

The translation into Polish, cultural adaptation, and initial validation of the Action Research Arm Test in subacute stroke patients

Joanna Małecka^{1,A–F}, Magdalena Goliwąg^{2,B,F}, Katarzyna Adamczewska^{1,B,F}, Jacek Lewandowski^{2,E,F}, Dawid Łochyński^{1,A,C–F}

¹ Department of Neuromuscular Physiotherapy, Poznan University of Physical Education, Poland

² Department of Musculoskeletal Rehabilitation, Poznan University of Physical Education, Poland

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation; D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online)

Adv Clin Exp Med. 2025;34(7):1165–1173

Address for correspondence

Joanna Małecka

E-mail: malecka@awf.poznan.pl

Funding sources

None declared

Conflict of interest

None declared

Received on October 26, 2023

Reviewed on March 18, 2024

Accepted on July 29, 2024

Published online on December 4, 2024

Abstract

Background. In Poland, there are limited validated outcome measures to evaluate upper extremity function in stroke patients for clinical and research use. The Action Research Arm Test (ARAT) aims to assess functional performance of the upper extremities.

Objectives. To translate and culturally adapt the original version of ARAT into Polish, and to determine its reliability and validity.

Materials and methods. A Polish version of ARAT (ARAT-PL) was developed using a forward-backward translation. The study then examined 60 patients with subacute stroke. Internal consistency (α), test–retest and inter-rater reliability (intra-class correlation (ICC), κ), standard error of measurement (SEM), minimal detectable change (MDC), and floor and ceiling effects were determined. The construct validity was evaluated using the method of hypothesis testing based on the results of correlations (ρ) between subscale and total scores of the ARAT-PL and the upper and lower extremity section of the Fugl–Meyer Assessment (FMA-UE and FMA-LE).

Results. The internal consistency of the total scores and subscale was excellent ($\alpha = 0.97–0.99$). Test–retest and inter-rater reliability scores were almost perfect ($\kappa = 0.85–1.0$) and excellent for the total and subscale scores (ICC = 0.99–1). The SEM and MDC for the test–retest and inter-rater reliability were 0.479, 1.327 points and 0.335, 0.930 points, respectively. The ceiling effect amounted to 48%. The validity levels with respect to FMA-UE and FMA-LE were found to be high (ρ ranging from 0.70 to 0.83) and moderate (ρ ranging from 0.53 to 0.68), respectively.

Conclusions. A Polish version of ARAT is a reliable and valid tool for assessing upper extremity function in subacute stroke patients in Poland. However, it appears to have a ceiling effect that limits differentiation of patients with mild upper limb impairment.

Key words: stroke, outcome measure, ARAT, FMA, upper extremity function

Cite as

Małecka J, Goliwąg M, Adamczewska K, Lewandowski J, Łochyński D. The translation into Polish, cultural adaptation, and initial validation of the Action Research Arm Test in subacute stroke patients. *Adv Clin Exp Med.* 2025;34(7):1165–1173. doi:10.17219/acem/191775

DOI

10.17219/acem/191775

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported (CC BY 3.0) (<https://creativecommons.org/licenses/by/3.0/>)

Background

Stroke is a leading cause of disability in modern societies.^{1,2} In a large number of patients, motor and functional deficits are observed after a stroke.³ Upper extremity dysfunction is present in approx. 30–66% of stroke survivors⁴; it manifests in limitations in reaching and grasping movements, resulting in serious deterioration in the ability to perform daily living activities.^{5,6}

There are various outcome measures to assess the level of upper limb functional capacity after stroke. Examples frequently seen in the literature and practice include the Fugl–Meyer Assessment for Upper Extremity (FMA-UE), the Jebsen Hand Function Test and the Chedoke Arm and Hand Activity Inventory.⁷ One of the commonly used upper extremity assessment measures for post-stroke patients is the Action Research Arm Test (ARAT).⁸ This test was described by Lyle in 1981,⁹ and was based on Carroll's Upper Extremity Function Test.^{10–12} It was designed for observation of the arm and hand during grasping, gripping, pinching and gross movements in people with cortical damage.¹³ Previous studies have shown good psychometric properties of this instrument in stroke patients.^{13–15} The ARAT has shown excellent internal consistency in stroke patients with mild-to-moderate hemiparesis ($\alpha = 0.98$).¹⁶ The test–retest and inter-rater reliability, as calculated using intraclass correlation (intra-class correlation (ICC): 0.92–0.99) for the total and all subscales' scores, was similarly excellent when tested in patients with subacute stroke.^{7,16–18} Studies examining the convergent validity of ARAT have reported moderate, good or excellent correlations between the absolute ($\rho = 0.77$ – 0.94) and subscale scores ($\rho = 0.67$ – 0.74) of ARAT and FMA.^{17–19} To date, the original version of the ARAT has been translated into Swedish,²⁰ Chinese²¹ and Spanish (in Chile).²² No studies have reported cultural adaptation of ARAT or assessed its reliability and validity on Polish stroke survivors; thus, there is a significant need to develop a Polish version of ARAT.

