

REVIEW ARTICLE

SYSTEMATIC REVIEW OF OUTCOME MEASURES USED IN THE EVALUATION OF ROBOT-ASSISTED UPPER LIMB EXERCISE IN STROKE

Manoj Sivan, MRCS^{1,2}, Rory J. O'Connor, MD, FRCP^{1,2,3}, Sophie Makower, BA(Hons), Grad. Dip. Phys^{1,3}, Martin Levesley, BEng., PhD⁴ and Bipinchandra Bhakta, MD, FRCP^{1,2}

From the ¹Academic Department of Rehabilitation Medicine, University of Leeds, ²National Demonstration Centre for Rehabilitation Medicine, Leeds Teaching Hospitals NHS Trust, ³Community Rehabilitation Unit, NHS Leeds Community Healthcare and ⁴Department of Mechanical Engineering, University of Leeds, Leeds, UK

Objective: To classify and evaluate outcome measures currently used in robot-assisted exercise trials (RAET) in stroke, and to determine selection criteria for outcome measures in future trials.

Methods: Outcome measures used in RAET were identified from MEDLINE, EMBASE, CINAHL, PubMed and PsychINFO databases. The scale items were categorized into International Classification of Functioning Disability and Health (ICF) domains. The psychometric properties of scale were rated using a standardized *pro forma*.

Results: Thirty outcome measures were identified from 28 published RAET. Commonly used ICF body function scales were: Fugl-Meyer (FM) (24 studies), Modified Ashworth Scale (13 studies), Medical Research Council (11 studies), Kinematic measures (8 studies) and Motor Status Score (6 studies); ICF activity scale was Functional Independence Measure (FIMTM) (9 studies); ICF participation, personal and environmental factors scales were rarely used. Standardized rating identified that FM, kinematic measures, Action Research Arm Test, Wolf Motor Function Test, FIMTM, and ABILHAND have adequate measurement properties for use in RAET.

Conclusion: Some of the currently used outcome measures seem appropriate for RAET. The use of the ICF framework enables selection of an appropriate combination of outcome measures depending on patient characteristics, such as severity of weakness and chronicity of stroke impairments.

Key words: rehabilitation; robotic devices; scales; ICF; psychometrics.

J Rehabil Med 2011; 43: 181–189

Correspondence address: Manoj Sivan, NIHR Academic Clinical Fellow, Academic Department of Rehabilitation Medicine, University of Leeds, D Floor, Martin Wing, Leeds General Infirmary, Leeds LS1 3EX, UK. E-mail: M.Sivan@leeds.ac.uk

Submitted August 15, 2010; accepted December 17, 2010

INTRODUCTION

Stroke is the commonest cause of severe physical disability. The annual incidence of new strokes in Europe is between 200 and 300 per 100,000 population (1). The recovery of upper limb function is generally slower and less complete than the return of mobility. Up to 85% of survivors experience

some degree of paresis of the upper limb at the onset (2) and 25% report difficulty using the affected limb 5 years post-stroke (3). This is partly due to the complexity of movement required for upper limb function (4, 5). Increasing the amount and frequency of physical therapy can improve some aspects of motor recovery; however, physical therapy resource is limited. As a consequence, intervention is often inadequate in terms of the intensity and frequency required to relearn motor skills. Robot-assisted exercise can supplement conventional physical therapy.

A number of robotic devices have been developed to assist upper limb training and rehabilitation (6). Some have been evaluated in robot-assisted exercise trials (RAET) with patients after stroke. A meta-analysis of randomized controlled trials showed a significant improvement in upper limb motor function and no significant change in activities of daily living (ADL) function with upper limb robotics (7). This apparent lack of effect in real-life activities may relate to the effect size of the intervention and the outcome measures used in the studies. Larger trials are needed to confirm or refute the findings of smaller scale robot studies done so far. In designing larger and more expensive studies it is vital that appropriate outcome measures are used.

There is lack of consensus on the outcome measures that should be used in RAET. Most published clinical trials have used comparable outcome measures. The rationale for this is to allow comparison across studies and pooling of data for systematic review purposes. There has been emphasis on measuring change at the impairment level (either through kinematic assessment or impairment-based rating scales) rather than change in functional activities of daily living. Some recently published clinical studies have incorporated outcomes that reflect day-to-day activities. There is limited literature describing how to select outcome measures based on the nature of the intervention and the patient's clinical features.

The aims of this systematic review are: (i) to identify the outcome measures that have been used in RAET, classify them using the International Classification of Functioning, Disability and Health (ICF) (8) and report on their psychometric properties; and (ii) to determine the factors that should be considered when selecting outcome measures in future trials. The domains described in the ICF conceptual framework of health condition, body functions (and structures), activities and participation,

and personal and environmental factors, although related, have no necessary causality between them, making measurement of all the domains necessary (9).

METHODS

The systematic review was undertaken in 3 stages:

Stage 1. Identify clinical trials involving robot-assisted arm therapy in patients after stroke and describe the outcome measures used

A search of MEDLINE, EMBASE, CINALH, PubMed and PsychINFO databases was performed to identify relevant RAET. The keywords used were: stroke, upper limb, arm, rehabilitation, motor, recovery, robot, computer, training, therapy, physiotherapy, function, study and trial. From the initial search, all abstracts were reviewed. The inclusion criteria for this review were: (i) study involving participants with diagnosis of a stroke; (ii) upper limb exercise assisted using a robot device. For this review, a robotic device was defined as any technology that has the ability to assist arm movement for therapeutic exercises; (iii) at least one outcome measure used in the study.

Studies of robot devices involving only healthy volunteers were excluded.

This stage was undertaken by authors MS, SM, ML and BB and their lists were cross-referenced with each other's to ensure all relevant RAET had been identified.

Stage 2. Classify the content of the outcome measures used in these RAET according to the three main ICF domains

The content of each scale (identified in stage 1) was classified in terms of the ICF categories (8):

- Body functions and structures: functions refer to physiological functions of body systems including psychological. Structures are anatomical parts or regions of body and their components. Impairments are problems in body function or structure.
- Activity: activity refers to execution of a task by an individual. Limitations of a task are defined as difficulties an individual might experience in completing a given activity.
- Participation: involvement of an individual in a life situation. Restrictions to participation describe difficulties experienced by the individual in a life situation or role.
- Contextual factors: these include personal and environmental factors that influence the relationship between the different components.

