Modifications/versions: Original version in English only	Developed with patients with varied neurological conditions (TBI, Stroke, other ABI and MS)
Climbing Stairs Questionnaire (CSQ) (36, 37)	Items: 15 items in one domain for ascending and descending stairs Scoring: Dichotomous variables Administration: Patient self-report or interview Modifications/versions: Original version in English only	Developed with patients with mixed lower limb impairment including stroke
Human Activity Profile (HAP) (38–50)	Items: 94 items in one domain of ‘activity’ Scoring: 0–3 scale Administration: Patient or carer self-report or interview Modifications/versions: Original version in English and Portuguese	Developed with patients of mixed aetiology including stroke
Lower Extremity Functional Scale (LEFS) (51–55)	Items: 20 items in one domain of lower limb activity Scoring: 0–5 scale of task difficulty Administration: Patient or carer self-report or interview Modifications/versions: Original version in English only	Developed in musculoskeletal problems, but applied in neurological conditions
Nottingham Extended ADL Index (N-ADL) (56–64)	Items: 21 items in 4 domains (mobility, kitchen, domestic, and leisure) relevant to activities of daily living (ADL) Scoring: 0–4 scale of undertaking task in past ‘few’ days Administration: Patient or carer self-report or interview Modifications/versions: Original version and extended version in English only	Developed with patients primarily with stroke (also multiple sclerosis)
Rivermead Mobility Index (RMI) (8, 65–78)	Items: 15 items in one domain of lower limb activity Scoring: yes/no (0–1 scale) Administration: 14 self-report items and one clinician observation item (independent standing) Modifications/versions: Original version in English, Italian, Dutch, German and Portuguese. A modified version of the measure is available, but is entirely scored by the clinician and is not self-reported by the patient	Developed with patients with acquired brain injury
Sickness Impact Profile (SIP) (9, 79–93)	Items: 30 items in 8 domains (Body Care and Movement, Social Interaction, Mobility, Communication, Emotional Behaviour, Household Management, Alertness Behaviour, Ambulation) Scoring: yes/no (0–1 scale) Administration: Patient self-report or interview Modifications/versions: Original 136 item version (SIP-136), which is burdensome to complete. A 68 item version (SIP-68) is also available. Original version and shortened versions in English and SIP-136 in German. An Italian version (SIP-23) for chronic pain is also available	Adapted from the SIP-68 to develop a new stroke specific version (SIP-30)
Stroke Impact Scale (SIS) (3, 4, 10, 11, 94–104)	Items: 60 items in 9 domains (physical problems, memory and thinking, mood, communication, activities, mobility, hand use, participate in activities, recovery) Scoring: 8 domains scored on a 5 point scale. 1 domain (recovery) scored with a single item on a numeric scale from 0 to 100 Administration: Patient self-report or interview Modifications/versions: SIS 2.0, SIS 3.0 and SIS-16. Translated into a number of different languages, originally developed in English	Developed with patients primarily with stroke

TBI: traumatic brain injury; ABI: acquired brain injury; MS: multiple sclerosis.

The administrative burden was adequate for 7 of the measures, but the HAP has a significant burden with time for completion (over 20 min in many studies) and contains 94 items. For the HAP construct validity was considered unclear according to the review criteria, with a resulting ‘doubtful’ rating.

DISCUSSION

In this systematic review of patient-reported measures of lower limb function, 8 measures were identified. All the measures had been applied in neurological populations, with 7 having under-

gone specific psychometric evaluation for this group (excluding the LEFS). Three of the measures specifically focused on evaluating lower limb function (CSQ, LEFS and RMI). The 5 remaining measures incorporated lower limb function in among self-report on other functional items or sub-scale dimensions. All the measures were self-reported by patients and some could also be completed by interview (see Table I). All measures identified following the search application, address active function within the activity domain of the ICF, with none evaluating passive function. Clinically-based measures of active function, such as the 10-m walk test or 6-min walk (24), are useful and valid

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

Measure	Admin Time	Content burden	Internal validity	Internal consistency	Construct validity	Floor/ceiling effect	Reliability	Agreement	Responsiveness	Interpretability	MCID
BICRO	+	+	+	+ ^b	+	+	+	+	?	?	?
CSQ	+	+	+	+	+	+	+	+	?	?	?
HAP	±	±	+	+	±	+	+	+	± ^a	± ^a	?
LEFS	+ ^a	+ ^a	+ ^a	?	+ ^a	+	+ ^a	+ ^a	± ^a	± ^a	± ^a
N-ADL	+	+	+	+	+	+	+	+	+	+	+
RMI	+	+	+	+	+	±	+	+	±	±	±
SIP	+	+	+	+	+	±	?	?	+	–	–
SIS	+	+	+	+	+	±	+	+	±	±	±

Method or result was rated as: + Adequate; ± Doubtful; – Poor; ? No data available.