Objectives

The principal aim of the present study was to estimate the reliability and construct validity of a translated and culturally adapted Polish version of ARAT in the population of subacute patients with stroke. Construct validity was estimated by the correlation of total and subscale ARAT scores with scores for the translated upper and lower extremity sections of the FMA.²³

Materials and methods

Translation and cultural adaptation

Forward, backward and final translation

The ARAT was translated into Polish by following the international guidelines.^{24–27} Permission for the outcome measure translation was obtained from the author, Ronald Lyle (Wolters Kluwer Health rights). In the 1st and 2nd stages, the English version of ARAT was independently translated using a process that encompassed semantic, idiomatic, experimental, and conceptual meaning. The translation was performed by 2 bilingual Polish translators fluent in English (1 specialized in the physiotherapy field). The 2 Polish versions were compared; the differences between the translations were discussed and corrected, and the draft of the common version was jointly established.

In the 3rd stage, this Polish draft was back translated into English independently by 2 certified English translators. A common retranslated English version was then created. This was compared to the original English version by 2 native-speaking translators specialized in health sciences and rehabilitation. Where necessary, corrections were made to the retranslated version.

In the 4th stage, a panel of judges consisting of a neurologist, a psychologist, 2 neurological physiotherapists, 1 clinical neurophysiology and 1 orthopedic physiotherapist, and all the translators compared and discussed the differences between the translated and original versions of the ARAT. Based on the detected differences, the emerging Polish version of ARAT was corrected to obtain a satisfactory harmony between cultural language requirements and the original English instrument. Lastly, the linguistic consistency between the final Polish and original English versions was verified very carefully to ascertain the equivalence of concepts. The final version of ARAT-PL was therefore established (stages 5 and 6).

Study design, participants and initial evaluation

This was a cross-sectional study that lasted for 7 months. We recruited 60 stroke patients from the Bonifraterskie Medical Center hospital in Piaski, Poland. The inclusion criteria were: 1) a diagnosis of stroke, as indicated by computed tomography (CT) scans or magnetic resonance imaging (MRI), 2) hemiparesis, and 3) no additional orthopedic or neurological disabling deficits. The exclusion criteria were: 1) total hemiparesis in the upper extremity (i.e., score = 4 on the Modified Ashworth Scale), 2) serious visual and hearing disorders, 3) cognitive decline that limited administration of the tests, 4) disorders of speech and language, and 5) a native language other than Polish.

During the initial evaluation, we collected demographic data such as age, gender, weight, height, and upper limb lateralization. We also collected clinical data such as duration of illness, type of lesion, location of lesion, involved side, presence of comorbidities, and duration of rehabilitation in the hospital.

The study was approved by the Bioethical Committee of Poznan University of Medical Sciences (approval No. 187/19) and was carried out in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants at the time of their enrolment in the study.

Procedure of assessment

The ARAT-PL and FMA were carried out by 2 experienced neurological physiotherapists trained in administration of each measure. Reproducibility, i.e., the degree to which the score is free from random error, was assessed with test–retest and inter-rater procedures.²⁸ To determine inter-rater reliability, 2 raters independently examined patients at the same time in a quiet hospital room.²⁹ Test–retest reliability was obtained by 1 observer examining the patients twice on the same day with a 2-h gap between assessments.²⁹ The results were collected for the total and subscales of ARAT and FMA.

Outcome measures

Action Research Arm Test

This clinical scale is an evaluative measure to assess dexterity and object-handling ability. It was initially designed for individuals who sustained stroke resulting in hemiplegia. The original ARAT consists of 4 subtests: Grasp, Grip, Pinch, and Gross Movement. Every item within the subtest is assessed on a 4-point ordinal scale and arranged with the most difficult task 1st and the easiest 2nd.¹⁷

Fugl–Meyer assessment

The FMA is a recommended clinical assessment of sensorimotor function of the upper and lower extremities; it has mostly been used after stroke.³⁰ The FMA has been translated into Polish but has not yet been cross-culturally adapted.²³ The present study administered only the motor domain of the (as yet unpublished) Polish version of FMA for the upper extremity (FMA-UE) and lower extremity (FMA-LE).²³ The maximum score for the total motor scale is 100 points (66 for FM-UE and 34 for FM-LE).^{18,31,32}

Statistical analyses

The statistical analysis was made using a software package in Statistica v. 13 (Tibco Software Inc Polska, Cracow, Poland) and R studio program (the psych package, v. 2.4.3).³³

Internal consistency

Internal consistency evaluates the homogeneity of the scale items.²⁸ This study used Cronbach's α to assess internal consistency for the subscales and total scale.^{34,35}

Reliability

The test–retest and inter-rater reliability of ARAT were determined using kappa coefficients, the ICC (ICC 2,k, absolute agreement, the command in RStudio: 1st line – choosing the psych package, 2nd line –library(psych) IC C(dane[,c(1,2)])\$results[5,]), and percentage of agreement (PA).^{8,36,37} Item reliability was established when more than 80% agreement was observed.⁸ The minimal detectable change and standard error of measurement were calculated for all scale items according to the following Equations 1,2:

$$SEM = SD\sqrt{1 - ICC} \quad (1)$$

$$MDC = 1.96 \times SEM \times \sqrt{2} \quad (2)$$

where ICC is the reliability of the test and SD is the standard deviation of all scores.

