



COMPARISON OF SELF-REPORTED VS OBSERVATIONAL CLINICAL MEASURES OF IMPROVEMENT IN UPPER LIMB CAPACITY IN PATIENTS AFTER STROKE

Eline C. C. VAN LIESHOUT, MSc^{1,2}, Johanna M. A. VISSER-MEILY, MD, PhD^{2,3}, Rinske H. M. NIJLAND, PhD⁴, Rick M. DIJKHUIZEN, PhD¹ and Gert KWAKKEL, PhD⁴⁻⁶

From the ¹Biomedical MR Imaging and Spectroscopy Group, Center for Image Sciences, University Medical Center Utrecht and Utrecht University, ²Center of Excellence for Rehabilitation Medicine, UMC Utrecht Brain Center, University Medical Center Utrecht and Utrecht University, ³Department of Rehabilitation, Physical Therapy Science and Sports, UMC Utrecht Brain Center, University Medical Center Utrecht and Utrecht University, Utrecht, ⁴Department of Neurorehabilitation, Amsterdam Rehabilitation Research Centre, Reade, Amsterdam, the Netherlands ⁵Amsterdam university Medical Centre, location VUmc, Department of Rehabilitation Medicine, Amsterdam Neurosciences and Amsterdam Movement Sciences, Amsterdam, The Netherlands and ⁶Department of Physical Therapy and Human Movement Sciences, Northwestern University, Chicago, IL, USA

Objective: Recovery of the paretic arm post-stroke can be assessed using observational and self-reported measures. The aim of this study was to determine whether the correspondence (match) or non-correspondence (mismatch) between observational and self-reported improvements in upper limb capacity are significantly different at 0–3 months compared with 3–6 months post-stroke.

Methods: A total of 159 patients with ischaemic stroke with upper limb paresis were included in the study. Recovery of arm capacity was measured with observational (Action Research Arm Test; ARAT) and self-reported measures (Motor Activity Log Quality of Movement; MAL-QOM and Stroke Impact Scale Hand; SIS-Hand) at 0–3 and 3–6 months post-stroke. The proportion of matches was defined (contingency tables and Fisher's exact test) and compared across the different time-windows using McNemar's test.

Results: The proportion of matches was not significantly different at 0–3 months compared with 3–6 months post-stroke for the ARAT vs MAL-QOM and SIS-Hand (all $p > 0.05$). In case of mismatches, patients' self-reports were more often pessimistic (86%) in the first 3 months post-stroke compared with the subsequent 3 months (39%).

Conclusion: The match between observational and self-reported measures of upper limb capacity is not dependent on the timing of assessment post-stroke. Assessment of both observational and self-reported measures may help to recognize possible over- or under-estimation of improvement in upper limb capacity post-stroke.

Key words: stroke; upper limb outcome; motor function recovery; patient-reported outcome measures; activities of daily living.

Accept Feb 21, 2020; Epub ahead of print Mar 3, 2020

J Rehabil Med 2020; 52: jrm00051

Correspondence address: Rinske Nijland, Reade, locatie Overtoom, Postbus 58271, 1040 HG Amsterdam, The Netherlands. E-mail: R.Nijland@reade.nl

Upper limb paresis is common after stroke and reduces a patient's independence, performance of activities of daily living (ADL) and self-reported quality

LAY ABSTRACT

One of the most common motor disturbances after stroke is a paretic arm, which may be of little functional use in activities of daily living. Recovery of the paretic arm can be assessed by a clinician (observational) or by the patient (patient-reported). It might be expected that observational and patient-reported measures will be strongly related to each other. The aim of this study was to determine whether the correspondence (matches) between those measures is different at 0–3 months post-stroke compared with 3–6 months post-stroke. The results showed that the time-frame post-stroke (0–3 or 3–6 months) did not seem to influence the correspondence between the observational and self-reported measures: there were more matches than mismatches found. Self-reported measures can be used in addition to observational measures to assess arm recovery. Information on the ability and use of the affected arm outside the treatment setting is valuable for clinicians, as it provides more insight into the patients' perspective.

of life (1). Different instruments can be used to assess the upper limb after stroke, based on the levels of function, activity (capacity and performance) and participation (World Health Organization's International Classification of Functioning, Disability and Health (WHO-ICF)) (2). Upper limb function and capacity can be scored by a clinician using standardized validated measurements. However, a number of these measurements have floor and ceiling effects. In addition, self-perceived difficulties with arm use are not reflected by these clinical performance tests (3). Self-reported upper limb outcome measures, however, require subjective assessment of arm functioning at the activity and participation level, as perceived by patients themselves (2, 4).

