# Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm

Stephen Ashford[1, 2] PhD MSc MCSP
Clinical Specialist and Research Physiotherapist and Honorary Research Fellow

Lynne Turner-Stokes[1, 2] DM FRCP
Herbert Dunhill Chair of Rehabilitation and Director Regional Rehabilitation Unit

Richard Siegert[2] PhD
Professor of Psychology and Rehabilitation

Mike Slade[3] PhD
Professor of Psychology

[1] Regional Rehabilitation Unit, Northwick Park Hospital
[2] King's College London, School of Medicine, Department of Palliative Care, Policy and Rehabilitation
[3] King's College London, Health Service and Population Research Department, Institute of Psychiatry
[4] AUT University, Department of Psychology, Faculty of Health and Environmental Sciences, New Zealand

Address for correspondence: Stephen Ashford, Regional Rehabilitation Unit, Northwick Park Hospital, Watford Road, Harrow, Middlesex HA1 3UJ, UK
Tel: +44 (0) 20 8869 2800
Fax: +44 (0) 20 8869 2803
Email: Stephen.Ashford@nhs.net

***Keywords:*** Psychometrics, Activities, Arm.

**Abstract**

**Objective**

To evaluate the psychometric properties of the Arm Activity Measure (ArmA), a patient-reported measure of active and passive function in the paretic upper limb.

**Design**

Psychometric evaluation study.

**Setting**

Two specialist rehabilitation and spasticity management services.

**Method**

Patients (n=92) with upper limb paresis were recruited from two specialist neurorehabilitation centres. Mean age 44.5(SD16.7). Diagnostic distribution: stroke 48(52%); other brain injury 28(31%); or other neurological condition 16(17%). Evaluation of convergent and divergent validity; unidimensionality, scaling, reliability (internal consistency and test-retest); responsiveness to change and feasibility of the ArmA were undertaken.

**Results**

Expected convergent and divergent relationships were seen with the Leeds Adult Spasticity Impact Scale and the Disabilities of Arm Shoulder and Hand (DASH) (rho 0.5–0.63). Principal components analysis (PCA) confirmed that active and passive function formed two separate constructs in each sub-scale. Mokken analysis corroborated the findings of the PCA and demonstrated scaling using the monotone homogeneity model (Item H >0.5 for all items). Cronbach's alpha was 0.85 and 0.96 respectively for the passive and active function sub-scales. Item level test-retest agreement ranged from 92-97.5% (quadratic-weighted Kappa 0.71-0.94). In the subgroup treated for spasticity with botulinum toxin (n=58), the ArmA passive function scale identified a significant difference between responder and non-responder groups

(Mann Whitney U=0.85, p<0.01) Respondents reported the ArmA to be relevant (77%), easy to use (90%) and timely to complete (83% under 10 minutes).

**Conclusion**

The ArmA is a valid and reliable tool feasible for use in the evaluation of upper limb function in the context of treatment for spasticity.

*Word count 246*

## Introduction

The complex nature of upper limb function presents a challenge for rehabilitation following neurological injury. Some patients, with relatively mild injury, have potential to recover useful function such as the ability to use the hand to hold and manipulate objects (active function). Others with more severe injury will continue to have a non-functional upper limb, and may require assistance from another person (or their own non-affected arm) to care for the affected limb (passive function) [1]. There is a requirement for psychometrically robust instruments capable of reflecting clinically important change in clinical and research practice. Outcome measures for treatments (such as botulinum toxin for spasticity) therefore need to cover both active and passive function, and should ideally reflect performance in real life, as opposed to that just observed in the clinic setting. Patient-reported measures offer that advantage as well as minimising the burden of data collection for clinicians in a busy clinical environment.

Reduction in spasticity has been demonstrated in a number of randomised trials following botulinum toxin (BoNT) and physical interventions [2-6]. Change in function has also been demonstrated by some authors for passive function [2, 7]. However due in part to limitations in currently available evaluation tools, these benefits have been harder to quantify [1, 8].

A previous systematic review of the published literature [9] demonstrated that, although a number of tools addressed aspects of either passive (Leeds Adult Spasticity Impact Scale - LASIS [2, 10]) or active function (Motor Activity Log - MAL [11]) none provided a comprehensive assessment of both parameters. This led to the development of the Arm Activity measure (ArmA) – a new self-report measure of active and passive

function in the upper limb. Conceptualisation and development has been presented for publication elsewhere (reference in press). The aim of this work was to evaluate its psychometric properties.

Tools to measure outcome in rehabilitation should be subject to rigorous evaluation to confirm that they provide a valid and reliable assessment of the clinical parameters in question and to understand their metric properties. The Scientific Advisory Board of the Medical Outcomes Trust has defined a set of attributes and review criteria against which the psychometric properties of health status and quality of life instruments may be judged, and these also form a useful framework for evaluation [12]. In the present article, we use this framework to provide a preliminary evaluation of the psychometric properties of the ArmA. Key criteria addressed were those for; construct validity, internal consistency, unidimensionality, scaling, test re-test reliability, responsiveness, interpretability and feasibility for use in routine clinical practice [12].

## Methods

Ethical approval for the research programme was received (number 05/Q1604/110).

The primary population for psychometric analysis was a consecutive cohort of patients (n=58) with upper limb spasticity presenting for treatment at two specialist rehabilitation units in London (Group 1). However as the majority of these had severe upper limb impairment and intervention was primarily focused on passive function improvement, a second group (n=36) was purposively selected to expand the range of scores in the active function sub-scale (Group 2). The participants in group 2 were recruited from patients with arm impairment seen at the same two rehabilitation units.

They were chosen on the basis that they were able to perform at least one of the active function activities within the ArmA.

The setting for this work utilised two specialist rehabilitation and spasticity management services, both providing inpatient rehabilitation. In addition, both services provided spasticity management clinics where botulinum toxin and physical interventions (e.g. splinting) were treatment options. One service also provided spasticity intervention in an outreach capacity across a geographical region.

Both groups were assessed at baseline (Time 1) and one day later (Time 2), for the evaluation of repeatability. In addition, group 1 were assessed at 8 weeks (Time 3), following treatment for upper limb spasticity with botulinum toxin-A (BT-A) and physical therapy (PT).

Data were collected at two sites. Both services provide inpatient rehabilitation, outpatient clinics including spasticity clinics, and managed patients through a spasticity management integrated care pathway (ICP).

Power Calculation: The sample size was based on the criteria by Terwee and colleagues for evaluation of construct validity and test re-test reliability in groups of at least 50 participants [12] and recommendations by Sim and Wright for sample size in reliability studies [13]. Assuming a null value for Kappa of 0.50 and a power of 90% a sample size of 88 was required. In addition, this would conform to the recommendations for minimal total sample size of 50 participants for principal components analysis (PCA) [14].

Patients included were adults aged between 18 and 85 years with hemiplegic upper limb impairment affecting either active or passive function. In addition for Group 1: undergoing treatment for spasticity of the upper limb requiring BT-A intervention and PT. Patients were excluded if they declined to participate or if their family and/or treating team declined on their behalf; or if they were unable to complete a questionnaire and no carer (professional or family) was available to undertake questionnaire completion on their behalf.