^aNot adequately evaluated in a neurodisability or acquired brain injury patient group.

^bDemonstrated on lower limb sub-scales.

Admin: Administrative; BICRO: Brain Injury Community Rehabilitation Outcome Scales; CSQ: Climbing Stairs Questionnaire; HAP: Human Activity Profile; LEFS: Lower Extremity Functional Scale; MCID: Minimal Clinically Important Difference; N-ADL: Nottingham Extended ADL Index; RMI: Rivermead Mobility Index; SIP: Sickness Impact Profile; SIS: Stroke Impact Scale.

measures of performance in the clinical environment under test conditions (105). However, they evaluate performance under an artificial test situation and not functional performance carried out under normal circumstances (real-life) by the individual (105). PROM's are therefore very relevant in understanding active and passive function ability in the day-to-day environment, though clinically based measures maybe a complementary approach.

Of the 3 measures which focused specifically on lower limb function (CSQ, LEFS and RMI), the CSQ only evaluated stair climbing. The LEFS and RMI evaluated mobility in a number of different common situations and were broadly hierarchical in nature starting with easier items and progressing to more difficult items. The RMI was shown to be a unidimensional scale with a hierarchy of easy-to-hard items (65). The difficulty of items in the RMI was demonstrated to be stable when applied to different groups of patients assessed on different occasions in the same study. The RMI was developed and tested specifically with a population of patients with neurological impairment, while the LEFS was developed with patients with musculoskeletal problems, although it has been used in neurological populations. The RMI is therefore recommended as

a hierarchical measure of lower limb active function, but may not address function at a very high level and therefore have a ceiling effect. The LEFS could be an option to evaluate high level function from a patient report perspective, but would need specific testing for patients with neurological impairment and therefore can not be recommended without further evaluation, which would be valuable.

Search strategy and PROM selection

The search strategy applied has enabled selection of measurement tools applied in research studies of patients with acquired brain injury including stroke. The use of two systematic searches has 1) facilitated the initial identification of relevant measures, and 2) been followed by a detailed evaluation of their published psychometric properties.

A range of patient-reported tools evaluating active function were identified, but no tool evaluating passive function. The search strategy focused on identifying patient reported tools, evaluating active and passive function. However, on reflection, the use of passive function as a search term, while appropriate, may have (in theory) excluded possible tools where authors

Table III. Categorisation of items included in each measure

Item	BICRO	CSQ	HAP	LEFS	N-ADL	RMI	SIP	SIS
Turning in bed				✓		✓		
Lying to sitting						✓		
Sitting						✓		
Transfer (bed to chair)	✓					✓		✓
Transfer (bath or car)				✓	✓			
Sit to stand						✓		
Standing			✓			✓		✓
Walking indoors	✓		✓	✓		✓		✓
Stairs		✓	✓		✓	✓		
Walking outdoors	✓		✓	✓	✓	✓		
Running				✓	✓	✓		
Jumping/hopping				✓				
Endurance			✓	✓				✓

BICRO: Brain Injury Community Rehabilitation Outcome Scales; CSQ: Climbing Stairs Questionnaire; HAP: Human Activity Profile; LEFS: Lower Extremity Functional Scale; N-ADL: Nottingham Extended ADL Index; RMI: Rivermead Mobility Index; SIP: Sickness Impact Profile; SIS: Stroke Impact Scale.

had not used this term or concept. Other search terms such as 'personal care' could have been considered. We are, however, not aware of such tools from our research or clinical experience. The search did not include all possible search engines (e.g. PEDro), but did include indexes of publications most relevant to measurement in this area of outcome evaluation. Other measures may have been relevant to evaluate of active function. For example the ABILOCO (106), though applied via interview and not directly patient-reported, would nevertheless be relevant to assessment of this dimension. However ABILOCO was not identified from the stage 1 search or subsequent evaluation. Alongside considering an increase in the breadth of search terms applied, inclusion of other search engines could also be considered in future expanding on this review.

