

## Review Article

# Action research arm test as a tool for assessment of motor recovery in patients with stroke: a critical review

Sukanya Parag Dandekar\*, Suvarna Ganvir

Department of Neurosciences, D. V. V. P. F's College of Physiotherapy, Ahmednagar, Maharashtra, India

**Received:** 14 June 2019

**Revised:** 28 June 2019

**Accepted:** 12 July 2019

### \*Correspondence:

Dr. Sukanya Parag Dandekar,

E-mail: [sukanya.dandekar@gmail.com](mailto:sukanya.dandekar@gmail.com)

**Copyright:** © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

### ABSTRACT

Action research arm test has been used widely clinically for the assessment of upper extremity function post stroke and in various other conditions. Measurement of recovery after stroke is becoming increasingly important with the advent of new treatment options under investigation in stroke rehabilitation research. The Action Research Arm Test scale was developed as the first quantitative evaluative instrument for measuring motor stroke recovery, based on an upper extremity test by Lyle. It is a well-designed, feasible and efficient clinical examination method that has been tested widely in the stroke population. Excellent interrater and intrarater reliability and construct validity have been demonstrated. Limitations of the motor domain include a ceiling effect. Further study should test performance of this scale in specific subgroups of stroke patients and better define its criterion validity, sensitivity to change, and minimal clinically important difference. Based on the available evidence, the Action Research Arm Test is recommended highly as a clinical and research tool for evaluating changes in motor impairment following stroke.

**Keywords:** Scales, Stroke recovery, Upper Extremity

### INTRODUCTION

Action Research Arm Test is considered by many in the field of stroke rehabilitation to be one of the most comprehensive quantitative measures of motor impairment following stroke, and its use has been recommended for clinical trials of stroke rehabilitation.

One of the common deficits after stroke is upper and lower extremity motor impairment. The upper extremity is used principally for reaching, grasping and manipulation, sometimes for lifting the body mass, during periods of postural instability for preserving balance. Effective functional use of upper limb is absolutely dependent on functional grasp and release. Thus, the proper evaluation of upper extremity after stroke is important.

ARAT has also been used for evaluating upper extremity impairment in Parkinson's disease, multiple sclerosis. The most effective use of ARAT is seen in stroke.<sup>1,2</sup> In this article, we review the development of this scale and its measurement properties, reliability, validity, and responsiveness. Limitations of the scale are discussed, and recommendations for future use are presented.

### DESCRIPTION OF THE SCALE

The test consists of 19 items grouped into subtests (grasp, grip, pinch, and gross arm movement) and performance of each item is rated on a 4-point scale ranging from 0 (no movement possible) to 3 (movement performed normally). If subjects score the maximum on the first, most difficult item of each subtest, they are credited with

having scored 3 on all items of the subtest without having to be tested. If the patient scores less than 3, then the second item is tested. This is the easiest item, and if patients score 0 then they are unlikely to achieve a score above 0 for the remainder of the items and are credited with a zero for the other items and the assessor moves onto the next subtest. For example, in the Grasp subtest the first item is lifting a 10 cm 3 block onto a shelf and the second item is lifting a 2.5 cm<sup>3</sup> block. If the patient scores less than 3 for the first item and more than 0 for the second item, then all items in the subtest should be assessed. The maximum obtainable score is 57. There are four subtests: Grasp, Grip, Pinch, Gross Movement. Items in each are ordered so that: if the subject passes the first, no more need to be administered and he scores top marks for that subtest; if the subject fails the first and fails the second, he scores zero, and again no more tests need to be performed in that subtest; otherwise he needs to complete all tasks within the subtest.