This stage was undertaken by authors MS, SM, ROC and BB.

Stage 3. Describe the measurement properties of the identified outcome measures in patients after stroke

A search of the same databases (used in Stage 1) was undertaken to identify RAET involving stroke participants and describing the properties of the identified scales from stage 1. The keywords used were: the name of the outcome measure, stroke, validity, reliability, questions, items, consistency, minimal clinically important difference (MCID), responsiveness, floor effect, ceiling effect and agreement. A measure-

Table I. Definition and standards for the evaluation criteria

Criterion	Definition	Standard
Reliability	Reproducibility is the extent to which the same results are obtained on repeated administrations of the same questionnaire by same person (test re-test) or different people (inter-rater). Internal consistency assesses the homogeneity of the scale items (23).	<i>Reproducibility</i> (test-retest or inter-observer) – <i>Intraclass correlation coefficient (ICC) or kappa value</i> – excellent or high >0.75, moderate 0.4–0.74 and poor <0.40 (23, 24). <i>Internal consistency</i> – <i>Cronbach's alpha</i> excellent >0.8, adequate 0.70–0.79 and low <0.70 (25, 26).
Validity	The extent to which the scale measures what it intends to measure. <i>Content validity</i> is the extent to which the measure is representative of the conceptual domain. <i>Criterion validity</i> (concurrent, convergent, predictive) is the degree to which the measure correlates with a gold standard. For most of the functional scales, there is no gold standard and hence construct validity is used. Construct validity is determined by examining the hypothetical relationship between the measure and other similar measures (23).	<i>Correlation coefficient value (r)</i> – excellent >0.60, adequate 0.3–0.6 and poor <0.3 (25). <i>ROC analysis</i> – <i>area under curve (AUC)</i> excellent >0.9, adequate 0.7–0.9 and poor <0.7 (27).
Responsiveness	The ability of the instrument to detect changes that have occurred accurately over time (28). <i>Minimal clinically important difference (MCID)</i> – the smallest difference in score in the domain of interest that patients perceive as beneficial or that would be clinically meaningful. <i>Floor and ceiling effects</i> – the extent to which scores cluster at the bottom or top, respectively, of the scale range.	<i>Change in score</i> – the <i>effect size</i> is calculated by the observed change in score divided by the standard deviation of baseline score. Large >0.8, moderate 0.5–0.8 and small <0.5 (29, 30). Other methods: Standardized response mean (SRM) ROC analysis – area under curve (AUC) Statistical significance <i>p</i> -value Correlation values of observed change compared with change in other scales. <i>MCID</i> – described as a score value. <i>Floor and ceiling effects</i> – expressed as <i>percentage of the number of scores</i> clustered at bottom/top. Excellent 0%, Adequate <20%, poor >20% (25).
Acceptability	<i>Respondent burden</i> – is the length and content acceptable to the intended participants (participants with disability)? <i>Administrative burden</i> – how easy is the tool to administer, score and interpret? Cost implications?	<i>Respondent burden</i> – Excellent: brief (<15 min) and acceptable, Adequate: either longer or some problems of acceptability. Poor: both lengthy and problems of acceptability (25). <i>Administrative burden</i> – Excellent: scoring by hand, easy to interpret, Adequate: computer scoring, obscure interpretation, Poor: costly and complex scoring/interpretation (25).

ment profile for each scale was constructed based on the evidence for the different properties mentioned above. This stage was undertaken by author MS and ROC. The criteria for defining the measurement properties are summarized in Table I.

Participants were considered as being in the *sub-acute* stage of recovery if within 6 months of their stroke and in the *chronic* stage if more than 6 months since their stroke.

Table II lists the abbreviations for the outcome measures used in this article.

RESULTS

Stage 1

A total of 28 RAET involving 16 robot devices met the inclusion criteria for this review. Table III summarizes the outcome measures used in these studies. The commonly used scores were Fugl-Meyer (FM; 24 studies), Modified Ashworth Scale (MAS; 13 studies), Medical Research Council power grading scale (MRC; 11 studies), Functional Independence Measure (FIM; 9 studies) kinematic measurements (8 studies) and Motor Status Score (MSS; 6 studies).

Stage 2

The individual items within each scale were classified into the 3 ICF domains (Table SI, available from: URL://http:jrm.medical-journals.se/article/abstract/10.2340.16501977-0674). Based on the overall item content, each scale was categorized as belonging to one of the ICF domains. Fig. 1 summarizes the classification of all the scales into impairment, activity, participation and contextual factor categories. Fifteen scales were identified as body function based outcome measures, 10 activities based, 2 participation based and 3 identified as contextual factor outcome measures.

Stage 3

Studies investigating the measurement properties of these scales in stroke participants were identified, and the evidence

Table II. Outcome measure abbreviations

Abbreviation	Outcome measure
AMAT	Arm Motor Ability Test
ARAT	Action Research Arm Test
AS	Ashworth Scale
BBT	Box and Block Test
BI	Barthel Index
CAHAI	Chedoke Arm and Hand Activity Inventory
CMSA	Chedoke-McMaster Stroke Assessment
EMG	Electromyogram
EQ-5D	EuroQol Quality of Life Scale
FAT	Frenchay Arm Test
FIM	Functional Independence Measure
FIM motor	Functional Independence Measure motor subscale
FM	Fugl-Meyer scale
FM motor	Fugl-Meyer motor subscale
fMRI	Functional Magnetic Resonance Imaging
MAS	Modified Ashworth Scale
MFT	Manual Function Test
Motor AS	Motor Assessment Scale
MRC	Medical Research Council
MSS	Motor Status Score
NHPG	Nine-Hole Peg Test
NSA	Nottingham Sensory Assessment
RLAFT	Rancho Los Amigos Functional Test
RMA	Rivermead Motor Assessment
ROM	Range of Motion/Movement
SCT	Star Cancellation Test
SIS	Stroke Impact Scale
TUG	Timed Up and Go
TCT	Trunk Control Test
UMAQS	University of Maryland Arm Questionnaire for Stroke
VAS	Visual Analogue Scale
WMFT	Wolf Motor Function Test

on different properties consolidated. The properties for each of the scales are described in this section. Tables IV and V summarize the properties of all the scales.