**Measures**

The measures used in this work were selected to allow comparison with the ArmA and test aspects of reliability and validity. Measures for this purpose were selected either on the basis of an earlier systematic review [9] in the case of the functional measures or were widely utilised tools in research and clinical practice.

The ArmA, LASIS and DASH were used to rate function by patients on the basis of activities performed over the preceding seven days.

The **Arm Activity Measure (ArmA)** version 1 [15] is a patient or carer-rated 20-item measure of difficulty in passive and active arm function. It comprises a seven-item Passive function subscale and thirteen-item Active function subscale, and uses a Likert scoring system between 0 (No difficulty) and 4 (unable to do task). The Passive function sub-scale scores range from 0 (high function) to 28, and the Active function sub-scale scores range from 0 (high function) to 52 and was rated in this evaluation by patient and/or carer.

The **Leeds Adult Spasticity Impact Scale (LASIS)** [2, 10] is a 12-item measure of the impact of spasticity on arm function. The LASIS has two sub-scales: a disability sub-scale consisting of 12 items and a carer burden sub-scale consisting of 9 of the same items. The LASIS uses a scale between 0 (No difficulty) and 4 (Unable to do task). The LASIS is applied as a structured interview with the patient completing all 12 items and the carer completing 9 of the same items. A modified approach to scoring was used. When patients and carers were both involved in the activity, the two scores were combined to produce a mean. Items 1 to 9 were classified as passive function and items 10 to 12 were classified as active function to allow comparison with the ArmA sub-scales.

The **Disabilities of the Arm Shoulder and Hand (DASH)** [16] questionnaire comprises 30 items, 21 of which are arm active function items, 5 are symptom related (for example pain) and the remaining 4 items review the impact of arm impairment on wellbeing and participation. The DASH uses a scale between 0- none and 5- extreme difficulty for functional tasks. The DASH is rated by the patient. In this study for comparison with the ArmA an overall total for the measure was not produced, instead a total sum was produced for the active function items (items 1 to 21) for comparison with ArmA.

The **Barthel Index** [17] is a measure of global disability and function. The Barthel Index self-completion version was completed by patient or carer [18]. The measure comprises 10 items relating to personal ADLs, with each item scored on either a scale of 0 to 2 or 0 to 3, giving a total score range from 0 (total dependence) to 20 (complete independence).

The **feasibility questionnaire** was used to evaluate ease of use, relevance and value in the clinical situation. It comprises one question each for a) time to complete, b)

relevance, b) usefulness of the active function section, d) usefulness of the passive function section and e) ease of completion. Each question is rated on a 5-point Likert scale (e.g. for ease of completion: Very easy, Easy, Moderate, Difficult, Very difficult).

A patient- and clinician-rating of goal attainment was used.

**Goal Attainment Scaling (GAS)** [19] is a method of evaluating the extent to which patient's individual goals are achieved in the course of intervention. Scoring of GAS followed the approach proposed by Turner-Stokes [20], but was used in this study to identify 'responders' and 'non responders' to treatment for spasticity without calculation of the 'T' score. Patients were categorised as responders if they achieved a score of 0 to +2 for their primary treatment goal, and non-responders of they achieved a score of -1 to -2.

The **Modified Ashworth Scale (MAS)** [21, 22] is a clinical measure of spasticity which forms a single item scale from 0 (no increase in muscle tone) to 4 (affected part rigid in flexion or extension), with an additional point at +1 (slight increase in muscle tone…) producing a six-point scale. The MAS therefore provides a single score to represent spasticity.

For group 1, goals were set prior to intervention using the GAS method for negotiating and recording goals [19, 20]. The ArmA, LASIS, DASH, Barthel Index, Modified Ashworth Scale and feasibility questionnaire were recorded at baseline. The ArmA was repeated one day later to enable the evaluation of test-retest reliability. Measures were then repeated at 8 weeks for group 1 to enable the testing of responsiveness following intervention. In addition, achievement of goals set was recorded on a 5-point scale (-2 to +2) [19, 20].

Patients gave written informed consent (n=63). For those patients unable to consent, assent was given by the next of kin and the treating team (n=29). A cleaned, validated dataset was exported to SPSS v15 [23] and STATA v10 [24] statistical packages for analysis.

**Analysis plan**

Whole sample:

Floor and ceiling effects were assessed in the study population by considering the percentage of participants at either extreme of the subscales according to the criteria by Terwee and colleagues [12].

Construct validity was evaluated at Time 1 by comparing the passive and active function subscales of the ArmA with the respective components of the LASIS and the DASH using the Spearman rank order correlation coefficient for convergent and divergent validity.

Cronbach's alpha (Time 1) was used to evaluate internal consistency. Applying the criteria of Terwee and colleagues [12], a positive rating for internal consistency was given when ratings for Cronbach's alpha were between 0.70 and 0.95 [12].

The unidimensionality of the ArmA sub-scales were initially evaluated using principal component analysis. The results from principal component analysis were evaluated by initially considering Scree plots and Eigenvalues above 1 according to the criteria by Kaiser [25]. To confirm these findings, a Monte Carlo analysis was carried out according to the method by Horn [26].

Dimensionality was further explored using Mokken analysis and the relationship between different items in a scale can be explored using four assumptions. The first is that items form a unidimensional scale. The second is that item scores are locally independent, which means that item scores are independent within a group of persons with the same degree of the construct of interest [27]. The third assumption is that the 'item response function' for each item is a non-decreasing function of the trait, which means each item would be expected to change in the same direction and to a related degree of change against the construct [27]. The fourth and final assumption measured is that 'item response functions' do not intersect, meaning that the items in the scale have an invariant hierarchical ordering across the latent trait when tested [27, 28]. Assumptions one to three need to be satisfied to accept differentiation between respondents (Monotone Homogeneity – indicated by the H coefficient), and in addition, assumption four is required to accept differentiation between items in an ordinal scale (Double Monotonicity – indicated by the *crit* value together with the H index). van Schuur identifies that no item in a unidimensional scale should have an item H below 0.3 [29]. Mokken analysis (monotone homogeneity) was applied in this analysis a) to confirm the constructs of the ArmA, and b) as a preliminary evaluation of the ordinal scale structure of the items in the passive and active sub-scales. Mokken analysis was applied as a preliminary and further evaluation of scaling properties is ongoing.

Reproducibility (test re-test reliability) of ArmA was evaluated between Time 1 (Baseline) and Time 2 (one day later) using Quadratic Weighted Kappa coefficients for test re-test reliability to comply with the recommendations of Terwee et al [12].

Feasibility was evaluated (Time 1) using a self-completed questionnaire administered following ArmA completion. Patients and carers rated the timeliness, ease of use, relevance, and value in the clinical situation of the ArmA.

Group 1 only:

Responsiveness of the ArmA was evaluated between Time 1 (baseline) and Time 3 (8 weeks) following BT-A injection. Responsiveness was determined by comparing the change in ArmA between responder and non-responder groups using the non-parametric Mann-Whitney U test. It was expected that the ArmA would identify a significant difference between the responder and the non-responder groups for passive function as defined by primary goal outcome (GAS) at Time 3 (8 weeks) following baseline. A significant change was not expected in the group for active function, due to the lack of change identified in groups for active function who receive BT-A [30, 31].