Evaluation of psychometric properties

The use of the COSMIN criteria in this systematic review has enabled a detailed and structured evaluation of the identified measures. The COSMIN criteria have built upon earlier work by Terwee and colleagues (107) and have incorporated both classical and item response methods into psychometric evaluation. In so doing, the COSMIN criteria make a useful contribution to the evaluation of patient-reported measures for clinically applied research and indeed practice.

In the current systematic review a wide range of psychometric evaluation of the different measures could be identified. In some cases, psychometric evaluation has been limited to one or two studies (e.g. BICRO); however for many measures evaluation was much more extensive. In some instances replication of psychometric evaluation had been possible largely confirming findings and supporting the properties being evaluated (e.g. RMI and N-ADL).

Passive vs. active function

While a number of measures addressing active function in the lower limb are available, no measure to evaluate passive function could be identified. This may be because active function improvement is more likely for greater numbers of patients in the lower limb, where smaller impairment improvement may lead to significant activity gain (e.g. small improvements in knee stability post stroke or brain injury may enable independent transfer and even walking).

Nevertheless from a clinical perspective, a group of patients can be identified who will not regain function and mobility. In a proportion of these patients, difficulty with issues such as spasticity and contracture may cause challenges in caring for the leg as seen in the upper limb (2, 7, 108). However, differences with the presentation in the upper limb are likely in the lower limb, and ease of care related to issues such as perineal hygiene may be more common in brain injury rather than stroke. For example, focal botulinum toxin intervention for lower limb spasticity (for example to hip adductors for perineal hygiene and catheter care), passive function improvement may be particularly important. In addition, passive function is

possibly a more significant issue in the lower limb for people following spinal cord injury or those with severe multiple sclerosis. Passive function may be more significant in these groups, because of the bilateral nature of impairment in some cases and the severity of related symptoms such as spasticity. A review to consider if any additional PROMS are available in these patient populations would therefore be valuable.

Summary and conclusions

The RMI is a practical and clinically applicable measure of mobility in neurologically impaired, and in particular acquired brain injury, patients. The RMI has robust psychometric measurement properties, some of which have been replicated in a number of studies. Its dimensionality and measurement scaling properties have been evaluated and demonstrated using Rasch analysis and it can therefore be applied in a hierarchical manner to rate patient ability. It does however have a ceiling effect and high function patients may require a different patient-reported tool. The LEFS could be considered in high function groups, but has not been tested in neurological populations and is therefore not recommended at present.

In this systematic review, no measure was identified that addressed passive function. While a number of patient-reported measures of active function are available, the lack of passive function measures remains a deficiency. There is therefore a need for a measure that can evaluate passive function changes following intervention and management in the lower limb.

ACKNOWLEDGEMENTS

Stephen Ashford is funded by the National Institute of Health Research (CAT-CL-03-2012-010) in the form of a Clinical Lectureship award.

This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department for Health, UK.

REFERENCES

1. Sheean G, Lannin NA, Turner-Stokes L, Rawicki B, Snow BJ. Botulinum toxin assessment, intervention and after-care for upper limb hypertonicity in adults: international consensus statement. *Euro J Neurol* 2010; 17: 74–93.
2. Ashford S, Slade M, Malaprada F, Turner-Stokes L. Evaluation of functional outcome measures for the hemiparetic upper limb: A systematic review. *J Rehabil Med* 2008; 40: 787–795.
3. Carod-Artal F, Coral L, Trizotto D, Moreira C. Self- and proxy-report agreement on the stroke impact scale. *Stroke* 2009; 40: 3308–3314.
4. Duncan P, Reker D, Kwon S, Lai S, Studenski S, Perera S, et al. Measuring stroke impact with the Stroke Impact Scale: telephone versus mail administration in veterans with stroke. *Med Care* 2005; 43: 507–515.
5. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr A. Routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010; 340: 464–467.
6. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported