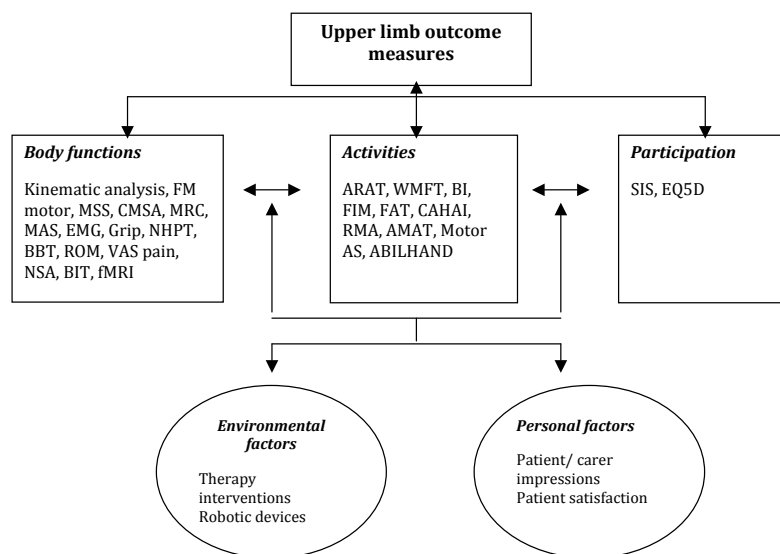


Fig. 1. International Classification of Functioning, Disability and Health (ICF) categorization of scales used in robot studies. (For abbreviations, see Table II).

Table III. Scales used in robot studies (in the order of number of studies and then year of publication)

Robot device	Reference	n	Type of patients	FM motor	MSS	MAS	MRC	FIM	Kinematic assessments	Robot measures	Others
MIT MANUS	Lo et al., 2010 (31)	127	Chronic	+		+					WMFT, SIS
	Posteraro et al., 2009 (32)	20	Chronic		+	+					ROM, VAS pain
	Krebs et al., 2008 (33)	47	Chronic	+							
	Rabadi et al., 2008 (34)	30	Subacute	+	+		+	+			FM – pain
	Fasoli et al., 2003 (35)	20	Subacute	+	+	+	+				
	Volpe et al., 2000 (36)	56	Subacute	+	+		+	+			
	Aisen et al., 1997 (37)	20	Subacute	+			+	+			
Bi Manu	Hesse et al., 2008 (38)	54	Subacute	+		+	+				BBT
track Arm	Hesse et al., 2005 (39)	44	Subacute	+		+	+				
trainer	Hesse et al., 2003 (13)	12	Chronic			+					RMA, Patient impressions
MIME	Lum et al., 2006 (40)	23	Subacute	+	+			+			
	Lum et al., 2002 (10)	27	Chronic	+				+	Reach extent		BI, Muscle power MVC
	Burgar et al., 2000 (41)	21	Chronic	+							
NeReBot	Masiero et al., 2007 (42)	35	Subacute	+		+	+	+			TCT
	Rosati et al., 2007 (43)	24	Subacute	+	+		+	+			
BATRAC	Luft et al., 2004 (11)	21	Chronic	+							WMFT, UMAQS, fMRI, Strength (Dynamometer)
	Whitall et al., 2000 (17)	14	Chronic	+							WMFT, UMAQS, Strength (Dynamometer)
GENTLE	Coote et al., 2008 (22)	20	Chronic	+		+					ROM, Motor AS, SCT, VAS (pain), NSA
ReoGo	Bovolenta et al., 2009 (15)	14	Chronic	+		+	+	+			BBT, FAT, ABILHAND, TUG, EQ-5D
BdF	Squeri et al., 2009 (44)	4	Chronic	+						Force, Time, Balance error	
Reo Therapy	Treger et al., 2008 (45)	10	Subacute	+							MFT, Patient satisfaction
HWARD	Takahashi et al., 2008 (20)	13	Chronic	+		+					ARAT, NHPT, BBT, SIS
BFIAMT	Chang et al., 2007 (12)	20	Chronic	+		+			Peak speed, Time, Jerk +	Push–pull strength	FAT, Grip strength
Therajoy/ drive	Johnson et al., 2007 (46)	16									EMG – muscle strength
REHAROB	Fazekas et al., 2007 (47)	30	Mixed	+		+		+			RMA, ROM, VAS (patient acceptance), VAS (pain)
REHA-SLIDE	Hesse et al., 2007 (48)	2	Subacute	+		+	+				
ARM Guide	Kahn et al., 2006 (49)	19	Chronic						Range, Smoothness, Path length	Stiffness, Range, Velocity	CMSA
In Motion S-E Robot	Daly et al., 2005 (50)	12	Chronic	+							AMAT

+: used in trial; BFIAMT: Bilateral Force Induced Isokinetic Arm Movement Trainer; MIME: Mirror Image Movement Enabler; BdF: Braccio di Ferro; ARM: Assisted Rehabilitation and Measurement Guide; HWARD: Hand Wrist Assistive Rehabilitation Device; NeReBot: Neuro Rehabilitation Robot; S-E: Shoulder–Elbow; MIT MANUS: Massachusetts Institute of Technology; BATRAC: Bilateral Arm Training with Rhythmic Auditory Cueing; FM: Fugl-Meyer scale; MSS: Motor Status Score; MAS: Modified Ashworth Scale; MRC: Medical Research Council; FIM: Functional Independence Measure. For other abbreviations, see Table II.

DISCUSSION

The impacts of stroke on the domains of the ICF are not always directly related to each other. The severity of impairment does not necessarily determine the limitation in activities and participation due to the varied interplay between these domains and the influence of contextual factors. Such differences may also be seen in relation to the effects of any intervention (e.g. changes at the impairment level do not necessarily translate into the other domains, e.g. participation). The selection of

outcome measures is therefore crucial in the design of RAET and should aim to capture the changes in all the aspects of the health condition (in this case, stroke). Using the ICF to describe scale content should enable researchers to compare different scales and select the most appropriate ones for their trial. Individual scale measurement properties may allow targeting of the most appropriate scale to the study participants.