Interpretability was based on the estimation of Minimal Important Change (MIC). MIC was calculated using two methods; a criterion-based method and a distribution-based method. The criterion-based method was produced by calculating the mean change in ArmA passive and active sub-scales in the responder group [32]. The distribution-based method [32] was calculated by using half the baseline (Time 1) standard deviation for ArmA as an estimate of MIC. Both of these methods use parametric assumptions and therefore can only provide a preliminary indication of interpretability for the ArmA (which is an ordinal scale).

## Results

A total of 92 patients (58 Group 1, 34 Group 2) participated from 103 patients approached (63 Group 1, 40 Group 2). Reasons for non-participation were not having BT-A intervention after assessment (n=4 in Group 1) or declining to participate (n=1 in Group 1, n=6 in Group 2). See figure 1 for numbers of participants included at each stage of analysis.

***Insert Figure 1 about here***

The demographic characteristics of participants are described in Table 1.

***Insert Table 1 about here***

The medians and interquartile ranges for the measures used are shown in Table 2.

***Insert Table 2 about here***

ArmA passive function scores were distributed over the full range of the measure from 0 to 28, with an increased frequency in the centre of the scale. The modal score was 13, rated by 10 (11%) of participants. In the active function sub-scale, a complete range of scores was produced at Time 1 from 0 to 52. However a ceiling effect occurred with 37% of scores for active function at the maximum point on the scale (52 – i.e. totally unable for all items).

**Construct validity (Assessed at time 1):** The passive function sub-scale of the ArmA was found to correlate with passive function items in LASIS (Rho 0.50; p=0.01) demonstrating convergent validity, but not with active function items (Rho 0.02; p=0.9) demonstrating divergent validity as expected. Similarly, the active function sub-scale correlated with LASIS active items (Rho 0.48; p=0.01) and DASH active items (Rho 0.63; p=0.01) but did not with the LASIS passive function items (Rho 0.23; p=0.078) again demonstrating convergent and divergent validity in the expected pattern.

**Internal consistency (Time 1):** Cronbach's alpha was 0.85 for the ArmA passive function sub-scale and 0.96 for active function demonstrating high internal consistency.

**Principal component analysis:** The passive function sub-scale had one component with an Eigenvalue above 1 [25]. These results indicate a relatively coherent single construct of passive function confirmed with parallel analysis [26]. The active function sub-scale had two components with an Eigenvalue above 1 [25]. But using Horn's method of parallel analysis, only the first factor was retained indicating a single principal component for active function [26].

To support the single construct interpretation for the active and passive function sub-scales, factor analysis was undertaken on the combined items from both sub-scales using principal components analysis followed by Promax rotation. Although three components had Eigen values above 1, a Monte Carlo analysis was again conducted and confirmed two components representing the constructs. Promax rotation was performed on the combined items and is shown in Table 3.

***Insert table 3 about here***

All 20 items were then plotted in 2 dimensional space shown in Figure 2.

***Insert figure 2 about here***

Figure 2 demonstrates that the passive function items group together as do the active function items, supporting two constructs. However, one active function item (a81) "difficulty with balance when walking due to your arm" is more closely related to passive function rather than active function (see figure 2).

**Mokken Analysis:** The items in the passive function sub-scale produced an overall H coefficient for the scale of 0.48 with no individual item H-coefficients below 0.3 and fits criteria for a moderately strong unidimensional scale [29, 33], shown in Table 4.

***Insert table 4 about here***

Only item 6, had a *crit* value above 40 (actual value 43) outside the recommended range of crit values [33] indicating possible violations of monotonicity. Although in this evaluation the aim was not to confirm double monotonicity, removal of items was explored to improve the overall item H and ensure *crit* values below 40. Removal of items 7 and 6 resulted in a 5-item scale with an overall item H of 0.56 indicating a strong scale and highest *crit* value of 15 indicating no violations of monotonicity.

The overall H coefficient for the active function total scale of 0.71 fits criteria for a strong unidimensional scale [29, 33] shown in Table 4. The item with the lowest H index was again "difficulty with balance when walking due to your arm", which had been discussed regarding its fit to either the passive or active function dimension. No item, had a *crit* value above 40 outside the recommended range of crit values [33] indicating no violations of monotonicity.

**Test re-test reliability**: Quadratic weighted Kappa coefficients for the ArmA sub-scale scores at time 1 and 2 were 0.90; CI 0.68-1.12 (passive function – Figure 3) and 0.93; CI 0.71-1.15 (active function – Figure 3).

***Insert Figure 3 about here***

Item-level agreement ranged from 92% to 99%, and Quadratic weighted Kappa co-efficients ranged from 0.71 to 0.94. The Kappa coefficients conformed to "substantial" or "almost perfect" criteria for all items [34].

The ArmA identified a significant difference between responder and non-responder groups for the passive function sub-scale at 8 weeks (U = 98.5; p = 0.01). In this respect the ArmA was more responsive than the LASIS passive items (U = 127.0; p = 0.07), LASIS active items (U = 167.0; p = 0.39), DASH active items (U = 176.5; p = 0.92) or Barthel Index (U = 200.5; p = 0.17) none of which demonstrated any difference between the two groups.

A significant difference was not shown for the active function sub-scale (U = 163.4; p = 0.35). However this was expected as only 4 of 58 participants in group 1 had a primary goal relating to active function. This finding supports the responsiveness of the active function sub-scale in reflecting lack of change in the study group for the active function domain.

**Interpretability:** Using the criterion-based method for assessing MIC, a clinically meaningful change would be 2.5 points on the passive function sub-scale and 1.1 on the active function sub-scale. Using the distribution-based method, the corresponding figures were 3 and 2.5 points respectively.

**Feasibility**, ease of completion was rated as 'Very easy' to 'Moderately easy' by 90% of patients or carers. Completion of the ArmA was undertaken by 83% of participants in 10 minutes or under. Relevance of the overall scale was rated by 77% of respondents as 'Very relevant' to 'Moderately relevant'. The active function sub-scale was rated as

'Very useful' to 'Moderately useful' by 71% of respondents and the passive function subscale by 88% of respondents.

## Discussion

In this evaluation we present a preliminary psychometric evaluation of the ArmA in relation to the Medical Outcomes Trust Quality Criteria.[12]. Construct validity is supported, with confirmation of predicted moderate correlations of the ArmA with comparison measures addressing broadly similar parameters. Internal consistency and test-retest reliability, were good in both ArmA sub-scales. Principal Components Analysis (PCA) demonstrated one principal component for each of the sub-scales. Following Mokken analysis the unidimensionality of the two sub-scales is supported, although should be seen as preliminary evaluation due to the limited number of participants and the ceiling effect present in the active function data. However the passive function sub-scale appears to be a unidimensional scale satisfying the monotone homogeneity model in its current form and may conform to double monotonicity when further evaluation is undertaken.

In this study group, responsiveness was demonstrated in passive function, but could not be fully addressed for the active function subscale as few patients showed any change in this respect. Our preliminary findings suggest that a change of 2-3 on the Passive function subscale represents clinically important change, but further evaluation is required for the active function subscale and in patient groups receiving other types of intervention.