Validity

Construct validity was evaluated using hypotheses testing according to the guidelines of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN).²⁷ A total of 10 independent hypotheses were formed. For each of them, we defined the anticipated Spearman's rank correlation direction, correlation strength, and rationale; upon these, we based the hypothesis (Table 1,2).³⁸ We assessed the relationships of ARAT-PL scores with scores in FMA-UE (5 hypotheses) and FMA-LE (5 hypotheses) to determine the degree to which they were consistent with the formulated hypotheses. The construct validity rating for ARAT-PL was assessed according to the total number of confirmed hypotheses: 8–10 ($\geq 75\%$) indicated high construct validity, while 5–7 ($\geq 50\%$) indicated a moderate level.²⁷ The threshold values for the correlations determined in the present study (Table 1,2) were based on those indicated by Prinsen et al.²⁷

Spearman's rank correlation is computationally identical to Pearson's product-moment coefficient. Therefore, we computed the required sample size for Spearman's correlation using the software G*Power (v. 3.1.9.2; Kiel University, Germany)³⁹ for estimating sample size for Pearson's correlation (bivariate normal model). We assumed correlation p H1: 0.70 (we expected moderate correlation), alpha of 0.05, a power of 0.95, and correlation p H0: 0.0 (we expected low correlation) for a 1-tailed test (we expected correlation between both measures to be positive). The calculation estimated that at least 44 participants were necessary; our study had a sample of 60.

Table 1. The method of hypothesis testing used to assess construct validity of the ARAT-PL based on its associations with FMA-UE

Hypotheses tested	Rationale	Correlation expected	FMA-UE	
			correlation actual (p < 0.001)	confirmed?
1. There will be at least a moderate-strong positive correlation between the overall result of ARAT-PL and FMA-UE	ARAT-PL and FMA-UE measure similar constructs but asked differently (actual functional vs motor performance)	≥0.50	0.83	yes
2. There will be at least a moderate-strong positive correlation between the result of the ARAT-PL subtest Grasp and the total FMA-UE result	ARAT-PL and FMA-UE measure similar constructs but asked differently (actual grasp versus motor performance)	≥0.50	0.83	yes
3. There will be at least a moderate-strong positive correlation between the result of the ARAT-PL subtest Grip and the total result of FMA-UE	ARAT-PL and FMA-UE measure similar constructs but asked differently (actual grip vs motor performance)	≥0.50	0.80	yes
4. There will be at least a moderate-strong positive correlation between the result of the ARAT-PL subtest called Pinch and the total result of FMA-UE	ARAT-PL and FMA-UE measure similar constructs but asked differently (actual pinch grip vs motor performance)	≥0.50	0.82	yes
5. There will be at least a moderate-strong positive correlation between the result of the ARAT-PL subtest Gross Movement and the total result of FMA-UE	ARAT-PL and FMA-UE measure similar constructs but asked differently (actual gross movement vs motor performance)	≥0.50	0.71	yes

ARAT-PL – Polish version of the Action Research Arm Test; FMA-UE – Fugl–Meyer Assessment for Upper Extremity.

Table 2. The method of hypothesis testing used to assess construct validity of the ARAT-PL based on its associations with FMA-LE

Hypotheses tested	Rationale	Correlation expected	FMA-LE	
			correlation actual (p < 0.001)	confirmed?
6. There will be none-minimal positive correlation between the overall result of ARAT-PL and FMA-LE	ARAT-PL and FMA-LE measure unrelated constructs (actual functional vs lower limb motor performance)	≤0.30	0.59	no
7. There is no or low correlation between the result of the ARAT-PL subtest Grasp and the total result of the FMA-LE	ARAT-PL and FMA-LE measure unrelated constructs (grasp vs lower limb motor performance)	≤0.30	0.68	no
8. There is no or low correlation between the result of the ARAT-PL subtest Grip and the total FMA-LE result.	ARAT-PL and FMA-LE measure unrelated constructs (grip vs lower limb motor performance)	≤0.30	0.63	no
9. There is no or low correlation between the result of the ARAT-PL subtest Pinch and the FMA-LE total score.	ARAT-PL and FMA-LE measure unrelated constructs (pinch grip vs lower limb motor performance)	≤0.30	0.58	no
10. There is no or low correlation between the result of the ARAT-PL subtest Gross Movement and the total result of FMA-LE	ARAT-PL and FMA-LE measure unrelated constructs (gross movement vs lower limb motor performance)	≤0.30	0.53	no

ARAT-PL – the Polish version of the Action Research Arm Test; FMA-LE – Fugl–Meyer Assessment for Lower Extremity.