Published reports of RAET indicate that criteria to categorize the study participants are helpful. Some studies have considered FM scores of less than 20 or 25 to indicate severe impair-

Table IV. Psychometric properties of impairment scales

Characteristics	FM motor	MSS	CMSA	MAS	MRC	Kinematics	Grip strength	NHPT	BBT
Time taken (min)	20	n/a	60	Varies	Varies	Varies	<1	2	1
Number of items	33	29	6	1	1	Varies	1	1	1
Type	3 point	6 point	7 point	6 point	6 point	Varies	Timed	Timed	Timed
Score range	0–66	0–82	6–42	0–5	0–5	Varies	Varies	Varies	Varies
Test-retest reliability	+++	+++	n/a	++	n/a	+++	+++	n/a	n/a
Inter-rater reliability	+++	+++	+++	++	+++	n/a	+++	+++	+++
Construct validity	+++	+++	+++	+	n/a	++	n/a	+++	+++
Responsiveness	++	n/a	n/a	n/a	n/a	+++	n/a	n/a	n/a
MCID	7	n/a	n/a	n/a	n/a	n/a	2.9 kg	32 s	6/min
Floor effect	Adeq	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Ceiling effect	Adeq	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Burden	Adeq	Adeq	Poor	Adeq	Adeq	Adeq	Nil	Nil	Nil
References	(23, 51–64)	(65)	(57, 66, 67)	(68–78)	(79–81)	(82–85)	(86–88)	(86–89)	(88, 90)

Scoring criteria as defined in Table I. For abbreviations, see Table II.

+++; high/excellent; ++; moderate; +; low/poor; n/a: no available evidence yet; adeq: adequate (acceptable) floor/ceiling effect/burden; poor: poor (unacceptable) floor/ceiling effect/burden; nil: minimal/no burden.

ment and more than 20 or 25 as moderate impairment (10, 11). Also, time since stroke has been used to indicate speed of recovery during rehabilitation. In acquired brain injury studies, participants are considered as being in the *sub-acute* stage of recovery if within 6 months since the event and *chronic* stage if more than 6 months (10, 12–14). Based on severity and time since stroke, we can therefore conceptualize participants in 4 categories (Fig. 2).

Severely impaired participants, particularly in the sub-acute stage of recovery, may need outcome measures with minimal floor effects to be able to discriminate between the score of individual participants. Kinematic measurement and the FM or MSS would be appropriate body function outcome measures for this group. Kinematic measurement, although more responsive than most of clinical scales, can be time-consuming and requires special equipment. The FM scale has been used in all RAET involving these participants and has been shown to

be responsive (Table III). Among activity measures, the FIM motor subscale is suitable for use in this category and has been used in RAET with such participants (Table III).

The use of FIM motor subscale and Barthel Index (BI) is limited by responsiveness in RAET involving patients who have severe impairments that persist beyond 6 months. Two studies involving chronic stroke survivors did not identify change in FIM scale, although changes in the FM scale were reported (10, 15). This finding may also relate to the BI and FIM being measures of global physical ability that may be affected by many factors other than arm impairments. For this category, Action Research Arm Test (ARAT) may be limited by its floor effects, the Chedoke Arm and Hand Activity Inventory (CAHAI) scale would seem to be an appropriate activity scale, but has not been used in RAET. One non-robot rehabilitation study has showed the CAHAI to be more sensitive to change than ARAT in chronic participants (16).

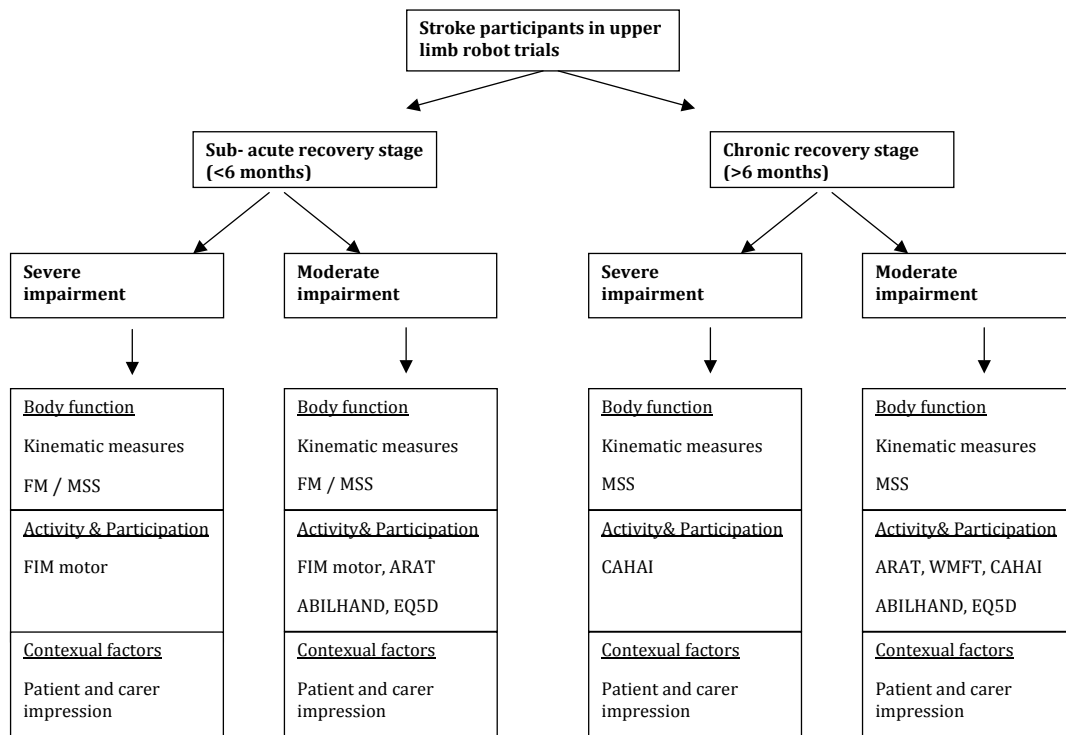
Table V. Psychometric properties of activity and participation scales