Recently, self-reported outcome measures (i.e. patient-reported outcome measures (PROMs)) have received increasing attention in mapping patients' perceived recovery in medical practice (5). PROMs can provide valuable insights into a patient's status outside the treatment setting, and can detect change in a patient's perceived health status (6–8). However, the use of PROMS for stroke patients may be affected by confounding factors, such as neglect, self-

awareness, mood, fatigue, social support, relationships, and encouragement from others, which may influence patients' expectations, and might lead to under- or over-estimations in self-reported assessments (9–13).

Moderate correlations have been shown between capacity measures and self-reported measures (14, 15). To date, only a few cross-sectional studies have investigated the correspondence (match) and non-correspondence (mismatch) between outcomes on observational and self-reported upper limb measures (10, 12, 16–18). As the time course of upper limb recovery is non-linear and is driven by poorly understood processes of neural recovery and compensation strategies, outcomes between observational and self-reported measures may deviate, depending on the timing of assessment post-stroke.

Therefore, the main aims of the present study were: (i) to determine whether the proportion of matches between observational and self-reported improvements on upper limb capacity differs between the early (0–3 months) and late subacute stages 3–6 months; and (ii) to identify whether the self-reported improvements may under- or over-estimate the observationally measured improvements in upper limb capacity at these stages.

For the first aim, it was hypothesized that there would be a higher proportion of matches at 3–6 months than at 0–3 months post-stroke, because patients gradually learn to deal better with and gain more experience of real-world limitations, with a better understanding of their own capabilities (19–21). For the second aim, it was hypothesized that more patients would overestimate (“optimistic”) than underestimate (“pessimistic”) their self-reported capacity at 0–3 months post-stroke than at 3–6 months post-stroke, because spontaneous neurological and functional recovery occurs within the first 3 months post-stroke. This can result in more subjectively experienced improvement than observationally measured improvement (22, 23).

METHODS

Data for the current study were collected during the EXPLICIT (EXplaining PLasticity) Stroke trial (24). The EXPLICIT Stroke trial was a multicentre, observer-blinded randomized controlled trial (RCT) to investigate the effects of modified constraint-induced movement therapy (mCIMT) and electromyography-triggered neuromuscular stimulation (EMG-NMS) on upper limb capacity. Eligible patients were screened and included in the first week post-stroke. The included patients had an upper limb paresis and were stratified into a poor (EMG-NMS) and favourable prognosis group (mCIMT) for upper limb recovery. Full details about randomization, treatment, and study design can be found elsewhere (24). Baseline assessments were performed within 2 weeks post-stroke. The data used in this study were taken from baseline, 12 and 26 weeks after stroke onset.

The EXPLICIT Stroke trial (24) was approved by the Medical Ethical Reviewing Committees of Leiden University Medical Centre (main reviewing committee: Dutch Central Commit-

tee on Research Involving Human Subjects, CCMO, protocol number NL21396.058.08), VU Medical Centre Amsterdam, Radboud University Medical Centre Nijmegen, and University Medical Centre Utrecht in the Netherlands. This trial is registered in the Netherlands Trial Register (NTR, <http://www.trialregister.nl>, NL1366).

Participants

All included patients met the following criteria: (i) first-ever, ischaemic stroke in one of the cerebral hemispheres; (ii) upper limb paresis according to National Institutes of Health Stroke Scale (NIHSS) item 5; (iii) baseline ARAT score ≤ 53 on a maximum of 57 points; (iv) ability to communicate and comprehend (Mini Mental State Examination ≥ 23 points on a maximum of 30 points); (v) ability to sit independently for at least 30 s; (vi) 18–80 years of age; (vii) no successful thrombolysis therapy resulting in upper limb motor recovery and attaining 0 points on NIHSS item 5 of the paretic arm; (viii) no musculoskeletal impairments of the upper paretic limb; (ix) no additional therapies, such as botulinum toxin injections or medication intake that may influence upper limb function in the previous 3 months; (x) willing to participate in an intensive rehabilitation treatment programme; and (xi) written informed consent.

Observational clinical testing

The current study used the ARAT as the observational measure of upper limb capacity (25). Observational measures require an independent assessor trained to measure a patient's skills to perform the tasks in the test. The ARAT assesses the ability to perform gross movements and the ability to grasp, move and release objects of different sizes, weights and shapes (WHO-ICF, Activity level). (2) The items are rated on 4-point scales (0–3), with a maximum score of 57 (best performance) (25). The ARAT is a reliable, valid and responsive test (26) in patients with stroke with mild to moderate motor severity and in the absence of severe cognitive impairment. The minimal clinically important difference (MCID) was set at 6 points, based on clinical experience and estimates, which is approximately 10% of the maximum score (27).