Floor and ceiling effects

Floor and ceiling effects were determined as the proportion of answers scoring beyond the lower (floor) and upper (ceiling) boundaries of the total ARAT score (0–57 points). The cutoff points for these boundaries were established at 5%, so that scores under 3 were considered as the floor and those above 54 as the ceiling. Floor and ceiling effects were established if more than 20% of patients fell outside either the set lower or upper boundaries.²¹ The level of significance selected throughout was $p < 0.05$.

Results

Clinical characteristics of the patients

A total of 60 subjects in the subacute stage of stroke participated in the examination, of whom 63.3% had left hemiplegia and 36.7% right hemiplegia. Of the patients, 31.6% were women and 68.4% were men. The mean age was 64 years (range: 31–85 years). The median length of time since stroke was 47 days (range: 22–138 days). Most of the patients were right-handed (93.3%); only 6.7% were ambidextrous.

Table 3. Changes introduced during the whole process of cultural adaptation

Forward translation		Backward translation		Final translation	
original words/ sentences	translated words/ sentences	original words/ sentences	translated words/ sentences	original words/ sentences	translated words/ sentences
Action Research Arm Test	Test assessing function of upper limb	Action Research Arm Test	Research Test of Upper Extremity Action	Research Test of Upper Extremity Action	Test of Upper Extremity Function
there are four subtests	consist of 4 subtests	ordered	arranged	rater	examiner
items in each are ordered	tasks are ordered	top	maximum	passess	execute correctly
zero	0	washer over bolt	bolt washer	no more need to be administered	the rest of the test is skipped
and again no more tests need to be performed in that subtest	and again there is no need to perform additional tasks in that subtest	1 st and 3 rd	index and ring	subject fails	subject does not complete the task
he	subject			no more test need to be performed	and the execution is skipped
wood	wooden			more	further
3 rd finger	ring finger			block, wood	wooden, block
2 nd finger	middle finger			pick up	lifting
1 st finger	index finger			pour water	decanting water
GM	gross movement			place hand	placing hand
grasp	static grip			hand to mouth	touching the mouth with the hand
grip	dynamic grip			pinch	pinch grip
pinch	pinch grip			numbers	addition of the word "points"
gross movement	total movement			grasp	
				static grip	precision grip
				dynamic grip	global movement

Translation and cultural adaptation

Multiple linguistic changes were required in the forward, backward, and final versions of the translation to obtain an ARAT-PL that was as consistent as possible with the original English version (Table 3).

Reliability

The ARAT-PL Grip, Grasp, Pinch, and Gross Movement Test subscale items exhibited almost perfect agreement: The calculated test–retest kappa values ranged from 0.95 to 1.00 (Table 4). The ICC coefficients for the subscales and total instrument score were in a range of 0.99–1.00, indicating excellent reliability (Table 4). The standard measurement error and minimal detectable change for the subscales ranged from 0 to 0.479 and 0 to 1.327 points, respectively; for the total ARAT-PL score calculated for the test–retest measurements (Table 5).

Inter-rater kappa values for the ARAT-PL subscale items ranged from 0.85 to 1.00 (Table 4); they exhibited almost perfect agreement. The ICC (2,k) coefficient values calculated for each subscale and the total score were above 0.99,

showing excellent reliability (Table 5). The intra-observer standard measurement error and minimal detectable change calculated for the subscales and total ARAT-PL score ranged from 0.112 to 0.335 and 0.310 to 0.930, respectively (Table 5).

Internal consistency

The total ARAT-PL score exhibited excellent internal consistency with a Cronbach's α value of 0.99. Similarly, the Cronbach's α values for the grasp, grip, pinch, and gross movement items amounted to 0.99, 0.98, 0.97, and 0.99, respectively. The Cronbach's α value for the FMA-UE was 0.93, also indicating excellent internal consistency.

Validity

Scatter plots of the correlations between the ARAT-PL and FMA scores are shown in Fig. 1. There were high correlations between ARAT-PL and FMA-UE absolute scores, and between all ARAT-PL subscale scores and FMA-UE absolute scores (Table 1). There were moderate correlations

Table 4. Kappa and percent agreement values for the test–retest and inter-rater reliability (n = 60)

Item		Test–retest		Inter-rater	
		κ	PA (%)	κ	PA (%)
Grasp	Block, wood, 10 cm cube	0.97	98.33	1.00	100
	Block, wood 2.5 cm cube	0.97	98.33	0.97	98.33
	Block, wood 5 cm cube	0.97	98.33	1.00	100
	Block, wood 7.5 cm cube	0.97	98.33	1.00	100
	Ball (cricket), 7.5 cm diameter	1.00	100	0.94	96.67
	Stone 10 × 2.5 × 1 cm	1.00	100	1.00	100
Grip	Pour water from glass to glass	1.00	100	1.00	100
	Tube 2.25 cm	0.97	98.33	0.94	96.67
	Tube 1 × 16 cm	0.97	98.33	0.97	98.33
	Washer (diameter: 3.5 cm) over bolt	1.00	100	1.00	100
Pinch	Ball bearing, 6 mm 3 rd finger and thumb	0.95	96.67	1.00	100
	Marble, 1.5 cm index finger and thumb	0.97	98.33	0.91	95.00
	Ball bearing 2 nd finger and thumb	1.00	100	0.95	96.67
	Ball bearing 1 st finger and thumb	0.91	95.00	0.97	98.33
	Marble 3 rd finger and thumb	1.00	100	1.00	100
	Marble 2 nd finger and thumb	0.97	98.33	1.00	100
Gross movement	Place hand behind head	1.00	100	0.96	98.33
	Place hand on top of head	1.00	100	1.00	100
	Hand to mouth	1.00	100	0.85	95.00

κ – kappa value; PA – percent agreement.