Characteristics	BI	FIM motor	ARAT	WMFT	CAHAI	AMAT	RMA arm	FAT	Motor AS	ABILH-AND	SIS part.	EQ5D
Time taken (min)	10–15	20	10	10–12	25	45	20	3	20–30	n/a	n/a	2–3
Number of items	10	13	19	15	13	17	15	5	9	23	8	5
Type	2–4 point ^a	7 point	4 point	6 point	7 point	6 point	2 point	2 point	7 point	3 point	5 point	3 point
Score range	0–100	13–91	0–57	0–75	13–91	0–85	0–15	0–5	0–54	logit	0–100	0–1
Test-retest reliability	+++	+++	+++	+++	n/a	+++	+++	+++	+++	n/a	+++	+++
Inter-rater reliability	+++	+++	+++	+++	+++	+++	n/a	+++	+++	+++	n/a	n/a
Construct validity	+++	+++	++	+++	+++	n/a	++	n/a	+++	++	+++	++
Responsiveness	++	++	++	++	+++	n/a	n/a	n/a	n/a	n/a	n/a	n/a
MCID	16	11	6	12	6.3	n/a	3	n/a	n/a	n/a	n/a	n/a
Floor effect	Poor	n/a	Poor	Poor	n/a	n/a	n/a	Poor	n/a	n/a	n/a	Adeq
Ceiling effect	Poor	Adeq	Poor	Poor	n/a	n/a	n/a	Poor	n/a	n/a	n/a	Adeq
Burden	Nil	Adeq	Adeq	Adeq	Adeq	Poor	Adeq	Nil	Adeq	Adeq	Nil	Adeq
References	(55, 91–99)	(98, 100, S1–S5)	(62–64, 85, 90, S6–S8)	(62, S9–S14)	(16, S15)	(59, S16, S17)	(S18–S24)	(87, S25, S26)	(58, S27, S28)	(S29, S30)	(S31–S34)	(S35–S37)

^aTwo, 3 or 4 response options per item.

Scoring criteria as defined in Table I. For abbreviations, see Table II.

+++; high/excellent; ++; moderate; +; low/poor; n/o: no available evidence yet; adeq: adequate (acceptable) floor/ceiling effect/burden; poor: poor (unacceptable) floor/ceiling effect/burden; nil: minimal/no burden.



Severe impairment – Fugl-Meyer < 25 / Unable to actively move robot arm without assistance

Moderate impairment – Fugl-Meyer > 25 / Able to actively move robot arm without assistance

Fig. 2. Proposed algorithm for selection of scales based on patient characteristics and International Classification of Functioning, Disability and Health (ICF) domains.

The third category is those with moderate impairments in the sub-acute stage of recovery. These studies require outcome measures with minimal ceiling effects to be able to discriminate between score changes observed in individual participants. Kinematic measurement and the FM or MSS are suitable body function outcome measures for this group. The ARAT, Wolf Motor Function Test (WMFT) (although both will be limited by their ceiling effects) and ABILHAND would be suitable activity scales. EuroQol Quality of Life Scale (EQ5D) will be a suitable participation outcome.

The final category of moderately impaired participants in the chronic stage will need outcome measures with high responsiveness to be able to capture the intervention effect. The FM scale seems to be less useful in such participants if the changes in impairment are small. Although the MSS was developed to be more responsive in therapeutic studies targeting arm impairments, we could not find evidence for this in RAET. The ARAT and WMFT may be suitable activity scales for these participants, as has been shown in RAET involving such participants (11, 17). ABILHAND might be useful in this group, as it captures real-life functional benefit, but the only RAET that included ABILHAND as an outcome measure did not observe significant change in scores (15).

Achievement of personalized goals can be used to capture change following intervention at an individual level (e.g. Goal Attainment Scale, Canadian Occupational Performance

Measure) (18). These are suitable for monitoring individual persons, but are not appropriate for group analysis (19), which limits their usefulness in evaluating robot-assisted exercise in the context of a randomized control trial.

Participation should be considered as an important part of the evaluation with RAET. There are few scales that measure participation (e.g. Stroke Impact Scale (SIS) and SF36) that can be considered for inclusion. One robot trial included the SIS as an outcome, but recorded only the hand motor subscale of the SIS, which is an activity measure (20). The participation subscale of the SIS should be considered in future trials.

Economic evaluation should be considered as an important part of any large-scale clinical investigation of robot-assisted exercise. Therefore, when designing the trial it is important to include the use of a health utility measure (e.g. EQ-5D) and health resource utilization questionnaires within the context of RAET. One robot trial involving chronic participants did not observe any statistically significant improvement in EQ-5D scores, although statistically significant improvements were found in FM and FIM scores (15). It is possible that the EQ-5D may have lower responsiveness than FM and FIM. The responsiveness of EQ-5D in stroke participants is currently unknown (Table VI). Other measures that capture dependency and provide an estimate of care costs saving through reduction in dependency for physical assistance following robot-assisted exercise treatments, such as the Northwick Park Dependency Scale (NPDS), should be considered (21).

Personal and environmental factors have a huge influence on any intervention in rehabilitation. Patient and carer perceptions of robot-assisted exercise are important outcome measures to allow design iteration and gain information about satisfaction with the delivery of robot-assisted therapy, which could relate to the look and feel of the system (13).

The other important factor to be considered when selecting outcome measures is whether the intervention is aimed at proximal or distal upper limb muscle groups. The 3 hand function tests (Grip strength, Nine-Hole Peg Test (NHPT) and Box and Block Test (BBT)) are quick to administer and may be suitable for studies where the intervention is directed at distal limb and hand function. The hand-based robot Hand Wrist Assistive Rehabilitation Device (HWARD) trial showed greater increase in BBT and NHPT scores when compared with proximal shoulder and elbow FM and ARAT (20).

The use of functional magnetic resonance imaging (fMRI) is at an experimental stage (11) and needs to be evaluated further in robot trials. It may provide interesting insights into the recovery process, but has limitations in terms of cost and feasibility. One aspect of recovery that is neglected in RAET is measuring changes in perceptual (sensory) function arising as a result of robot-assisted arm exercise. Perceptual function is a vital part of normal movement, and evidence suggests that recovery of functional motor ability is dependent on intact sensation, spatial awareness and attention. Interactive robot-assisted exercise may improve perceptual deficits or potentially confound the benefits that might be identified in RAET. Only one RAET used a sensory assessment tool as one of the outcome measures (22). The extent of sensory impairment did not seem to influence the overall benefit from robot-assisted therapy in this study. Changes in the perceptual function need to be further researched in RAET.

Fig. 2 describes our proposed algorithm for selecting outcome measures based on the type of participants recruited to RAET. Apart from the factors mentioned in the algorithm, other factors that should be considered in selecting outcome measures are the type of assistance that the robot provides (proximal, distal or both) and available resources (e.g. to undertake measurements in terms of research staff cost and participant time). We propose that at least 4 suitable outcome measures covering the different domains of ICF could be considered as essential to understand the effects of robot-assisted exercise on arm impairments in people with stroke.