The test takes approximately 10 minutes to administer and while no special training is necessary it does require considerable non-standard equipment (various sized blocks of wood, cricket ball, stone, jug and glass, tube, washer and bolt, ball bearing, marble).<sup>3,4</sup>

#### DEVELOPMENT OF THE SCALE

It was first described in 1981 as a modification of an earlier test, the Upper Extremity Function Test (UEFT) and was designed to assess recovery in the upper limb following cortical damage.<sup>5</sup> Impairment of the hand is prime obstacle to re-acquiring competency in performing ADLs. Carroll (1965) made contribution with Upper Extremity Function Test to develop sensitive measure of changes in upper extremity function.<sup>5</sup> The main reasons why Lyle decided to change the UEFT and to construct the ARA test, were:

- The time needed to complete the UEFT, which consisted of 33 items,
- The complexity of certain items,
- Perceived redundancy of other items, notably those involving repetitive finger-thumb opposition movements.

Using Guttman Scale or Scalogram analysis, this scale was re-standardized resolving all the problems. Twenty patients with cortical damage (stroke, road traffic accident, assault, surgery) were included in the study. Average of age of the participants was 53.2 years and average duration of the sustained cortical damage was 46 months. Two independent teams were used to assess. Each patient was tested once by each team. Certain items were eliminated to overcome the problems faced by Carroll. The amount of time used to distinguish a score of 2 versus 3 was not specified by Lyle.<sup>4</sup> A specific time limit was first suggested by Wagenaar et al, 11 who advocated using the mean  $\pm$  2 SDs, as determined from age-matched healthy control subjects. As an extension

of this, we define “takes abnormally long” as 5 to 60 seconds.<sup>6</sup>

During development, Lyle (1981) used Guttman scale analysis to ensure that items were truly hierarchical. This shortens by over 50% the time taken to complete the test. This is an advantage over an alternative outcome measure, the MAS (Hand Movements and Advanced Hand Activities Scales), where each item must be tested as the ordering of items is not truly hierarchical (Sabari et al 2005).<sup>7</sup> Although the scoring of the ARAT appears complex, experience with the test confirms the comment by Lyle.<sup>4</sup> Since Lyle’s publication of the ARA test in 1981, its validity and reliability have been reconfirmed, and this test has been used as an outcome measure in a number of clinical studies.

The responsiveness of the ARA test, which is another important characteristic, in addition to reliability and validity, has been shown to be adequate in the first eight weeks post stroke as well as in chronic stroke patients undergoing forced use therapy.<sup>8</sup>

#### ITEM GENERATION AND REDUCTION

The ARAT, like most motor assessments, requires a human examiner to transform observations of a patient’s movement into a score. Reliance on a human examiner leaves room for variability in scoring, particularly given the innumerable patterns of motor exam abnormality that arise after stroke. Reliance on a human examiner also emphasizes the need for clear methods for testing and rules for scoring; however, little information is available to guide ARAT administration and scoring, although some strides have been made in this regard. In order to assess the reliability and validity of the currently presented method of ARAT scoring 12 subjects with stroke were examined. The ARAT is capable of detecting changes that are in the range of clinically significant values. This assertion is based on the fact that a test is capable of detecting a difference that is equal to the mean  $\geq$  2 SDs of the difference between 2 ratings of the same subject.

This increases confidence that clinically significant changes detected by ARAT are not a result of measurement error. The aspect of the ARAT that could be improved is specification of the amount of time used to define “abnormally long,” which distinguishes a score of 2 versus 3.

Another aspect of the ARAT that requires greater standardization is the source, material, weight, and size of the materials used for examining subjects, variability in which likely influences ARAT scores. In addition, many of the fine details of test administration are not stated in the original report and are open to interpretation, such as body position/posture, test item positioning, and a maximum time allowed completing each ARAT test item. This could be an additional source of score variance across centers and time. The study gives definite

instructions as to how to perform and score each component as well as the dimensions of the materials used in the test. The study has also given the positioning for the test with the normal and abnormal patterns.<sup>9</sup>