Table 5. Test–retest reliability (n = 60)

Subtest scores (points)	Lower 95% CI	ICC	95% CI	SEM	MDC
Test–retest reliability					
Grasp (0–18)	0.997	0.998	0.999	0.258	0.716
Grip (0–12)	0.999	0.999	0.999	0.129	0.358
Pinch (0–18)	0.997	0.999	0.999	0.214	0.593
Gross movement (0–9)	1.000	1.000	1.000	0.000	0.000
Total ARAT-PL (0–57)	0.999	0.999	0.996	0.479	1.327
Inter-rater reliability					
Grasp (0–18)	0.999	0.999	0.999	0.112	0.310
Grip (0–12)	0.998	0.999	0.999	0.144	0.400
Pinch (0–18)	0.999	0.999	0.999	0.129	0.358
Gross movement (0–9)	0.997	0.998	0.998	0.129	0.358
Total ARAT-PL (0–57)	0.999	0.999	0.999	0.335	0.930

95% CI – 95% confidence interval; ICC – intraclass correlation coefficient; SEM – standard measurement error; MDC – minimal detectable change; ARAT-PL – Polish version of the Action Research Arm Test.

between the ARAT-PL and FMA-LE absolute scores, and between all ARAT-PL subscales' scores and FMA-LE absolute scores (Table 2). Results for the hypotheses testing correlations are shown in Table 1 for the associations with FMA-UE and Table 2 for the associations with FMA-LE. Based on the absolute scoring method, ARAT-PL has 5 out of 10 hypotheses confirmed (50%), indicating moderate construct validity (Tables 1,2).

Floor and ceiling effects

The Polish version of ARAT had a significant ceiling effect, spanning 48% of tested patients, but no floor effect (12% of patients). It has been demonstrated that both FMA-UE and FMA-LE have significant ceiling effects (50% and 30% of patients, respectively) but no floor effect (0% of patients).

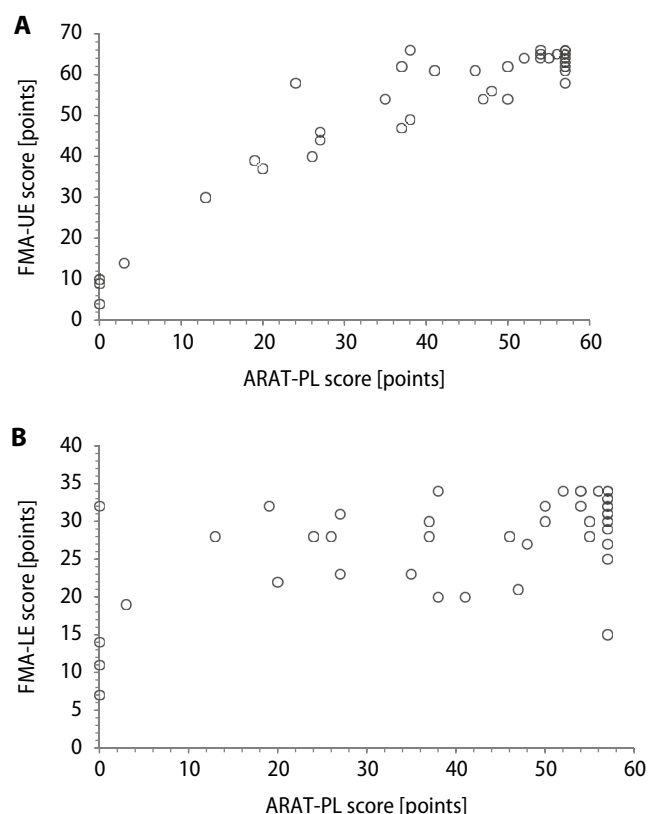


Fig. 1. The relationship between the scores of the ARAT-PL and FMA-UE (A) and the ARAT-PL and FMA-LE (B). See results in Table 1 and Table 2

ARAT-PL – the Polish version of the Action Research Arm Test;
FMA-UE – Fugl-Meyer Assessment for Upper Extremity;
FMA-LE – Fugl-Meyer Assessment for Lower Extremity.

Discussion

This is probably the first reported cross-cultural translation and adaptation based on rigorous methodology and strict regulation of this process. This study assessed the reliability and construct validity of a Polish version of ARAT. The hypotheses tested to evaluate construct validity showed that ARAT-PL had excellent reliability and moderate construct validity.