To our knowledge, the ArmA is the only published self-report measure in the current literature that addresses both passive and active function of the paretic upper limb in a comprehensive manner. The ArmA is designed for self-completion, making it potentially useful for patient and/or carer completion at a clinic or return by post following clinic visits with low clinician burden. The ArmA but may also be used as a basis for structured interview undertaken at the clinic visit. The findings presented here in the context of treatment for upper limb spasticity, provide preliminary psychometric support for its further testing and preliminary use.

As noted above, a number of other tools have been developed and used in this context, but they also have limitations. The LASIS [2, 10] primarily evaluates passive function, with a limited focus on active function and has not received formal psychometric testing to date. The Disability Assessment Scale (DAS) developed by Brashear and colleagues [21] is a clinician reported assessment and therefore does not represent the views of patients and carers. Again DAS also does not address active function for the small minority of patients where active function improvement is possible following BT-A and physical interventions.

The authors recognise a number of limitations to this study.

Firstly, the addition of a purposively selected sample of participants (Group 2) to our consecutive cohort. This was done to extend representation of active function in our baseline evaluation of validity and reliability, but may have introduced some selection bias. Moreover, the fact that the additional sample was not undergoing specific treatment for their upper limb impairment meant that it was not possible to include them in the evaluation of responsiveness.

Secondly, the period during which this study was undertaken the knowledge base has grown with respect to the nature of goals for treatment in focal interventions for spasticity. A number of studies have demonstrated that goals related to passive tasks are both set more often and are more likely to be achieved than those relating to active function [35, 36]. Nevertheless qualitative analysis of goal attainment has demonstrated change over a range of patient experience, including active tasks [35]. It is therefore pertinent to have a tool that measures both aspects of function.

Thirdly, previous work in patient-reported outcome measures [37],standardised measures in general practice [38] and rehabilitation [39] has emphasised that introduction of outcome measures into practice require clinicians to have ownership of the use of such measures.

Fourthly, a further limitation is that the evaluation of feasibility was focused on patients and carers but not clinicians. Patients and carers were prioritised as the most important and appropriate sources of data as they are the main users of the measure. Although the views of clinicians were obtained during the ArmA development and were used extensively in item reduction and confirmation, they were not sought again formally during the testing of the ArmA. Views from professionals may have been valuable in considering the impact of the ArmA in practice from the clinician's perspective.

Following psychometric evaluation of the ArmA, one passive function item excluded during development has been identified which merits further consideration regarding its place in the tool. During item reduction, 'cleaning around the affected elbow' was

removed during the first round of Delphi consultation. This item was removed on the recommendation of eight members of the consultation group (n=10). However from a clinical perspective 'Ease of elbow crease hygiene' continued to be set as a goal (n=6) for participants in the psychometric evaluation. This item has been added to the current version of the ArmA and revaluation of the properties of the modified measure is on-going. Further evaluation of the measurement scaling properties of the ArmA are also ongoing, including the active function sub-scale of the measure.

Despite these limitations the study provides preliminary support for the ArmA as a valid, reliable and potentially responsive tool for the evaluation of treatment in upper limb spasticity. Further evaluation, including additional exploration of its scaling and measurement properties is now underway. The ArmA is freely available to use and can be obtained with full instructions for completion, from the King's College London, Department of Palliative Care, Policy and Rehabilitation web site (http://www.csi.kcl.ac.uk/arma.html).

**Clinical messages**

- To our knowledge, the ArmA is the only published self-report measure that addresses both passive and active function of the paretic upper limb in a comprehensive manner.
- The ArmA is designed for self-completion, making it potentially useful for patient and/or carer completion at a clinic or return by post following clinic visits with low clinician burden.

- In the context of treatment for upper limb spasticity, the ArmA has demonstrated preliminary psychometric support for its further testing and initial use.

## Declarations of interest

The authors report no declarations of interest.

## Acknowledgements

We would like to thank patients, carers and colleagues who helped with this work.

## References


1. Sheean, G.L., *Botulinum treatment of spasticity: Why is it difficult to show a functional benefit?* Trauma and Rehabilitation, 2001: p. 771-776.
2. Bhakta, B.B., J.A. Cozens, M.A. Chamberlain, and J.M. Bamford, *Impact of botulinum toxin type A on disability and carer burden due to arm spasticity after stroke: a randomised double-blind placebo-controlled trial.* Journal of Neurology, Neurosurgery and Psychiatry, 2000a. **69**(2): p. 217-221.
3. Bakheit, A.M., S. Pittock, A.P. Moore, M. Wurker, S. Otto, F. Erbguth, and L. Coxon, *A randomised, double-blind, placebo-controlled study of the efficacy and safety of botulinum toxin type A in upper limb spasticity in patients with stroke.* European Journal of Neurology, 2001. **8**(6): p. 559-565.
4. Brashear, A., M.F. Gordon, E. Elovic, V.D. Kassicieh, C. Marciniak, M. Do, C. Lee, S. Jenkins, and C. Turkel, *Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke.* New England Journal of Medicine, 2002. **347**(6): p. 395-400.
5. Bakheit, A.M.O., N.V. Fedorova, A.A. Skoromets, S.L. Timerbaeva, B.B. Bhakta, and L. Coxon, *The beneficial antispasticity effect of botulinum toxin type A is maintained after repeated treatment cycles.* Journal of Neurology, Neurosurgery & Psychiatry, 2004b. **75**(11): p. 1558-1561.
6. Bhakta, B.B., R.J. O'Connor, and J.A. Cozens, *Associated reactions after stroke: a randomised controlled trial of the effect of botulinum toxin type A.* Journal Rehabilitation Medicine, 2008. **40**: p. 36-41.
7. Francis, H.P., D.T. Wade, L. Turner-Stokes, R.S. Kingswell, C.S. Dott, and E.A. Coxon, *Does reducing spasticity translate into functional benefit? An exploratory meta-analysis.* Journal of Neurology, Neurosurgery & Psychiatry, 2004. **75**(11): p. 1547-1551.
8. Sheean, G., N.A. Lannin, L. Turner-Stokes, B. Rawicki, and B.J. Snow, *Botulinum toxin assessment, intervention and after-care for upper limb hypertonicity in adults: international consensus statement.* European Journal of Neurology, 2010. **17**(Supplement 2): p. 74-93.
9. Ashford, S., M. Slade, F. Malaparade, and L. Turner-Stokes, *Evaluation of functional outcome measures for the hemiparetic upper limb: a systematic review* Journal of Rehabilitation Medicine, 2008. **40**(9): p. 787-795.
10. Bhakta, B.B., J.A. Cozens, J.M. Bamford, and M.A. Chamberlain, *Use of botulinum toxin in stroke patients with severe upper limb spasticity.* Journal of Neurology, Neurosurgery & Psychiatry., 1996. **61**(1): p. 30-35.
11. van der Lee, J.H., H. Beckerman, D.L. Knol, H.C.V. de Vet, and L.M. Bouter, *Clinimetric properties of the Motor Activity Log for the assessment of arm use in hemiparetic patients.* Stroke, 2004. **35**: p. 1-5.
12. Terwee, C.B., S.D. Bot, M.R. de Boer, D.A. van der Windt, D.L. Knol, J.H. Dekker, L.M. Bouter, and H.C. de Vet, *Quality criteria were proposed for measurement properties of health status questionnaires.* Journal of Clinical Epidemiology, 2007. **60**(1): p. 34-42.
13. Sim, J. and C.C. Wright, *The Kappa statistic in reliability studies: Use, interpretation and sample size requirements.* Physical Therapy, 2005. **85**(3): p. 257-268.
14. Barrett, P. and P. Kline, *A comparison between Rasch analysis and factor analysis of items in the EPQ.* Journal of Personality and Group Behaviour, 1981. **1**: p. 1-21.