I-Ping Hsueh et al 2000 conducted a study by performing ARAT on 3 different heighted tables on stroke patients. Sixty-one patients who had had only one stroke (mean age 63.3 years; median time since stroke onset 81 days; mean ARAT score administered at the standard table 33.8) participated in this study. Each subject was tested three times with the ARAT while sitting at three different tables: a table specially designed for the test and two generally available tables similar in height to the standard table. The specially designed table, or standard table,

has dimensions of 92 cm × 45 cm × 83 cm high and with a shelf of 93 cm × 10 cm positioned 37 cm above the main surface of the table. Two commonly used tables were chosen for administering the ARAT in addition to the original formal setup. One of the tables used in this study (table A) was 2 cm higher than the standard table, and the other (table B) was 3 cm lower. Both of these tables could be considered a common height, and tables of comparable height could most likely be found in any treatment facility. The patients were randomly and equally assigned to three different raters and to three different tables in accordance with a counterbalanced design. All evaluations were completed within a two-day period. The intraclass correlation coefficient (ICC) for the total scores obtained using the ARAT at the different tables was 0.99, indicating very high agreement. The ICCs were also very high in each of the subscales.<sup>10</sup>

## RELIABILITY AND VALIDITY

The ability to measure attributes in a reproducible and consistent manner is a necessary requirement of a valid and responsive measurement instrument. Excellent intrarater and interrater reliability have been established for the ARAT scale. There is good evidence from several validation studies that the FM scale is indeed measuring what it is intended to measure. Most studies have established construct validity by comparing the FM scale with measures of stroke recovery based on various functional scales reflecting independence in ADL or disability level following stroke.

Ching et al included 50 stroke subjects and three raters evaluated the subjects with the use of ARAT within a three day period. The 3-day period was established to minimize the effect of a possible spontaneous recovery, a confounding variable that could affect the result. All the raters received 30 minute training on administration of ARAT. UEMAS (upper extremity motor assessment scale) was used for the validity part of the study. Intra-class correlation coefficient (ICQ for the total score was 0.98 indicating very high inter-rater reliability. ICCs were also very high in each of the subscales. The score of the ARAT was closely correlated with that of the upper

extremity part of the motor assessment scale, the arm sub-score of the motricity index and the upper extremity movements of the modified motor assessment chart (Pearson  $r = 0.96, 0.87$  and  $0.94$ , respectively)<sup>(11)</sup>.

Johanna et al conducted a study on 20 chronic stroke patients (median age, 62yr; median time since stroke onset, 3.6yr; mean intake ARA score, 29.2). Intrarater reliability of the sum scores and of individual items was assessed by comparing (1) the ratings of the laboratory measurements of 20 patients with the ratings of the same measurements recorded on videotape by the original rater, and (2) the repeated ratings of videotaped measurements by the same rater. Interrater reliability was assessed by comparing the ratings of the videotaped measurements of 2 raters. The resulting limits of agreement were compared with the MCID. All intra- and interrater Spearman's rho and ICC values were higher than .98. The mean difference between ratings was highest for the interrater pair (.75; 95% confidence interval, .02–1.48), suggesting a small systematic difference between raters. Intrarater limits of agreement were 21.66 to 2.26; interrater limits of agreement were 22.35 to 3.85. Median weighted kappas exceeded .92. . A small systematic difference was noted between two raters in one study with a mean difference of 0.75 points and 95% CI 0.02 to 1.48. This same study also proposed a somewhat arbitrary value of 10% of the total range of the scale (i.e. 5.7 points) as the minimum clinically important difference, and then confirmed that a difference of this magnitude could be distinguished from measurement error. Concurrent validity has been confirmed by comparison with the upper limb component of the Fugl-Meyer Assessment and the Motor Assessment Scale (MAS).<sup>6</sup>