Therefore, this result provides an official, transculturally validated ARAT for wide and consistent clinical use across Poland, and for research across the world.

Reliability

The total scores and sub-scores of ARAT-PL showed excellent inter-rater and test–retest reliability. This agrees with the results of previous studies, which have reported ICC coefficients of 0.98 and 0.99 for inter-rater reliability^{13,40} and test–retest reliability²¹ in poststroke hemiparetic patients. Moreover, the agreement for individual ARAT-PL items assessed with Cohen's kappa coefficients was almost perfect, and the interobserver agreement measured via the percentage agreement was ≥ 90 . The latter result is even higher than reported in another study, which

found percentage agreement ≥ 70 .²⁰ Therefore, our study has shown that ARAT-PL has excellent reliability, comparable to the original scale.

Minimum detectable change and measurement error

The values for standard error of measurement and minimal detectable change were 0.34 and 0.93 for inter-rater, and 0.48 and 1.33 for test–retest measurements. Similar comparisons in past studies have shown higher values. One example produced standard error of measurement and minimal detectable change values for the test–retest assessment of post-stroke patients with ARAT of 1.3 and 3.5, respectively.⁴⁰ Another study reported minimal detectable change values of 13.1 and 3.5 for inter-rater and test–retest measurements performed with ARAT.¹³ The minimum detectable change captures the amount of change that must be observed in order to exceed measurement error, for assessments administered by the same or by different observers. The results suggest that ARAT-PL can produce very reliable data in subacute stroke patients, both across multiple sessions by the same experienced rater and for measurements performed by 2 different experienced raters.

Internal consistency

The Polish version of ARAT showed excellent internal consistency for both the total and subscale scores ($\alpha = 0.97$ – 0.99). These results are consistent with previous studies, which have reported excellent internal consistency for the original ARAT ($\alpha \geq 0.98$)^{39,16} and for the Chinese version ($\alpha = 0.98$)²¹ in subacute and chronic stroke patients. Our results show that the particular items of ARAT-PL have been well translated into Polish; this version is highly consistent with the original and other foreign adaptations.

Validity

This study found high correlation ($r = 0.71$ – 0.83) between the total and subtest scores of ARAT-PL and the total score of FMA-UE-PL in subacute post-stroke patients. These results agree with other studies, one of which indicated coefficients of 0.77 within 72 h of patient admission to the rehabilitation unit, and 0.87 in the 24 h before discharge.¹⁹ Another reported coefficients in the range of 0.71–0.74 for correlations between ARAT and FMA-UE in chronic patients with stroke.^{40,41} However, a further study found slightly higher correlation coefficients of 0.91 after 2 weeks and 0.94 8 weeks after stroke onset^{18,42} for the original ARAT and FMA-UE. Higher coefficient values of 0.90, 0.90, 0.82, and 0.92 have also been demonstrated for correlations between ARAT and FMA-UE performed 14, 30, 90, and 180 days after stroke, respectively.⁴⁰ However, the latter study had a smaller sample. Lastly, Wei et al.⁴³ found somewhat higher coefficient values of 0.93. However, they evaluated chronic stroke subjects before and after

upper-extremity rehabilitation robotic training. It seems that the strength of interdependence between ARAT and FMA-UE may be affected by many different factors, including 1) the size of the study sample, 2) the time of the administration of outcome measures after stroke, 3) the type of rehabilitation therapy to which studied subjects are subjected, and 4) translation-related differences between versions of the same instrument. Both ARAT and FMA-UE evaluate the degree of impairment of the upper limbs in patients with stroke. However, ARAT assesses the functioning of upper extremities using observational methods, while the FMA measures motor impairment. Therefore, collectively, these studies show that the ARAT score may effectively assess not only function, but also indirectly some motor impairment of the upper extremity.

Compared to the FMA, ARAT has a smart scoring system. Subjects with both severe and minor upper limb dysfunction may get minimum or maximum scores, and then no more tests need to be administered for them to receive a score for that subtest. This shortens the total time of evaluation. The advantage of ARAT is that it can very precisely evaluate hand movements and indicate the specific functional problem of the extremity, even if the patient seems to be in generally good functional shape. Our results show that ARAT is an appropriate tool for assessing people with moderate-to-severe stroke.

Floor and ceiling effects

We did observe a significant ceiling effect of ARAT-PL. The studied patients were in a range of 22–138 days after recovery from stroke. It was perhaps possible for many patients who had had minor strokes and longer histories of recovery, and had reached high functional status, to gain the highest scores in the ARAT-PL. Therefore, it seems that ARAT is a less useful outcome measure for people who substantially recover from stroke. For example, in cases of mild stroke we did not observe difficulties with completing the specific tasks; the only exception was the ability to pinch a marble with the 3rd finger and thumb. Therefore, a relatively large number of patients with mild stroke achieved maximum points. This may suggest that the scoring system of ARAT is not well designed for people with mild upper limb dysfunctions. In parallel, we observed a significant ceiling effect for FMA-UE; 50% of patients had total scores ≥ 64 . However, no floor effect was demonstrated. Hence, as with ARAT-PL, half the patients had near the maximal FMA-UE score. The FMA assesses some additional skills, such as movement coordination or reflex activity, and requires greater mobility skills than ARAT. Again, this shows that many of the studied patients had recovered well from stroke; for such patients, FMA-UE is not a challenging evaluation. The consistency between the results with ARAT and FMA-UE also suggests that recovery in movement coordination and muscle reflex activity is paralleled by upper extremity functional independence in subacute patients with stroke.^{17,44}