15. Royal College of Physicians, British Society of Rehabilitation Medicine, Chartered Society of Physiotherapy, and Association of Chartered Physiotherapist Interested in Neurology, *Spasticity in adults: management using botulinum toxin - National Guidelines*, 2009, Royal College of Physicians, Clinical Effectiveness and Evaluation Unit: London.
16. Hudak, P., P. Amadio, and C. Bombardier, *Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand).* American Journal of Industrial Medicine, 1996. **29**: p. 602-608.
17. Wade, D.T. and C. Collin, *The Barthel ADL index: a standard measure of physical disability?* International Disability Studies, 1988. **10**: p. 64-67.
18. Gompertz, P., C. Pound, and S. Ebrahim, *A postal version of the Barthel index.* Clinical Rehabilitation, 1994. **8**(3): p. 233-239.
19. Ashford, S. and L. Turner-Stokes, *Goal attainment for spasticity management using botulinum toxin.* Physiotherapy Research International, 2006. **11**(1): p. 24-34.
20. Turner-Stokes, L., *Goal attainment scaling (GAS) a practical guide.* Clinical Rehabilitation, 2009b. **23**(4): p. 362-370.
21. Brashear, A., R. Zafonte, M. Corcoran, N. Galvez-Jimenez, J.M. Gracies, M.F. Gordon, A. McAfee, K. Ruffing, B. Thompson, M. Williams, C.H. Lee, and C. Turkel, *Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb post-stroke spasticity.* Archives of Physical Medicine & Rehabilitation, 2002. **83**(10): p. 1349-54.
22. Wade, D.T., *Measurement in neurological rehabilitation*1992b, Oxford: Oxford University Press.
23. SPSS, *Statistical package for the social sciences*, 2000, SPSS: Chicago.
24. Stata, *Stata statistical software*, 2001, Stata Corporation: College Station.
25. Kaiser, H.F., *The application of electronic computers to factor analysis.* Educational and Psychological Measurement, 1960. **20**: p. 141–151.
26. Horn, J.L., *A rationale and test for the number of factors in factor analysis.* Psychometrika, 1965. **32**: p. 179-185.
27. Van der Lee, J.H., L.D. Roorda, H. Beckerman, G.J. Lankhorst, and L.M. Bouter, *Improving the action research arm test: a unidimensional hierarchical scale.* Clinical Rehabilitation, 2002. **16**: p. 646-653.
28. Zwinkels, A., C. Geusgens, P. van de Sande, and C. van Heugten, *Assessment of apraxia: inter-rater reliability of a new apraxia test, association between apraxia and other cognitive deficits and prevalence of apraxia in a rehabilitation setting.* Clinical Rehabilitation, 2004. **18**: p. 819-827.
29. van Schuur, W.H., *Mokken scale analysis: Between the Guttman scale and parametric item response theory.* Political Analysis, 2003. **11**: p. 139-163.
30. McCrory, P., L. Turner-Stokes, I. Baguley, S. De Graaff, P. Katrak, J. Sandanam, L. Davies, M. Munns, and A. Hughes, *Botulinum toxin A for treatment of upper limb spasticity following stroke: A multi-centre randomized placebo-controlled study of the effects on quality of life and other person-centred outcomes.* Journal of Rehabilitation Medicine, 2009. **41**: p. 536-544.
31. Shaw, L., H. Rodgers, C. Price, F. Van Wijck, P. Shackley, N. Steen, M. Barnes, G. Ford, L. Graham, and BoTULS_investigators, *BoTULS: a multicentre randomised-controlled trial to evaluate the clinical effectiveness and cost-effectiveness of treating upper limb spasticity due to stroke with botulinum toxin A.* Health Technologies Assessment, 2010. **14**(26): p. 1-113, iii-iv.

32. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes on health related quality of life. The remarkable universality of half a standard deviation.* Medical Care, 2003. **41**: p. 582-592.
33. Molenaar, I.W., K. Sijtsma, and P. Boer, *Users Manual MSP5 for Windows; A program for Mokken scale analysis for polytomous items*2000: iec ProGAMMA, Groningen, Netherlands.
34. Landis, J. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**: p. 159-174.
35. Turner-Stokes, L., I. Baguley, S. De Graaff, P. Katrak, L. Davies, P. McCrory, and A. Hughes, *Goal attainment scaling in the evaluation of treatment of upper limb spasticity with botulinum toxin: A secondary analysis from a double-blind placebo-controlled randomised clinical trial.* Journal of Rehabilitation Medicine, 2010. **42**: p. 81-89.
36. Bakheit, A.M., B. Zakine, P. Maisonobe, C. Aymard, K. Fhedoroff, and H. Hefter, *The profile of patients and current practice of treatment of upper limb muscle spasticity with botulinum toxin type A: an international survey.* International journal of rehabilitation research, 2010: p. 199-204.
37. Greenhalgh, J., A.F. Long, and R. Flynn, *The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory?* Social Science and Medicine, 2005. **60**(4): p. 833-843.
38. Meadows, K.A., F. Twidale, and D. Rodgers, *Action research - a model for introducing standardised health assessment in general practice: an exploratory study.* Journal of Evaluation in Clinical Practice 1998. **4**(3): p. 225-229.
39. Turner-Stokes, L., H. Williams, K. Sephton, H. Rose, S. Harris, and A. Thu, *Engaging the hearts and minds of clinicians in outcome measurement – the UK Rehabilitation Outcomes Collaborative approach.* Disability and Rehabilitation, 2012;In press.