- outcome measures in routine practice: a structured review. *Eval Clin Pract* 2005; 12: 559–568.
7. Ashford S, Slade M, Turner-Stokes L. Conceptualisation and development of the Arm Activity measure (ArmA) for assessment of activity in the hemiparetic arm. *Disabil Rehabil* 2013; 35: 1513–1518.
 8. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Studies* 1991; 13: 50–54.
 9. Van de Port IG, Ketelaar M, Schepers VPM, Gam, Lindeman E. Monitoring the functional health status of stroke patients: the value of the Stroke-Adapted Sickness Impact Profile-30. *Disabil Rehabil* 2004; 26: 635–640.
 10. Lai S, Perera S, Duncan PW, Bode R. Physical and social functioning after stroke: comparison of the Stroke Impact Scale and Short Form-36. *Stroke* 2003; 34: 488–493.
 11. Moriello C, Byrne K, Cieza A, Nash C, Stolee P, Mayo N. Mapping the Stroke Impact Scale (SIS-16) to the International Classification of Functioning, Disability and Health. *J Rehabil Med* 2008; 40: 102–106.
 12. Hendricks H, van Limbeek J, Geurts AC, Zwarts MJ. Motor recovery after stroke: A systematic review of the literature. *Arch Phys Med Rehabil* 2002; 83: 1629–1637.
 13. Duncan PW, Min Lai S, Keighley J. Defining post-stroke recovery: implications for design and interpretation of drug trials. *Neuropharm* 2000; 39: 835–841.
 14. Yagura H, Miyai I, Seike Y, Suzuki T, Yanagihara T. Benefit of inpatient multidisciplinary rehabilitation up to 1 year after stroke. *Arch Phys Med Rehabil* 2003; 84: 1687–1691.
 15. Wade D. Goal setting in rehabilitation: an overview of what, why and how. *Clin Rehabil* 2009; 23: 291–295.
 16. Wade DT, Collin C. The Barthel ADL index: a standard measure of physical disability? *Int Disabil Studies* 1988; 10: 64–67.
 17. Wade DT, Langton-Hewer R, Wood V, Skilbeck CE, Ismail IM. The hemiplegic arm after stroke: measurement and recovery. *J Neurol Neurosurg Psych* 1983; 46: 521–524.
 18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford P, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19: 539–549.
 19. 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.
 20. Ashford S, Turner-Stokes L. Goal attainment for spasticity management using botulinum toxin. *Physio Res Int* 2006; 11: 24–34.
 21. Royal College of Physicians. Guidelines for the use of botulinum toxin (BTX) in the management of spasticity in adults. 2002; London, UK: Royal College of Physicians, Clinical Effectiveness and Evaluation Unit; 2002.
 22. Royal College of Physicians, British Society of Rehabilitation Medicine, Chartered Society of Physiotherapy, Association of Chartered Physiotherapists Interested in Neurology. Spasticity in adults: management using botulinum toxin - National Guidelines. 2009; London: Royal College of Physicians, Clinical Effectiveness and Evaluation Unit; 2009.
 23. Wade D. Evaluating outcome in stroke rehabilitation. *Scand J Rehabil Med* 1992; 26: 97–104.
 24. Wade DT. Measurement in neurological rehabilitation. Oxford: Oxford University Press; 1992.
 25. Jones L. *Jebson test of hand function (British Version)*. London: National Hospital for Neurology and Neurosurgery; 1990.
 26. WHO. *International Classification of Functioning, Disability and Health*. Geneva: World Health Organisation; 2002.
 27. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med J* 2009; 6: e1000097.
 28. Slade M. Routine outcome assessment in mental health services. *Psych Med* 2002; 32: 1339–1343.