Limitations

The main limitations of the study were differences in rehabilitation protocol and in time of recovery after stroke; these might have affected the sample homogeneity. However, at the time of the study, we had limited access to a more homogenous group of stroke survivors. Future research with ARAT-PL and FMA-UE should separately analyze patients in the acute or chronic stage of stroke to improve the conditions of observational studies aimed at determining the interdependence of particular outcome measures. To show the construct validity of ARAT, we examined correlations with FMA-LE, finding a significant but lower correlation coefficient (0.59) as compared to the FMA-UE (0.83) for the total score relationship. This may falsely indicate that the level of upper extremity function was moderately related to the level of lower extremity function.

Conclusions

It can be concluded that ARAT-PL is a reliable and valid tool for assessing upper extremity function in subacute stroke survivors. Its only drawback is that it appears to have a ceiling effect, limiting the differentiation of patients with mild upper limb impairment after stroke. Despite this, our results support the clinical and research use of ARAT-PL in the Polish population of patients with stroke.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

ORCID iDs

Joanna Małecka  <https://orcid.org/0000-0002-3341-7385>
 Magdalena Goliwás  <https://orcid.org/0000-0002-3661-654X>
 Katarzyna Adamczewska  <https://orcid.org/0000-0002-8063-1253>
 Jacek Lewandowski  <https://orcid.org/0000-0002-3417-2084>
 Dawid Łochyński  <https://orcid.org/0000-0002-4274-0950>

References

1. Krakauer JW. Arm function after stroke: From physiology to recovery. *Semin Neurol.* 2005;25(4):384–395. doi:10.1055/s-2005-923533
2. Meng G, Meng X, Tan Y, et al. Short-term efficacy of hand-arm bimanual intensive training on upper arm function in acute stroke patients: A randomized controlled trial. *Front Neurol.* 2018;8:726. doi:10.3389/fneur.2017.00726
3. Verheyden G, Nieuwboer A, De Wit L, et al. Time course of trunk, arm, leg, and functional recovery after ischemic stroke. *Neurorehabil Neural Repair.* 2008;22(2):173–179. doi:10.1177/1545968307305456
4. Koh CL, Hsueh IP, Wang WC, et al. Validation of the action research arm test using item response theory in patients after stroke. *J Rehabil Med.* 2006;38(6):375–380. doi:10.1080/16501970600803252