## Tables

**Table 1 Demographic characteristics of the study population (n=92)**

| Groups | Group 1 (n=58) | Group 2 (n=34) | Combined (n=92) |
|---|---|---|---|
| Mean age (years) | 47 (SD=17.5) | 42 (SD=15.8) | 44.5 (SD=16.7) |
| Male/female ratio | 32:26 | 22:12 | 54:38 |
| Barthel Index | - | - | 12 (2-15) |
| DIAGNOSIS | | | |
| Stroke | 30 (52%) | 18 (53%) | 48 (52%) |
| Right hemisphere | 13 (22%) | 10 (28%) | 23 (25%) |
| Left hemisphere | 17 (30%) | 8 (24%) | 25 (27%) |
| Acquired brain injury | 22 (38%) | 6 (18%) | 28 (31 %) |
| Traumatic | 16 (28%) | 5 (15%) | 21 (23%) |
| Anoxic | 6 (10%) | 1 (3%) | 7 (8%) |
| Other | 6 (10%) | 10 (29%) | 16 (17 %) |
| Multiple Sclerosis | 4 (6%) | 2 (6 %) | 6 (7%) |
| Motor neurone disease | 1 (2%) | | 1 (1%) |
| Encephalitis | 1 (2%) | | 1 (1%) |
| CNS Tumour | | 4 (12%) | 4 (4%) |
| Spinal cord injury | | 2 (6%) | 2 (2%) |
| Vasculitus | | 1 (3%) | 1 (1%) |
| Critical care neuropathy | | 1 (3%) | 1 (1%) |

**Table 2: Median and inter-quartile range (IQR) for the study measures**

| **Measure** | | **Time 1**<br>**Baseline**<br>Median (IQR) | **Time 2**<br>**1 day**<br>Median (IQR) | **Time 3**<br>**8 weeks**<br>Median (IQR) |
|---|---|---|---|---|
| **Group 1+2 (n=92)** | **ArmA Passive** | 12 (6-17) | 13 (6-17) | - |
| | **ArmA Active** | 48 (35-52) | 47 (30-52) | - |
| **Group 1 Only (n=58)** | **ArmA Passive** | 14 (10-18) | - | 13 (7-17) |
| | **ArmA Active** | 52 (48-52) | - | 51 (48-52) |
| | **Barthel Index** | 5 (0-15) | - | 5 (0-15) |
| | **LASIS Passive items** | 8 (4-16) | - | 8 (2-11) |
| | **LASIS Active items** | 0 (0-2) | - | 0 (0-2) |
| | **DASH Active items** | 105 (100-105) | - | 105 (99-105) |

**Table 3 Item loadings for combined active and passive function sub-scales following Promax rotation with Kaiser normalisation.**

| Item | **Loading 1st component** | **Loading 2nd component** |
|---|---|---|
| Passive Function | | |
| 1 Cleaning the palm of the hand | 0.445 | **0.860** |
| 2 Cutting finger nails | 0.368 | **0.681** |
| 3 Cleaning the armpit | 0.449 | **0.615** |
| 4 Positioning arm | 0.396 | **0.784** |
| 5 Putting arm through garment sleeve | 0.353 | **0.830** |
| 6 Putting on a glove | 0.338 | **0.653** |
| 7 Putting on a splint | 0.329 | **0.651** |
| Active Function | | |
| 8 Do up buttons on clothing | **0.860** | 0.493 |
| 9 Pick up a glass, bottle, or can | **0.848** | 0.524 |
| 10 Use a key to unlock the door | **0.889** | 0.484 |
| 11 Write on paper | **0.841** | 0.360 |
| 12 Open a previously opened jar | **0.853** | 0.521 |
| 13 Eat with a knife and fork | **0.856** | 0.435 |
| 14 Hold an object still | **0.708** | 0.476 |
| 15 Difficulty with balance when walking | 0.492 | **0.548** |
| 16 Dial a number on home phone | **0.934** | 0.477 |
| 17 Tuck in your shirt | **0.934** | 0.527 |
| 18 Comb or brush your hair | **0.884** | 0.466 |
| 19 Brush your teeth | **0.880** | 0.377 |
| 20 Drink from a cup or mug | **0.883** | 0.385 |

**Bold** indicates the highest loading of each item onto one of the two components

**Table 4 Mokken Analysis**

| | Passive function sub-scale (n=92) | |
|---|---|---|
| **Item** | **Summary per item** | |
| | **Mean** | **Item H** |
| 1 | 1.3 | 0.60 |
| 2 | 2.2 | 0.48 |
| 3 | 1.6 | 0.53 |
| 4 | 1.2 | 0.40 |
| 5 | 1.6 | 0.55 |
| 6 | 2.2 | 0.43 |
| 7 | 1.5 | 0.42 |
| | **Scale H** | 0.48 |
| | **Rho** | 0.85 |

| | Active function sub-scale (n=92) | |
|---|---|---|
| **Item** | **Summary per item** | |
| | **Mean** | **Item H** |
| 1 | 2.5 | 0.51 |
| 2 | 3.1 | 0.64 |
| 3 | 3.4 | 0.73 |
| 4 | 3.1 | 0.72 |
| 5 | 3.3 | 0.73 |
| 6 | 3.2 | 0.72 |
| 7 | 3.2 | 0.79 |
| 8 | 3.4 | 0.69 |
| 9 | 3.3 | 0.72 |
| 10 | 3.3 | 0.80 |
| 11 | 3.3 | 0.73 |
| 12 | 3.2 | 0.74 |
| 13 | 3.4 | 0.76 |
| | **Scale H** | 0.71 |
| | **Rho** | 0.97 |