 29. Slade M, Thornicroft G, Glover G. The feasibility of routine outcome measures in mental health. *Soc Psych Psych Epidemiol* 1999; 34: 243–249.
 30. Streiner D, Norman G. *Health measurement scales; a practical guide to their development and use*. 3rd ed: Oxford University Press; 2003.
 31. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psych Bulletin* 1955; 52: 281–302.
 32. Schmitt N. Uses and abuses of coefficient alpha. *Psych Assess* 1996; 8: 350–353.
 33. Liang MH. Longitudinal construct validity: Establishment of clinical meaning in patient evaluation instruments. *Med Care* 2000; 38 Suppl 2: S84–S90.
 34. Nunnally J, Bernstein IH. *Psychometric Theory*. 3rd ed. New York, USA: McGraw-Hill; 1994.
 35. Powell JH, Beckers K, Greenwood RJ. Measuring progress and outcome in community rehabilitation after brain injury with a new assessment instrument – The BICRO-39 Scales. *Arch Phys Med Rehabil* 1998; 79: 1213–1225.
 36. de Laat FA, Rommers GM, Geertzen JH, Roorda LD. Construct validity and test-retest reliability of the climbing stairs questionnaire in lower-limb amputees. *Arch Phys Med Rehabil* 2010; 91: 1396–1401.
 37. Roorda LD, Roebroek ME, van Tilburg T, Lankhorst GJ, Bouter LM, and the Measuring Mobility Study Group. Measuring activity limitations in climbing stairs: Development of a hierarchical scale for patients with lower extremity disorders living at home. *Arch Phys Med Rehabil* 2004; 85: 967–971.
 38. Bennell KL, Hinman RS, Crossley KM, Metcalf BR, Buchbinder R, Green S, et al. Is the Human Activity Profile a useful measure in people with knee osteoarthritis? *J Rehabil Res Develop* 2004; 41: 621–630.
 39. Bilek LD, Venema DM, Camp KL, Lyden ER, Meza JL. Evaluation of the human activity profile for use with persons with arthritis. *Arthritis Rheum* 2005; 53: 756–763.
 40. Daly AE, McGinley J, Martin C, Bialocerkowski AE. The clinimetric properties of the Radboud Skills Questionnaire and the human activity profile for adult upper limb complex regional pain syndrome. *Anaesthesia and Intensive Care, Melbourne, VIC Australia*; 2012, p. 535–536.
 41. Davidson M, de Morton N. A systematic review of the human activity profile. *Clin Rehabil* 2007; 21: 151–162.
 42. Herzberg PY, Heussner P, Mumm FHA, Horak M, Hilgendorf I, von Harsdorf S, et al. Validation of the human activity profile questionnaire in patients after allogeneic hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant* 2010; 16: 1707–1717.
 43. Hetzler RK, Stickley CD, Shikuma CM, Chow D, Takahashi CE, Kimura IF, et al. Validation of the human activity profile in HIV-positive individuals on HAART. *HIV Clin Trials* 2009; 10: 254–260.
 44. Johansen KL, Painter P, Kent-Braun JA, Ng AV, Carey S, Da Silva M, et al. Validation of questionnaires to estimate physical activity and functioning in end-stage renal disease. *Kidney Int* 2001; 59: 1121–1127.
 45. Nield M, Hoo GS, Roper J, Santiago S, Dracup K. Usefulness of the human activity profile, a functional performance measure, in people with chronic obstructive pulmonary disease. *J Cardio Rehabil* 2005; 25: 115–1121.
 46. Souza AC, Magalhaes LDC, Teixeira-Salmela LF. Cross-cultural adaptation and analysis of the psychometric properties in the Brazilian version of the Human Activity Profile. *Cadernos de Saude Publica* 2006; 22: 2623–2636.
 47. Teixeira-Salmela LF, Devaraj R, Olney SJ. Validation of the human activity profile in stroke: A comparison of observed, proxy and self-reported scores. *Disabil Rehabil* 2007; 29: 1518–1524.