The main limitation of this review is that we have analysed in detail only those outcome measures that have been used in RAET so far. This does not necessarily mean that outcome measures not used in the trials are not suitable for use in future trials. However, this review provides a system for the selection of outcome measures, which should enable researchers to apply these criteria to the outcome they wish to explore in future trials. The future of robot-assisted rehabilitation after stroke is influenced by accurate analysis and interpretation of the observed effects. This could be accomplished by using the most appropriate outcome measures.

In conclusion, we are proposing an approach to assist researchers in selecting outcome measures in the design of future

clinical trials of robot-assisted rehabilitation. We feel that a basket of outcome measures covering all domains of ICF is crucial, as it is important to measure change in each domain. The selection of outcome measures should also be based on the focus of the intervention, severity of arm impairments, time since stroke, and psychometric properties of the scales.

REFERENCES

1. Stroke prevention and educational awareness diffusion (SPREAD). The Italian guidelines for stroke prevention and treatment. Milano: Ed. Hyperphar Group; 2003.
2. Skilbeck CE, Wade DT, Hewer RL, Wood VA. Recovery after stroke. *J Neurol Neurosurg Psychiatry* 1983; 46: 5–8.
3. Geddes JM, Fear J, Tennant A, Pickering A, Hillman M, Chamberlain MA. Prevalence of self reported stroke in a population in northern England. *J Epidemiol Community Health* 1996; 50: 140–143.
4. Gresham GE, Duncan PW, Stason WB, Adams HPJ, Adelman AM, Alexander DN, et al. Post-stroke rehabilitation. Clinical Practice Guideline No. 16. Rockville: US Dept Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, AHCPR;1995. Report No. 95-0662.
5. Kwakkel GKB, Wagenaar RC. Therapy impact on functional recovery in stroke rehabilitation. *Physiotherapy* 1999; 85: 377–391.
6. Riener R, Nef T, Colombo G. Robot-aided neurorehabilitation of the upper extremities. *Med Biol Eng Comput* 2005; 43: 2–10.
7. Kwakkel G, Kollen BJ, Krebs HI. Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. *Neurorehabil Neural Repair* 2008; 22: 111–121.
8. World Health Organization (WHO). The International Classification of Functioning, Disability and Health – ICF2001. Geneva: WHO.
9. Muller U, Roeder C, Dubs L, Duetz MS, Greenough CG. Condition-specific outcome measures for low back pain. Part II: scale construction. *Eur Spine J* 2004; 13: 314–324.
10. Lum PS, Burgar CG, Shor PC, Majumdar M, Van der Loos M. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Arch Phys Med Rehabil* 2002; 83: 952–959.
11. Luft AR, McCombe-Waller S, Whittall J, Forrester LW, Macko R, Sorkin JD, et al. Repetitive bilateral arm training and motor cortex activation in chronic stroke: a randomized controlled trial. *JAMA* 2004; 292: 1853–1861.
12. Chang JJ, Tung WL, Wu WL, Huang MH, Su FC. Effects of robot-aided bilateral force-induced isokinetic arm training combined with conventional rehabilitation on arm motor function in patients with chronic stroke. *Arch Phys Med Rehabil* 2007; 88: 1332–1338.
13. Hesse S, Schulte-Tiggas G, Konrad M, Bardeleben A, Werner C. Robot-assisted arm trainer for the passive and active practice of bilateral forearm and wrist movements in hemiparetic subjects. *Arch Phys Med Rehabil* 2003; 84: 915–920.
14. Sivan M, Neumann V, Kent R, Stroud A, Bhakta BB. Pharmacotherapy for treatment of attention deficits after non-progressive acquired brain injury. A systematic review. *Clin Rehabil* 2010; 24: 110–121.
15. Bovolenta F, Goldoni M, Clerici P, Agosti M, Franceschini M. Robot therapy for functional recovery of the upper limbs: a pilot study on patients after stroke. *J Rehabil Med* 2009; 41: 971–975.
16. Barreca SR, Stratford PW, Masters LM, Lambert CL, Griffiths J. Comparing 2 versions of the Chedoke Arm and Hand Activity Inventory with the Action Research Arm Test. *Phys Ther* 2006; 86: 245–253.