5. Jonsdottir J, Thorsen R, Aprile I, et al. Arm rehabilitation in post stroke subjects: A randomized controlled trial on the efficacy of myoelectrically driven FES applied in a task-oriented approach. *PLoS One*. 2017;12(12):e0188642. doi:10.1371/journal.pone.0188642
6. Nakayama H, Stig Jørgensen H, Otto Raaschou H, Skyhøj Olsen T. Recovery of upper extremity function in stroke patients: The Copenhagen stroke study. *Arch Phys Med Rehabil*. 1994;75(4):394–398. doi:10.1016/0003-9993(94)90161-9
7. Hsueh IP, Hsieh CL. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. *Clin Rehabil*. 2002;16(6):617–624. doi:10.1191/0269215502cr530oa
8. Barreca SR, Stratford PW, Lambert CL, Masters LM, Streiner DL. Test-retest reliability, validity, and sensitivity of the Chedoke Arm and Hand Activity Inventory: A new measure of upper-limb function for survivors of stroke. *Arch Phys Med Rehabil*. 2005;86(8):1616–1622. doi:10.1016/j.apmr.2005.03.017
9. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research: *Int J Rehabil Res*. 1981;4(4):483–492. doi:10.1097/00004356-198112000-00001
10. Carroll D. A quantitative test of upper extremity function. *J Chronic Dis*. 1965;18(5):479–491. doi:10.1016/0021-9681(65)90030-5
11. McDonnell M. Action Research Arm Test. *Aust J Physiother*. 2008;54(3):220. doi:10.1016/S0004-9514(08)70034-5
12. 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 Phys Med Rehabil*. 2001;82(1):14–19. doi:10.1053/apmr.2001.18668
13. 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):107–113. doi:10.1093/ageing/27.2.107
14. Lin JH, Hsu MJ, Sheu CF, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Phys Ther*. 2009;89(8):840–850. doi:10.2522/ptj.20080285
15. Chen HF, Lin KC, Chen CL. Rasch validation and predictive validity of the Action Research Arm Test in patients receiving stroke rehabilitation. *Arch Phys Med Rehabil*. 2012;93(6):1039–1045. doi:10.1016/j.apmr.2011.11.033
16. Van Wegen E, Nijland R, 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):694–696. doi:10.2340/16501977-0560
17. Yozbatiran N, Der-Yeghian L, Cramer SC. A standardized approach to performing the Action Research Arm Test. *Neurorehabil Neural Repair*. 2008;22(1):78–90. doi:10.1177/1545968307305353
18. Page SJ, Hade E, Persch A. Psychometrics of the Wrist Stability and Hand Mobility Subscales of the Fugl–Meyer Assessment in moderately impaired stroke. *Phys Ther*. 2015;95(1):103–108. doi:10.2522/ptj.20130235
19. 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 Phys Med Rehabil*. 2006;87(7):962–966. doi:10.1016/j.apmr.2006.02.036
20. Nordin Å, Murphy M, Danielsson A. Intra-rater and inter-rater reliability at the item level of the Action Research Arm Test for patients with stroke. *J Rehabil Med*. 2014;46(8):738–745. doi:10.2340/16501977-1831
21. Zhao JL, Chen PM, Li WF, et al. Translation and initial validation of the Chinese version of the Action Research Arm Test in people with stroke. *Biomed Res Int*. 2019;2019:5416560. doi:10.1155/2019/5416560
22. Doussoulain SA, Rivas RS, Campos RV. Validation of “Action Research Arm Test” (ARAT) in Chilean patients with a paretic upper limb after a stroke [in Spanish]. *Rev Med Chile*. 2012;140:59–65.
23. Goliwas M, Małecka J, Lewandowski J, Kamińska E, Adamczewska K, Kocur P. Analysis of dependencies between Fugl–Meyer Assessment Scale test and Berg Balance Scale test as an assessment of the increased muscle tone in chronic-phase patients after a ischemic stroke. *Med Rehabil*. 2022;26(1):4–9. doi:10.5604/01.3001.0015.8241
24. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417–1432. doi:10.1016/0895-4356(93)90142-N
25. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–3191. doi:10.1097/00007632-200012150-00014
26. Szczechowicz J, Lewandowski J, Sikorski J. Polish adaptation and validation of Burn Specific Health Scale – Brief. *Burns*. 2014;40(5):1013–1018. doi:10.1016/j.burns.2013.11.026
27. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157. doi:10.1007/s11136-018-1798-3
28. Salter K, Jutai J, Teasell R, Foley N, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disabil Rehabil*. 2005;27(6):315–340. doi:10.1080/09638280400008545
29. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–745. doi:10.1016/j.jclinepi.2010.02.006
30. Cecchi F, Carrabba C, Bertolucci F, et al. Transcultural translation and validation of Fugl–Meyer assessment to Italian. *Disabil Rehabil*. 2021;43(25):3717–3722. doi:10.1080/09638288.2020.1746844
31. 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(3):232–240. doi:10.1177/154596802401105171
32. Rech KD, Salazar AP, Marchese RR, Schifino G, Cimolin V, Pagnussat AS. Fugl–Meyer Assessment scores are related with kinematic measures in people with chronic hemiparesis after stroke. *J Stroke Cerebrovasc Dis*. 2020;29(1):104463. doi:10.1016/j.jstrokecerebrovasdis.2019.104463
33. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. R package v. 2.1.3. Evanston, USA: Northwestern University; 2021. <https://cran.r-project.org/web/packages/psych/index.html>. Accessed August 15, 2023.
34. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil*. 2000;81(12 Suppl 2):S15–S20. doi:10.1053/apmr.2000.20619
35. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York, USA: McGraw-Hill; 1994. ISBN:978-0-07-047849-7.
36. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropract Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
37. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282. doi:10.11613/BM.2012.031
38. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. 2nd ed. Boston, USA: Houghton Mifflin; 1988. ISBN:978-0-395-36911-1.
39. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175–191. doi:10.3758/BF03193146
40. Simpson LA, Eng JJ. Functional recovery following stroke: Capturing changes in upper-extremity function. *Neurorehabil Neural Repair*. 2013;27(3):240–250. doi:10.1177/1545968312461719
41. Hsieh YW, Wu CY, Lin KC, Chang YF, Chen CL, Liu JS. Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. *Stroke*. 2009;40(4):1386–1391. doi:10.1161/STROKEAHA.108.530584
42. De Weerd WJG. Measuring recovery of arm-hand function in stroke patients: A comparison of the Brunnstrom–Fugl–Meyer test and the Action Research Arm test. *Physiother Can*. 1985;37(2):65–70. doi:10.3138/ptc.37.2.065
43. Wei XJ, Tong KY, Hu XL. The responsiveness and correlation between Fugl–Meyer Assessment, Motor Status Scale, and the Action Research Arm Test in chronic stroke with upper-extremity rehabilitation robotic training. *Int J Rehabil Res*. 2011;34(4):349–356. doi:10.1097/MRR.0b013e32834d330a
44. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl–Meyer assessment and its implications for clinical trials. *Neurorehabil Neural Repair*. 2013;27(8):732–741. doi:10.1177/1545968313491000