48. Webster S, Khan A, Nitz JC. A brief questionnaire is able to measure population physical activity levels accurately: A comparative validation study. *J Clin Gerontol Geriatrics* 2011; 2: 83–87.
49. Wellard S. Validation of physical activity measurement for people on dialysis treatment. *EDTNA-ERCA J* 2003; 29: 140–142.
50. Williams K, Frei A, Vetsch A, Dobbels F, Puhan MA, Rüdell K. Patient-reported physical activity questionnaires: A systematic review of content and format. *Health Qual Life Outcomes* 2012; 10: 1–18.
51. Binkley J, Stratford P, Lott S, Riddle D, The North American Orthopaedic Rehabilitation Research Network. The Lower Extremity Functional Scale: Scale development, measurement properties, and clinical application. *Phys Ther* 1999; 79: 4371–4383.
52. Stratford PW, Kennedy DM, Maly MR, MacIntyre NJ. Quantifying Self-Report Measures' Overestimation of Mobility Scores Post arthroplasty'. *Phys Ther* 2010; 90: 1288–1296.
53. Stratford PW, Kennedy DM, Riddle DL. New study design evaluated the validity of measures to assess change after hip or knee arthroplasty. *J Clin Epidemiol* 2009; 62: 347–352.
54. Watson CJ, Propps M, Ratner J, Zeigler DL, Horton P, Smith SS. Reliability and responsiveness of the lower extremity functional scale and the anterior knee pain scale in patients with anterior knee pain. *J Orthop Sports Phys Ther* 2005; 35: 136–146.
55. Yeung TM, Wessel J, Stratford P, Macdermid J. Reliability, validity, and responsiveness of the lower extremity functional scale for inpatients of an orthopaedic rehabilitation ward. *J Orthop Sports Phys Ther* 2009; 39: 468–477.
56. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997; 50: 79–93.
57. Wu CY, Chuang LI, Lin KC, Lee SD, Hong W. Responsiveness, minimal detectable change, and minimal clinically important difference of the Nottingham extended activities of daily living scale in patients with improved performance after stroke rehabilitation. *Arch Phys Med Rehabil* 2011; 92: 1281–1286.
58. das Nair R, Moreton BJ, Lincoln NB. Rasch analysis of the Nottingham Extended Activities of Daily Living Scale. *J Rehabil Med* 2011; 43: 944–950.
59. Garrod R, Bestall JC, Paul EA, Wedzicha JA, Jones PW. Development and validation of a standardized measure of activity of daily living in patients with severe COPD: the London Chest Activity of Daily Living scale (LCADL). *Resp Med* 2000; 94: 589–596.
60. Green S, Forester A, Young J. A test-retest reliability study of the Barthel Index, the Rivermead Mobility Index, the Nottingham extended Activities of Daily Living Scale and the Frenchay Activities Index in stroke patients. *Disabil Rehabil* 2001; 23: 670–676.
61. Harwood RH, Ebrahim S. The validity, reliability and responsiveness of the Nottingham Extended Activities of Daily Living scale in patients undergoing total hip replacement. *Disabil Rehabil* 2002; 24: 371–377.
62. Lincoln NB, Gladman RF. Extended Activities of Daily Living scale: a further validation. *Disabil Rehabil* 1992; 14: 41–43.
63. Nicholl CR, Lincoln NB, Playford ED. The reliability and validity of the Nottingham Extended Activities of Daily Living Scale in patients with multiple sclerosis. *Mult Scler* 2002; 8: 372–376.
64. Nouri FM, Lincoln NB. An extended activities of daily living scale for stroke patients. *Clin Rehabil* 1987; 1: 301–305.
65. Antonucci G, Aprile T, Paolucci S. Rasch analysis of the Rivermead Mobility Index: a study using mobility measures of first-stroke inpatients. *Arch Phys Med Rehabil* 2002; 83: 1442–1449.
66. Franchignoni F, Brunelli S, Orlandini D, Ferriero G, Traballese M. Is the Rivermead Mobility Index a suitable outcome measure in lower limb amputees? A psychometric validation study. *J Rehabil Med* 2003; 35: 141–144.
67. Franchignoni F, Tesio L, Benevolo E, Ottonello M. Psychometric properties of the Rivermead Mobility Index in Italian stroke rehabilitation inpatients. *Clin Rehabil* 2003; 17: 273–282.