Nijland et al, 2010 conducted a study comparing the Wolf Motor Function Test and the Action Research Arm Test for upper limb function after stroke. Forty patients were included in this study. Concurrent validity was determined with Spearman's rank correlation coefficients. Reproducibility was assessed with interclass correlation coefficients (ICCs) and Bland-Altman plots, internal consistency by means of Cronbach's alphas, and floor and ceiling effects were considered to be present if more than 20% of patients fell outside a preliminary set lower and upper boundary. Spearman's rank correlation coefficients ranged from 0.70 to 0.86. ICCs for inter-rater and intra-rater reliability ranged from 0.92 to 0.97. Bland-Altman plots showed a less stable way of scoring for the WMFT, compared with the ARAT. Cronbach's alpha was  $> 0.98$  for both scales. No floor and ceiling effects were found. The study concluded good clinimetric properties for both assessments. The high concurrent validity suggests that ARAT and WMFT have significant overlap with regard to the underlying construct that is being measured.<sup>12</sup>

Rabadi et al, 2006, conducted a study comparing ARAT and FMA as measures of Upper extremity motor weakness in 104 stroke patients. The Spearman rank correlation statistic indicated that the 2 upper-limb motor

scales (ARAT, FMA) correlated highly with one another, both on admission ( $P=0.77$ ,  $P<0.001$ ) and on discharge ( $P=0.87$ ,  $P=0.001$ ). The mean change in score from admission to discharge was  $10\pm 15$  for the ARAT and  $10\pm 13$  for the FMA motor score.

The responsiveness to change as measured by the standard response mean was 0.68 for the ARAT and 0.74 for the FMA motor score. The Spearman rank correlation of each upper-limb motor scale with the FIMADL at the time of admission was as follows: ARAT,  $p$  equal to .32 ( $P<0.001$ ) and FMA motor score,  $P$  equal to 0.54 ( $P<0.001$ ). They concluded that both the FMA motor score and the ARAT were equally sensitive to change during inpatient acute rehabilitation and could be routinely used to measure recovery of upper-extremity motor function.<sup>13,14</sup>

Chen et al in 2012, performed a validation and predictive validity of the Action Research Arm Test in 191 patients receiving stroke rehabilitation in seven medical rehabilitation centers. The study analyzed that the 4-point ARAT scale had a disordered rating scale structure. Further Rasch modeling suggested revising the original 4-point scale into a 3-point scale.

The 19 items measured 1 construct. The item difficulty hierarchy indicated that excluding the gross subtest, a score of 3 on the first item of any other subtest indicated the highest motor ability, and a score of 1 (the revised lowest rating) on the second item indicated the lowest motor ability. Tasks of “place hand behind head” and “place hand on top of head” showed poor item fit and item bias relevant to participants’ ages. The study concluded that the ARAT items can reliably separate participants into 5.44 strata. Moderate to good correlations indicated good predictive validity. The ARAT possesses good psychometric properties in stroke patients with mild to moderate motor severity and without severe cognitive impairment, and has evidence of unidimensionality, predictive validity, and reliability. The revised 3-point rating scale is recommended when the ARAT is administered on this population.<sup>15</sup>

Keh-chung Lin et al, conducted a study comparing the responsiveness and validity of the Box and Block Test (BBT), the Nine-Hole Peg Test (NHPT), and the Action Research Arm Test (ARAT). There were 59 patients with stroke were randomized into one of three rehabilitation treatments for 3 weeks. Standardized response mean (SRM) was used to examine responsiveness and the Spearman rank correlation coefficient ( $\rho$ ) was used to examine concurrent validity. The BBT, NHPT, and ARAT were moderately responsive to change and not significantly different ( $SRM=0.64-0.79$ ). The correlations within the BBT, NHPT, and ARAT were moderate to good at pretreatment ( $\rho= -0.55$  to  $-0.80$ ) and posttreatment ( $\rho= -0.57$  to  $-0.71$ ).