17. Whittall J, McCombe Waller S, Silver KH, Macko RF. Repetitive bilateral arm training with rhythmic auditory cueing improves motor function in chronic hemiparetic stroke. *Stroke* 2000; 31: 2390–2395.
18. Turner-Stokes L. Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clin Rehabil* 2009; 23: 362–370.
19. Tennant A. Goal attainment scaling: current methodological challenges. *Disabil Rehabil* 2007; 29: 1583–1588.
20. Takahashi CD, Der-Yeghiaian L, Le V, Motiwala RR, Cramer SC. Robot-based hand motor therapy after stroke. *Brain* 2008; 131: 425–437.
21. Turner-Stokes L, Shaw A, Law J, Rose H. Development and initial validation of the Northwick Park Therapy Dependency Assessment. *Clin Rehabil* 2009; 23: 922–937.
22. Coote S, Murphy B, Harwin W, Stokes E. The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke. *Clin Rehabil* 2008; 22: 395–405.
23. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. *Disabil Rehabil* 2005; 27: 191–207.
24. Fleiss J. *Statistical methods for rates and proportions*. New York: Wiley; 1981.
25. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000; 81: S15–S20.
26. Nunnally J. *Psychometric theory*. New York McGraw-Hill; 1967.
27. McDowell I, Newell C. *A guide to rating scales and questionnaires*. New York: Oxford University Press; 1996.
28. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997; 50: 239–246.
29. Cohen J. *Statistical power analysis for the behavioural sciences*. New York: Academic Press; 1977.
30. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: S178–S189.
31. Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med* 2010; 362: 1772–1783.
32. Posteraro F, Mazzoleni S, Aliboni S, Cesqui B, Battaglia A, Dario P, et al. Robot-mediated therapy for paretic upper limb of chronic patients following neurological injury. *J Rehabil Med* 2009; 41: 976–980.
33. Krebs HI, Mernoff S, Fasoli SE, Hughes R, Stein J, Hogan N. A comparison of functional and impairment-based robotic training in severe to moderate chronic stroke: a pilot study. *NeuroRehabilitation* 2008; 23: 81–87.
34. Rabadi M, Galgano M, Lynch D, Akerman M, Lesser M, Volpe B. A pilot study of activity-based therapy in the arm motor recovery post stroke: a randomized controlled trial. *Clin Rehabil* 2008; 22: 1071–1082.
35. Fasoli SE, Krebs HI, Stein J, Frontera WR, Hogan N. Effects of robotic therapy on motor impairment and recovery in chronic stroke. *Arch Phys Med Rehabil* 2003; 84: 477–482.
36. Volpe BT, Krebs HI, Hogan N, Edelstein OL, Diels C, Aisen M. A novel approach to stroke rehabilitation: robot-aided sensorimotor stimulation. *Neurology* 2000; 54: 1938–1944.
37. Aisen ML, Krebs HI, Hogan N, McDowell F, Volpe BT. The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke. *Arch Neurol* 1997; 54: 443–446.
38. Hesse S, Werner C, Pohl M, Mehrholz J, Puzich U, Krebs HI. Mechanical arm trainer for the treatment of the severely affected arm after a stroke: a single-blinded randomized trial in two centers. *Am J Phys Med Rehabil* 2008; 87: 779–788.
39. Hesse S, Werner C, Pohl M, Rueckriem S, Mehrholz J, Lingnau ML. Computerized arm training improves the motor control of the severely affected arm after stroke: a single-blinded randomized trial in two centers. *Stroke* 2005; 36: 1960–1966.
40. Lum PS, Burgar CG, Van der Loos M, Shor PC, Majmundar M, Yap R. MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: a follow-up study. *J Rehabil Res Dev* 2006; 43: 631–642.
41. Burgar CG, Lum PS, Shor PC, Machiel Van der Loos HF. Development of robots for rehabilitation therapy: the Palo Alto VA/Stanford experience. *J Rehabil Res Dev* 2000; 37: 663–673.
42. Masiero S, Celia A, Rosati G, Armani M. Robotic-assisted rehabilitation of the upper limb after acute stroke. *Arch Phys Med Rehabil* 2007; 88: 142–149.
43. Rosati G, Gallina P, Masiero S. Design, implementation and clinical tests of a wire-based robot for neurorehabilitation. *IEEE Trans Neural Syst Rehabil Eng* 2007; 15: 560–569.
44. Squeri V, Casadio M, Vergaro E, Giannoni P, Morasso P, Sanguineti V. Bilateral robot therapy based on haptics and reinforcement learning: Feasibility study of a new concept for treatment of patients after stroke. *J Rehabil Med* 2009; 41: 961–965.
45. Treger I, Faran S, Ring H. Robot-assisted therapy for neuromuscular training of sub-acute stroke patients. A feasibility study. *Eur J Phys Rehabil Med* 2008; 44: 431–435.
46. Johnson MJ, Feng X, Johnson LM, Winters JM. Potential of a suite of robot/computer-assisted motivating systems for personalized, home-based, stroke rehabilitation. *J Neuroeng Rehabil* 2007; 4: 6.
47. Fazekas G, Horvath M, Troznai T, Toth A. Robot-mediated upper limb physiotherapy for patients with spastic hemiparesis: a preliminary study. *J Rehabil Med* 2007; 39: 580–582.
48. Hesse S, Schmidt H, Werner C, Rybski C, Puzich U, Bardeleben A. A new mechanical arm trainer to intensify the upper limb rehabilitation of severely affected patients after stroke: design, concept and first case series. *Eura Medicophys* 2007; 43: 463–468.
49. Kahn LE, Zygmant ML, Rymer WZ, Reinkensmeyer DJ. Robot-assisted reaching exercise promotes arm movement recovery in chronic hemiparetic stroke: a randomized controlled pilot study. *J Neuroeng Rehabil* 2006; 3: 12.
50. Daly JJ, Hogan N, Perepezko EM, Krebs HI, Rogers JM, Goyal KS, et al. Response to upper-limb robotics and functional neuromuscular stimulation following stroke. *J Rehabil Res Dev* 2005; 42: 723–736.
51. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. I. A method for evaluation of physical performance. *Scand J Rehabil Med* 1975; 7: 13–31.
52. Fugl-Meyer AR, Jaasko L, Norlin V. The post-stroke hemiplegic patient. II. Incidence, mortality, and vocational return in Goteborg, Sweden with a review of the literature. *Scand J Rehabil Med* 1975; 7: 73–83.
53. Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. *Stroke* 2000; 31: 1429–1438.
54. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys Ther* 1993; 73: 447–454.
55. Wood-Dauphinee SL, Williams JI, Shapiro SH. Examining outcome measures in a clinical study of stroke. *Stroke* 1990; 21: 731–739.
56. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair* 2002; 16: 232–240.
57. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hullenaar S, et al. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke* 1993; 24: 58–63.
58. Malouin F, Pichard L, Bonneau C, Durand A, Corriveau D. Evaluating motor recovery early after stroke: comparison of the Fugl-Meyer Assessment and the Motor Assessment Scale. *Arch Phys Med Rehabil* 1994; 75: 1206–1212.
59. Chae J, Labatia I, Yang G. Upper limb motor function in hemiparesis: concurrent validity of the Arm Motor Ability test. *Am J Phys Med Rehabil* 2003; 82: 1–8.
60. Dettmann MA, Linder MT, Sepic SB. Relationships among walking performance, postural stability, and functional assessments of the hemiplegic patient. *Am J Phys Med* 1987; 66: 77–90.
61. Shelton FD, Volpe BT, Reding M. Motor impairment as a predictor