68. Hsieh CL, Hsueh IP, Mao HF. Validity and responsiveness of the Rivermead mobility index in stroke patients. *Scand J Rehabil Med* 2000; 32: 140–142.
69. Johnson L, Selfe J. Measurement of mobility following stroke: a comparison of the Modified Rivermead Mobility Index and the Motor Assessment Scale. *Physiotherapy* 2004; 90: 132–138.
70. Lennon S, Johnson L. The modified Rivermead Mobility Index: validity and reliability. *Disabil Rehabil* 2000; 22: 833–839.
71. Pavan K, da Cruz LCA, Nunes MF, Menezes LG, Marangoni BEM. Cross-cultural adaptation and validation of the Rivermead Mobility Index in stroke patients within the Brazilian cultural and language context. *Arquivos de Neuro-Psiquiatria* 2010; 68: 52–55.
72. Roorda LD, Green J, De Kluis KRA, Molenaar IW, Bagley P, Smith J, et al. Excellent cross-cultural validity, intra-test reliability and construct validity of the Dutch Rivermead Mobility Index in patients after stroke undergoing rehabilitation. *J Rehabil Med* 2008; 40: 727–732.
73. Roorda LD, Green JR, Houwink A, Bagley PJ, Smith J, Molenaar IW, et al. Item hierarchy-based analysis of the Rivermead Mobility Index resulted in improved interpretation and enabled faster scoring in patients undergoing rehabilitation after stroke. *Arch Phys Med Rehabil* 2012; 93: 1091–1096.
74. Roorda LD, Green JR, Houwink A, Bagley PJ, Smith J, Molenaar IW, et al. The Rivermead Mobility Index allows valid comparisons between subgroups of patients undergoing rehabilitation after stroke who differ with respect to age, sex, or side of lesion. *Arch Phys Med Rehabil* 2012; 93: 1086–1090.
75. Ryall NH, Eyres SB, Neumann VC, Bhakta BB, Tennant A. Is the Rivermead Mobility Index appropriate to measure mobility in lower limb amputees? *Disabil Rehabil* 2003; 25: 143–153.
76. Schindl MR, Forstner C, Kern H, Zipko HT, Rupp M, Zifko UA. Evaluation of a German version of the Rivermead Mobility Index (RMI) in acute and chronic stroke patients. *Euro J Neurol* 2000; 7: 523–528.
77. Sommerfeld DK, Johansson H, Jonsson A-L, Murray V, Wessari T, Holmqvist LW, et al. Rivermead mobility index can be used to predict length of stay for elderly persons, 5 days after stroke onset. *J Geriatr Phys Ther* 2011; 34: 64–71.
78. Walsh JM, Barrett A, Murray D, Ryan J, Moroney J, Shannon M. The Modified Rivermead Mobility Index: reliability and convergent validity in a mixed neurological population. *Disabil Rehabil* 2010; 32: 1133–1139.
79. Bergner M. The Sickness Impact Profile (SIP). Health Services Research and Development Center, The Johns Hopkins University, 624 North Broadway, Baltimore, MD 21205-1901; 1977, p. 1–15.
80. Bowers S, Cannizzaro K, Gregus J, Scott Q, Eason J. Outcomes in cardiopulmonary physical therapy: sickness impact profile. *Cardio Phys Ther J* 2009; 20: 19–24.
81. de Bruin AF, de Witte LP, Stevens F, Diederiks JPM. Sickness Impact Profile: the state of the art of a generic functional status measure. *Soc Sci Med* 1992; 35: 1003–1014.
82. Deyo RA. Pitfalls in measuring the health status of Mexican Americans: comparative validity of the English and Spanish Sickness Impact Profile: a clinical study of low back pain. *Am J Pub Health* 1984; 74: 569–573.
83. Gilson BS, Gilson JS, Bergner M, Bobbitt RA, Kressel S, Pollard WE, et al. The Sickness Impact Profile Development of an Outcome Measure of Health Care. *Am J Public Health* 1975; 65: 1304–1310.
84. Hutter BO, Wurtemberger G. Reliability and validity of the German version of the Sickness Impact Profile in patients with chronic obstructive pulmonary disease. *Psych Health* 1997; 12: 149–159.
85. Jansen L, Steultjens MP, Holtslag HR, Kwakkel G, Dekker J. Psychometric properties of questionnaires evaluating health-related quality of life and functional status in polytrauma patients with lower extremity injury. *J Trauma Manag Outcomes* 2010; 4: 7.
86. Lipsett PA, Swoboda SM, Campbell KA, Cornwell E, III, Dorman T, Pronovost PJ. Sickness Impact Profile Score versus a modified short-form survey for functional outcome assessment: acceptabil-