The BBT and ARAT showed fair to moderate correlations with the FMA, MAL, and SIS hand function

domain at pretreatment and posttreatment ( $\rho= 0.31-0.59$ ). Results indicate that the BBT, NHPT, and ARAT are suitable to detect changes over time.<sup>16</sup>

Lang et al, 2006 conducted a study to examine the responsiveness and validity of the Action Research Arm Test (ARAT) in 50 subjects with mild-to-moderate hemiparesis within the first few months after stroke. Data were collected as part of the Very Early Constraint- Induced Therapy for Recovery from Stroke trial, an acute, single blind randomized controlled trial of constraint-induced movement therapy. Subjects were studied at baseline (day 0), after treatment (day 14), and after 90 days (day 90) poststroke. At each time point, subjects were tested on: (1) the ARAT, (2) clinical measures of sensorimotor impairments, (3) in the kinematics laboratory where they performed reach and grasp movements, and (4) clinical measures of disability. Blinded raters performed all evaluations. Analyses at each time point included calculating effect size as indicators of responsiveness, and correlation and regression analyses to examine relationships between ARAT scores and other measures. The ARAT is responsive to change, with effect sizes greater than 1.0 and responsiveness ratios of 7.0 at 3 months poststroke. ARAT scores were related to sensorimotor impairment measures, 3-dimensional kinematic measures of movement performance, and disability measures at all 3 time points. The study concluded that the ARAT is a responsive and valid measure of upper-extremity functional limitation and therefore may be an appropriate measure for use in acute upper-extremity rehabilitation trials.<sup>17</sup>

Chia-Lin Koh, conducted a study to measure the construct validity of ARAT. The aim of the study was to validate the unidimensionality of the Action Research Arm Test (ARAT) using Mokken analysis and to examine whether scores of the ARAT can be transformed into interval scores using Rasch analysis. 351 stroke patients were included in the study. The results supported a unidimensional scale of the 19-item ARAT by Mokken analysis, with the scalability coefficient  $H = 0.95$ . Except for the item “pinch ball bearing 3rd finger and thumb”, the remaining 18 items have a consistently hierarchical order along the upper extremity function’s continuum. Thus, from the study it can be concluded that the 19 item scale has a good construct validity and measures what it intends to measure.<sup>18</sup>

Demonstration of responsiveness, the ability of an instrument to detect clinical change over time, is essential for any evaluative measure. In contrast to reliability and validity, however, formal assessment of responsiveness for outcome measures used in stroke rehabilitation has received relatively little attention

## DISCUSSION

Action research arm test is one of the most commonly used tests used for monitoring the recovery of hand function in pts with hemiplegic stroke. Its reliability and validity has

been found with moderate to excellent in the different research studies conducted on a variety of stroke patients. The characteristic feature of the test is that, it assesses the gross motor as well the fine motor activities of hand which are similar to the functional activities of daily living. After its initial development, the test has been used for many other conditions as and well. However, the exact method of its development and item generation is not readily available in the literature. Positive aspects about the test being its ability to measure the functional aspect of hand and not only the structural impairment, as suggested by International Classification of Function.<sup>19,20</sup>

Ability or inability to perform a task is more important during the course of rehabilitation rather than the structural and functional impairment, according to ICF.<sup>21,22</sup> Though this link between functional impairment and activity limitation is understood indirectly in the test, it is not studied scientifically in the literature. Though it is designed for assessing the motor recovery, the studies related to the responsiveness or detecting the minimal clinical difference are rare. Hence, the floor or ceiling effect of the test cannot be commented upon. This might pose as a hurdle for choosing this test as an outcome measure in the experimental studies especially of a shorter duration. Though sensation is a major contributory factor in the hand function it has not been taken into consideration in the test.

Hence, its choice as an outcome measure is limited especially in the cognition where the sensations are expected to be involved. Previous literature available for the use of ARAT in patients with stroke reveals the heterogeneous population of stroke.

Hence, the generalization of results across all stages of stroke is little difficult and the interpretation of results either of a single study or systematic review should be done cautiously. Administration of the test does not require cognitive training and also its administration on the patient population is not a time consuming process but the small tools that need to be used during the administration of the test needs standardization at a local setup.

Further studies can be taken on the homogeneous stroke population investigating the responsiveness and detecting minimal clinical difference in short term as well as long term research studies. Overall, the test is reliable and valid for its use in patients with stroke.

## CONCLUSION

Action Research Arm Test is a reliable and valid tool and can be used for the assessment of hand functions in patients with stroke.