**Figure 1 Patients screened and response rate at each time point**

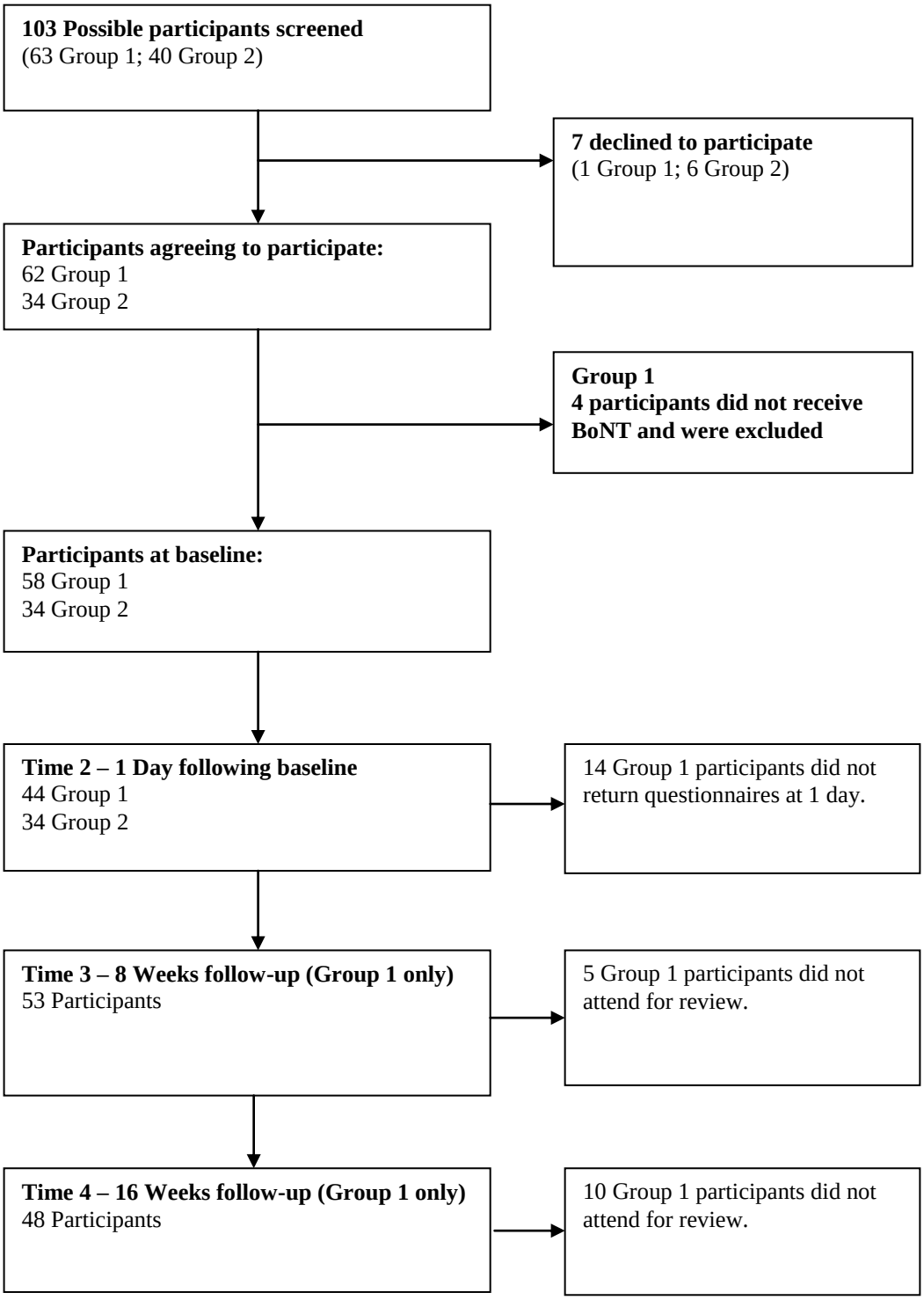
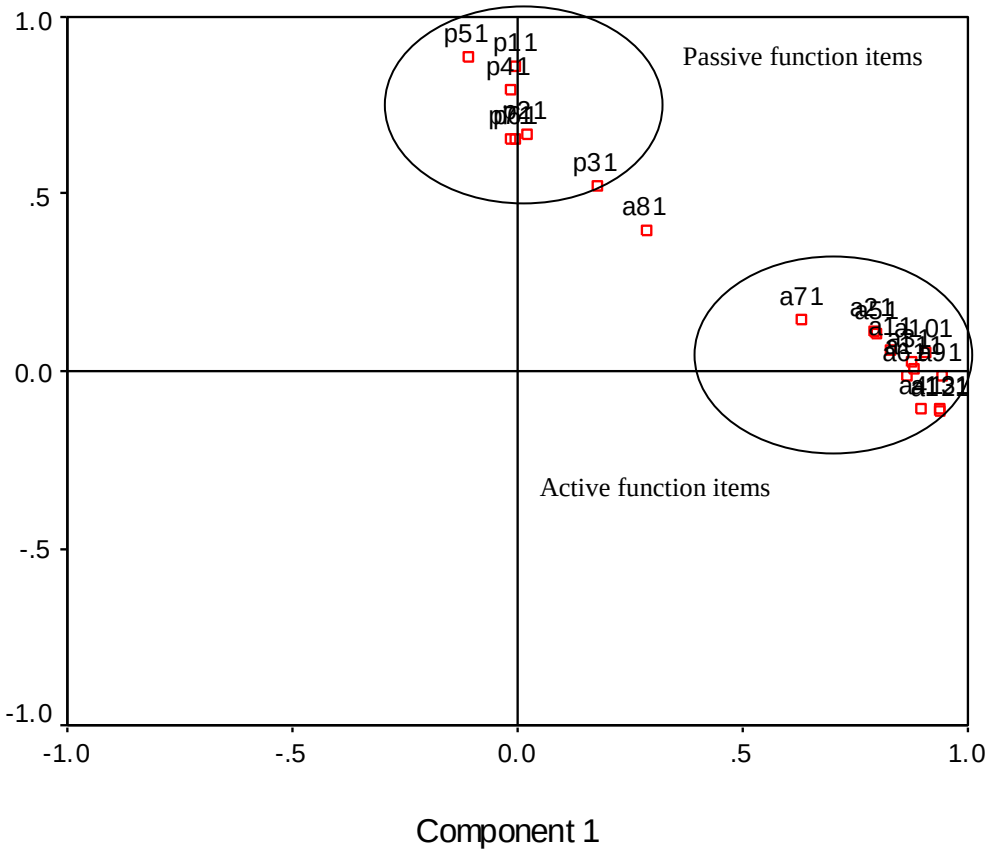
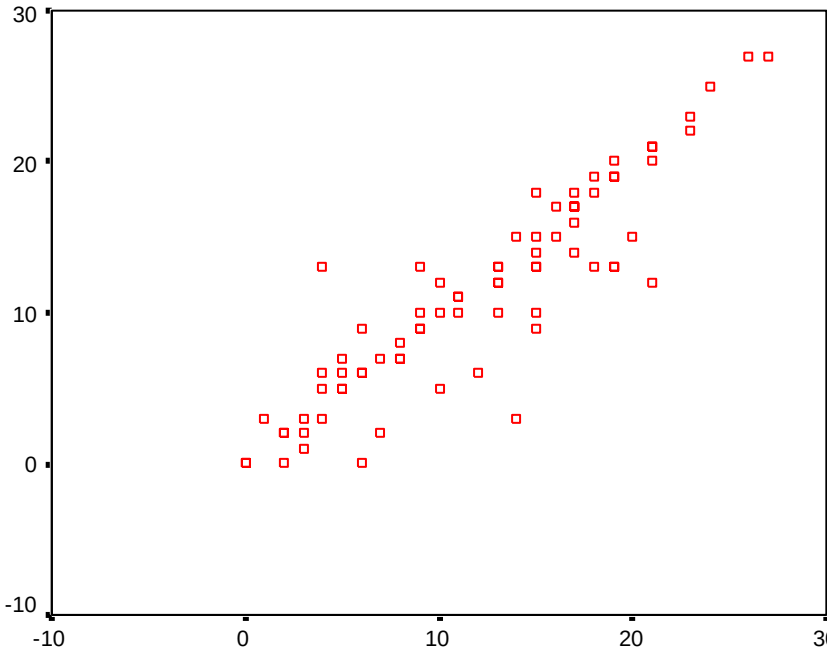


**Figure 2 Item plot in two-dimensional space for active and passive function sub-scales combined (n=92)**

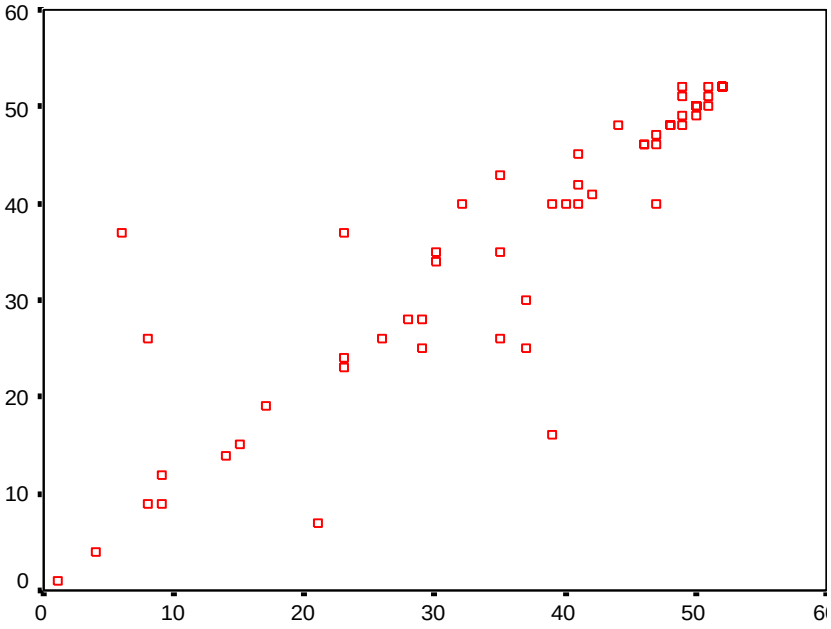



Legend: The active function and passive items form two distinct groups apart from active function item a81 ("difficulty with balance when walking due to your arm"), which is more closely related to the passive function items.