- of functional recovery and guide to rehabilitation treatment after stroke. *Neurorehabil Neural Repair* 2001; 15: 229–237.
62. Lin JH, Hsu MJ, Sheu CF, Wu TS, Lin RT, Chen CH, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Phys Ther* 2009; 89: 840–850.
 63. Hsueh IP, Hsieh CL. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. *Clin Rehabil* 2002; 16: 617–624.
 64. van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehabil Med* 2001; 33: 110–113.
 65. Ferraro M, Demaio JH, Krol J, Trudell C, Rannekleiv K, Edelstein L, et al. Assessing the motor status score: a scale for the evaluation of upper limb motor outcomes in patients after stroke. *Neurorehabil Neural Repair* 2002; 16: 283–289.
 66. Poole JL, Whitney SL. Assessments of motor function post stroke: a review. *Phys Occupat Ther Geriatr* 2001; 19: 1–22.
 67. Valach L, Signer S, Hartmeier A, Hofer K, Steck GC. Chedoke-McMaster stroke assessment and modified Barthel Index self-assessment in patients with vascular brain damage. *Int J Rehabil Res* 2003; 26: 93–99.
 68. Bohannon RW, Smith MB. Interrater reliability of a Modified Ashworth Scale of muscle spasticity. *Phys Ther* 1987; 67: 206–207.
 69. Allison SC, Abraham LD, Petersen CL. Reliability of the Modified Ashworth Scale in the assessment of plantarflexor muscle spasticity in patients with traumatic brain injury. *Int J Rehabil Res* 1996; 19: 67–78.
 70. Bodin PG, Moris ME. Inter rater reliability of the Modified Ashworth Scale for wrist flexors spasticity following stroke. In: *Proceedings of the 11th Congress of the World Confederation for Physical Therapy*; 1991.
 71. Sloan RL, Sinclair E, Thompson J, Taylor S, Pentland B. Interrater reliability of the modified Ashworth Scale for spasticity in hemiplegic patients. *Int J Rehabil Res* 1992; 15: 158–161.
 72. Hass BM, Bergstrom E, Jamous A, Bennie A. The inter-rater reliability of the original and of the modified Ashworth Scale for the assessment of spasticity in patients with spinal cord injury. *Spinal Cord* 1996; 34: 560–564.
 73. Pandyan AD, Johnson GR, Price CI, Curless RH, Barnes MP, Rodgers H. A review of the properties and limitations of the Ashworth and Modified Ashworth Scales as measures of spasticity. *Clin Rehabil* 1999; 13: 373–383.
 74. Pandyan AD, Price CI, Barnes MP, Johnson GR. A biomechanical investigation into the validity of the Modified Ashworth Scale as a measure of elbow spasticity. *Clin Rehabil* 2003; 17: 290–293.
 75. Skold C, Harms-Ringdahl K, Hultling C, Levi R, Seiger A. Simultaneous Ashworth measurements and electromyographic recordings in tetraplegic patients. *Arch Phys Med Rehabil* 1998; 79: 959–965.
 76. A M O Bakheit AMO, Maynard VA, Curnow J, Hudson N, Kodapala S. The relation between Ashworth scale scores and the excitability of the α motor neurones in patients with post-stroke muscle spasticity. *J Neurol Neurosurg Psychiatry* 2003; 74: 646–648.
 77. Leonard CT, Stephens JU, Stroppel SL. Assessing the spastic condition of individuals with upper motoneuron involvement: validity of the myotonometer. *Arch Phys Med Rehabil* 2001; 82: 1416–1420.
 78. McCrea PH, Eng JJ, Hodgson AJ. Linear spring-damper model of the hypertonic elbow: reliability and validity. *J Neurosci Methods* 2003; 128: 121–128.
 79. Medical Research Council. *Aids to the investigation of peripheral nerve injuries*. London: Medical Research Council, Committee NI; 1943.
 80. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Reliability of measurements of muscle tone and muscle power in stroke patients. *Age Ageing* 2000; 29: 223–228.
 81. MacAvoy MC, Green DP. Critical reappraisal of Medical Research Council muscle testing for elbow flexion. *J Hand Surg Am* 2007; 32: 149–153.
 82. Caimmi M, Carda S, Giovanzana C, Maini ES, Sabatini AM, Smania N, et al. Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients. *Neurorehabil Neural Repair* 2008; 22: 31–39.
 83. Chang JJ, Wu TI, Wu WL, Su FC. Kinematic measure for spastic reaching in children with cerebral palsy. *Clin Biomech (Bristol, Avon)* 2005; 20: 381–388.
 84. Bosecker C, Dipietro L, Volpe B, Krebs HI. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabil Neural Repair* 2010; 24: 62–69.
 85. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: properties of the action research arm test. *Arch Phys Med Rehabil* 2006; 87: 1605–1610.
 86. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. *J Hand Surg Am* 1984; 9: 222–226.
 87. Heller A, Wade DT, Wood VA, Sunderland A, Hewer RL, Ward E. Arm function after stroke: measurement and recovery over the first three months. *J Neurol Neurosurg Psychiatry* 1987; 50: 714–719.
 88. Chen HM, Chen CC, Hsueh IP, Huang SL, Hsieh CL. Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. *Neurorehabil Neural Repair* 2009; 23: 435–440.
 89. Kellor M, Frost J, Silberberg N, Iversen I, Cummings R. Hand strength and dexterity. *Am J Occup Ther* 1971; 25: 77–83.
 90. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil* 2005; 19: 404–411.
 91. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965; 14: 61–65.
 92. Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. *J Formos Med Assoc* 2001; 100: 526–532.
 93. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil* 2000; 81: S21–S29.
 94. Wolfe CD, Taub NA, Woodrow EJ, Burney PG. Assessment of scales of disability and handicap for stroke patients. *Stroke* 1991; 22: 1242–1244.
 95. Cup EHC, Scholte OP, Reimer WJM, Thijssen MCE, Van Kuyk-Minis MAH. Reliability and validity of the Canadian Occupational Performance Measure in stroke patients. *Clin Rehabil* 2003; 17: 402–409.
 96. Duncan PW, Samsa GP, Weinberger M, Goldstein LB, Bonito A, Witter DM, et al. Health status of individuals with mild stroke. *Stroke* 1997; 28: 740–745.
 97. Dromerick AW, Edwards DF, Diringer MN. Sensitivity to changes in disability after stroke: a comparison of four scales useful in clinical trials. *J Rehabil Res Dev* 2003; 40: 1–8.
 98. Wallace D, Duncan PW, Lai SM. Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *J Clin Epidemiol* 2002; 55: 922–928.
 99. Van der Putten JMF, Hobart JC, Freeman JA, Thompson AJ. Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *J Neurol Neurosurg Psychiatry* 1999; 66: 480–484.
 100. Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Tashman JS. A uniform national data system for medical rehabilitation. In: Fuhrer MJ, editor. *Rehabilitation outcomes: analysis and measurement*. Baltimore: Brookes Publishing Co.; 1987: p. 137–147.