*Funding: No funding sources*

*Conflict of interest: None declared*

*Ethical approval: Not required*

## REFERENCES

1. Song CS. Intrarater reliability of the action research arm test for individuals with Parkinson's disease. *J Physic Therapy Sci.* 2012;24(12):1355-7.
2. Carpinella I, Cattaneo D, Ferrarin M. Quantitative assessment of upper limb motor function in Multiple Sclerosis using an instrumented Action Research Arm Test. *J Neuroengineer Rehabil.* 2014;11(1):67.
3. De Weerd WJ, Harrison MA. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm test. *Physio Canada.* 1985;37(2):65-70.
4. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *International J Rehabil Res.* 1981;4(4):483-92.
5. Carroll D. A quantitative test of upper extremity function. *J Chronic Dis.* 1965;18(5):479-91.
6. Van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. *Arch Physic Med Rehabil.* 2001;82(1):14-9.
7. Sabari JS, Lim AL, Velozo CA, Lehman L, Kieran O, Lai JS. Assessing arm and hand function after stroke: a validity test of the hierarchical scoring system used in the motor assessment scale for stroke. *Arch Physic Med Rehabil.* 2005;86(8):1609-15.
8. van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. Improving the Action Research Arm test: a unidimensional hierarchical scale. *Clinic Rehabil.* 2002;16(6):646-53.
9. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabilit Neural Repair.* 2008;22(1):78-90.
10. Hsueh IP, Lee MM, Hsieh CL. The Action Research Arm Test: is it necessary for patients being tested to sit at a standardized table?. *Clinical Rehabil.* 2002;16(4):382-8.
11. Hsieh CL, Hsueh IP, Chiang FM, Lin PH. Inter-rater reliability and validity of the action research arm test in stroke patients. *Age Ageing.* 1998;27(2):107-13.
12. Nijland R, van Wegen E, Verbunt J, van Wijk R, van Kordelaar J, Kwakkel G. A comparison of two validated tests for upper limb function after stroke: The Wolf Motor Function Test and the Action Research Arm Test. *J Rehabil Med.* 2010;42(7):694-6.
13. Rabadi MH, Rabadi FM. Comparison of the action research arm test and the Fugl-Meyer assessment as measures of upper-extremity motor weakness after stroke. *Arch Physic Med Rehabil.* 2006;87(7):962-6.
14. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical

- review of its measurement properties. *Neurorehabilitat Neural Repair.* 2002;16(3):232-40.
15. Chen HF, Lin KC, Wu CY, Chen CL. Rasch validation and predictive validity of the action research arm test in patients receiving stroke rehabilitation. *Arch Physic Med Rehabilat.* 2012;93(6):1039-45.
  16. Lin KC, Chuang LL, Wu CY, Hsieh YW, Chang WY. Responsiveness and validity of three dexterous function measures in stroke rehabilitation. *J Rehabil Res Dev.* 2010;47(6):563-71.
  17. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: properties of the action research arm test. *Arch Physic Med Rehabilat.* 2006;87(12):1605-10.
  18. Koh CL, Hsueh I, Wang WC, Sheu CF, Yu TY, Wang CH, Hsieh CL. Validation of the action research arm test using item response theory in patients after stroke. *J Rehabil Med.* 2006;38(6):375-80.
  19. Tempest S, McIntyre A. Using the ICF to clarify team roles and demonstrate clinical reasoning in stroke rehabilitation. *Disability Rehabilat.* 2006;28(10):663-7.
  20. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF Participation. *Disability Rehabilat.* 2005;27(9):507-28.
  21. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disability Rehabilat.* 2005;27(6):315-40.
  22. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF Body Functions. *Disability Rehabilat.* 2005;27(4):191-207.

**Cite this article as:** Dandekar SP, Ganvir S. Action research arm test as a tool for assessment of motor recovery in patients with stroke: a critical review. *Int J Res Med Sci* 2019;7:3224-9.