- ity, reliability, and validity in critically ill patients with prolonged intensive care unit stays. *J Trauma* 2000; 49: 737–743.
87. Lyden PD, Hantson L. Assessment scales for the evaluation of stroke patients. *J Stroke Cerebrovasc Dis* 1998; 7: 113–127.
 88. Monticone M, Baiardi P, Nava T, Rocca B, Foti C. The Italian version of the Sickness Impact Profile-Roland Scale for chronic pain: cross-cultural adaptation, reliability, validity and sensitivity to change. *Disabil Rehabil* 2011; 33: 1299–1305.
 89. Post MWM, Gerritsen J, Diederiks JPM, De Witte LP. Measuring health status of people who are wheelchair-dependent: validity of the Sickness Impact Profile 68 and the Nottingham Health Profile. *Disabil Rehabil* 2001; 23: 245–253.
 90. Streppel KRM, van Harten WH, Warmerdam CGM. Short version of the Sickness Impact Profile for evaluating rehabilitation programs. *J Rehabil Sci* 1996; 9: 66–71.
 91. van Baalen B, Odding E, van Woensel MPC, van Kessel MA, Roebroek ME, Stam HJ. Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. *Clin Rehabil* 2006; 20: 686–700.
 92. van Straten A, de Haan RJ, Limburg M, Schuling J, Bossuyt PM, Gam. A stroke-adapted 30-item version of the Sickness Impact Profile to assess quality of life (SA-SIP30). *Stroke* 1997; 28: 2155–2161.
 93. Wilson JR, Hashimoto RE, Dettori JR, Fehlings MG. Spinal cord injury and quality of life: a systematic review of outcome measures. *Evid Based Spine Care J* 2011; 2: 37–44.
 94. Carod-Artal F, Coral L, Trizotto D, Moreira C. The Stroke Impact Scale 3.0: evaluation of acceptability, reliability, and validity of the Brazilian version. *Stroke* 2008; 39: 2477–2484.
 95. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil* 2003; 84: 950–963.
 96. Duncan PW, Lai SM, Tyler D, Perera S, Reker DM, Studenski S. Evaluation of proxy responses to the Stroke Impact Scale. *Stroke* 2002; 33: 2593–2599.
 97. Duncan PW, Reker DM, Horner RD, Samsa GP, Hoenig H, LaClair BJ, et al. Performance of a mail-administered version of a stroke-specific outcome measure, the Stroke Impact Scale. *Clin Rehabil* 2002; 16: 493–505.
 98. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale Version 2.0: evaluation of reliability, validity, and sensitivity to change. *Stroke* 1999; 30: 2131–2140.
 99. Duncan PW, Wallace D, Studenski S, Lai SM, Johnson D. Conceptualization of a new stroke-specific outcome measure: the stroke impact scale. *Top Stroke Rehabil* 2001; 8: 19–33.
 100. Fulk GD, Ludwig M, Dunning K, Golden S, Boyne P, West T. How much change in the stroke impact scale-16 is important to people who have experienced a stroke? *Top Stroke Rehabil* 2010; 17: 477–483.
 101. Igl, Leenes K, Sellmeijer D, Zuidgeest A, Kwakkel G. [Reliability and criterion validity of the Dutch Stroke Impact Scale 2.0 in patients with stroke.] *Nederlands Tijdschrift Voor Fysiotherapie* 2008; 118: 12–18 (in Dutch).
 102. Lin K, Fu T, Wu C, Wang Y, Liu J, Hsieh C, et al. Minimal detectable change and clinically important difference of the Stroke Impact Scale in stroke patients. *Neurorehabil Neural Repair* 2010; 24: 486–492.
 103. Savini S, Alvaro R, Vellone E. [Stroke impact scale: a specific instrument to assess the quality of life in stroke survivors.] *Int Nursing Perspectives* 2010; 10: 13–20 (in Italian).
 104. Vellone E, Venturini G, Caramia M, Alvaro R. Preliminary results from the reliability and validity of the stroke impact scale in Italy. Fourth European Nursing Congress. *J Clin Nurs* 2010; 19: 85.
 105. Dobkin BH. Short-distance walking speed and timed walking distance: redundant measures for clinical trials? *Neurology* 2006; 66: 584–586.
 106. Caty DC, Arnould C, Stoquart GG, Thonnard J, Lejeune TM. ABILOCO: A Rasch-Built 13-Item Questionnaire to Assess Locomotion Ability in Stroke Patients. *Arch Phys Med Rehabil* 2008; 89: 284–290.
 107. 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.
 108. Ashford S, Slade M, Turner-Stokes L. Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm. *Clin Rehabil* 2013; 27: 728–740.