**Figure 3 Test re-test reliability (Time 1 to Time 2) (n=78).**




Passive Function Time 2




Active Function Time 2

## Declarations of interest

The authors undertook the initial development and construction of the ArmA.

## Acknowledgements

We would like to thank patients, carers and colleagues who helped with this work.

**Appendix 1 - Summary of ArmA psychometric properties.**

| Attribute | Criteria | Evaluation |
|---|---|---|
| **Validity** | Face | Confirmed during pilot testing. |
| | Content | ***Aim; population and target concepts:*** The ArmA was designed to provide a low burden measure of difficulty in active and passive arm function for patients undergoing spasticity management in the upper limb.<br>***Item selection and reduction;*** Item selection used a systematic review and patient selected items followed by Delphi consensus process with specialist clinicians and confirmed with a wider consultation.<br>***Interpretability of items:*** Understanding was confirmed during pilot testing [12]. |
| | Criterion-related | Not testable - no accepted gold standard measure for comparison currently exists. |
| | Construct (Correlation with other measures). | ***Convergent***: Passive function sub-scale correlated with passive function items from LASIS (Rho 0.5) and active function was correlated with active function items from LASIS (Rho 0.48) and active function items from DASH (Rho 0.63).<br>***Divergent:*** Passive function did not significantly correlate with DASH active items and LASIS active items. The active function subscale was not correlated with LASIS passive function items. |
| **Reproducibility** | Agreement | Percentage agreement ranged between 91.99 - 96.87 for the passive function scale and 92.15 - 97.52 for the active function scale. |
| | Test re-test Reliability (Weighted Kappa > 0.70 [12]) | Quadratic Weighted Kappa coefficients for the passive function scale were between 0.71 - 0.90 and 0.70 - 0.94 for the active function scale. |

| | | |
|---|---|---|
| **Responsiveness** | | A significant difference was identified between responders and non-responders using the ArmA passive function sub-scale at 8 weeks (U=98.5, P = 0.01), the active function sub-scale demonstrated that no change occurred in this domain as expected. |
| **Interpretability** | | Preliminary MIC was calculated using a criterion-based method (2.5) and a distribution-based method (3.0). |
| **Floor/ceiling effects** | | No significant floor or ceiling effects in the passive function sub-scale, but >15% ceiling effect in the active function sub-scale in the patient group used in this evaluation. |
| **Feasibility and Burden** | Time to administer | Mean time for completion of ArmA was between 5-10 minutes and 83% were completed in 10 minutes or under. |
| | Ease of use | Ease of completion was rated as very easy, easy or moderate by 90 % of patients or carers. |
| | Relevance | Relevance of the overall scale was rated by 77% of respondents as very relevant to moderately relevant.<br>The active function sub-scale was rated as very useful to moderately useful by 71% of respondents and the passive function subscale by 88% of respondents. |
| | Acceptability | Completion of ArmA was rated as very easy to moderate by 90 % of respondents and 77% of respondents rated it as relevant with a mean time for completion of between 5-10 minutes. |
| | Value | The passive function subscale was rated by 88% of respondents as useful (very useful, useful or moderately useful).<br>The active function subscale was rated by 71% of respondents as useful (very useful, useful or moderately useful). |
| **Alternative modes of administration** | | Has been administered during testing as a self-completion questionnaire or as an interview (face-to-face or over the telephone). Only a small minority completed by interview or telephone and further validation of these methods will be needed. |
| **Cultural and language adaptations** | | None currently available |

MIC - Minimal Important Change

**Appendix 2 – Arm Activity Measure (ArmA)**

**In <u>each column</u>, please <u>CIRCLE</u> the amount of <u>difficulty</u> that you or your carer have experienced in doing the activity, over the <u>last 7 days</u>.**

| Activities (affected arm) | Difficulty<br>0 = no difficulty<br>1 = mild<br>2 = moderate<br>3 = severe difficulty<br>4 = Unable to do activity | | | | |
|---|---|---|---|---|---|
| **Section A Caring for your affected arm (not using it in tasks or activities)** | | | | | |
| 1. Cleaning the palm of the hand | **0** | **1** | **2** | **3** | **4** |
| 2. Cutting finger nails | **0** | **1** | **2** | **3** | **4** |
| 3. Cleaning the armpit | **0** | **1** | **2** | **3** | **4** |
| 4. Cleaning the elbow crease | **0** | **1** | **2** | **3** | **4** |
| 5. Positioning arm on a cushion or support in sitting (If never done circle 0) | **0** | **1** | **2** | **3** | **4** |
| 6. Putting arm through a garment sleeve | **0** | **1** | **2** | **3** | **4** |
| 7. Putting on a glove (If never done circle 0) | **0** | **1** | **2** | **3** | **4** |
| 8. Putting on a splint (If never done circle 0) | **0** | **1** | **2** | **3** | **4** |
| **Section B Independently completing tasks or activities using your affected arm** | | | | | |
| 1. Difficulty with balance when walking <u>due to your arm</u> | **0** | **1** | **2** | **3** | **4** |
| 2. Hold an object still while using unaffected hand | **0** | **1** | **2** | **3** | **4** |
| 3. Open (affected hand) a previously opened jar | **0** | **1** | **2** | **3** | **4** |
| 4. Pick up a glass, bottle, or can | **0** | **1** | **2** | **3** | **4** |
| 5. Drink from a cup or mug | **0** | **1** | **2** | **3** | **4** |
| 6. Brush your teeth | **0** | **1** | **2** | **3** | **4** |
| 7. Tuck in your shirt | **0** | **1** | **2** | **3** | **4** |
| 8. Write on paper | **0** | **1** | **2** | **3** | **4** |
| 9. Eat with a knife and fork | **0** | **1** | **2** | **3** | **4** |
| 10. Dial a number on home phone | **0** | **1** | **2** | **3** | **4** |
| 11. Do up buttons on clothing | **0** | **1** | **2** | **3** | **4** |
| 12. Comb or brush your hair | **0** | **1** | **2** | **3** | **4** |
| 13. Use a key to unlock the door | **0** | **1** | **2** | **3** | **4** |
| Total Score | Section A ☐☐<br>Section B ☐☐ | | | | |

Totalling section A and B separately produces a total score for each sub-scale of the measure.
*The sub-scales should not be combined.*

**Item 4 inserted following preliminary psychometric testing.**