Self-reported testing

Dutch versions of the MAL-QOM and SIS-Hand were used to describe the motor performance from the viewpoint of the patient (WHO-ICF, Activity level).

A Dutch version of the 14-item MAL was used to assess how well (Quality of Movement; QOM scale) the paretic arm was used spontaneously during 14 activities of daily living outside the laboratory. The patient is asked to indicate how well he/she used his/her affected arm during certain activities in the past week (e.g. pick up a glass, comb your hair, button a shirt). A 6-point ordinal scale (range 0–5) was used, in which half ratings can also be given. A higher score indicates better performance: maximum score: 5 (transformed scale: overall score (0–70) divided by 14, resulting in a 0–5 scale). This 6-point scale contains scoring from “The weaker arm was not used at all for that activity (never)”, to “The ability to use the weaker arm for that activity was as good as before the stroke (normal)”. Reliability and validity of the MAL has been shown (14). A MCID of 0.5 points was used, based on clinical experience and estimates, which reflects 10% of the maximum score (14). Version 3.0 of the SIS is a stroke-specific, self-report, health status measure containing 8 domains related

to hand function, strength, activities of daily living, communication, emotion, memory and thinking. The SIS is a valid and reliable measure for a diverse group of stroke survivors (28). The Hand domain of the SIS consists of 5 questions and each item is scored on a 5-point Likert scale (transformed from 5 to 25 to a scale from 0 to 100). In the questions patients must rate how difficult it was to use their affected hand in a range of activities in the past 2 weeks (e.g. turn a doorknob, tie a shoelace). Higher scores indicate a low(er) impact of hand problems on health and life. A MCID of 10 points was used, based on clinical experience and estimates, i.e. 10% of the maximum score (29).

Data analysis

To calculate change scores for the time-window 0–3 months post-stroke, the baseline scores from the ARAT, MAL-QOM and SIS-Hand were subtracted from the follow-up scores at 3 months. Some of the patients had a baseline score less than a MCID short of the maximum score on one of the outcome measures. Therefore, reaching the maximum score at the follow-up measurement 3 months post-stroke was considered a clinically meaningful change. Change scores for the time-window 3–6 months post-stroke were computed by subtracting the follow-up scores measured at 3 months post-stroke from the follow-up scores 6 months post-stroke. Changes were marked as successful when maximum scores were reached at follow-up, or when changes were beyond the known MCID (10% of the maximum score). Change scores smaller than the MCID were marked as unsuccessful. Subsequently, patients with a successful change (improvement) on the ARAT, as well on a self-reported measure (MAL-QOM and SIS-Hand), and patients with unsuccessful changes (no improvement) on the ARAT and on a self-reported measure (MAL-QOM and SIS-Hand) were grouped as matchers (i.e. true positives). Patients with a successful change on the ARAT, but not on a self-reported measure (MAL-QOM and SIS-Hand), and vice versa, were grouped as mismatches (i.e. true negatives).

Fisher's exact test was used to examine the significance of the association between matches and mismatches (i.e. overall fraction correct). The tested null hypothesis was that a successful or unsuccessful change on the ARAT is equally likely to have a successful change on the MAL-QOM or SIS-Hand. The percentage of false-negatives reflects the degree of underestimation, i.e. observed change on the ARAT without reported change on the MAL-QOM or SIS-Hand. The percentage of false-positives reflects the degree of overestimations, i.e. reported change on the MAL-QOM or SIS-Hand without observed change on the ARAT. The percentages of false-positives and false-negatives were deduced from the contingency tables. In addition, using

Table I. Patients' characteristics at baseline

Characteristics	Baseline	3 months	6 months
Females/males, <i>n</i>	63/96		
Age, years, mean (SD)	60.0 (12.3)		
Affected hemisphere, right/left, <i>n</i>	105/54		
Affected dominant side, <i>n</i>	54		
Handedness, right/left, <i>n</i>	139/17		
Time between stroke onset and baseline assessment, days, mean (SD)	8.3 (4.1)		
Level of education (low/high), <i>n</i>	113/28		
Partner (yes/no), <i>n</i>	111/48		
Mood (below/above normative score), <i>n</i>	89/45		
Dexterity (minimal/some), <i>n</i>	118/41		
MI (0–100), mean (SD)	30.3 (30.0)		
FMA (0–66), mean (SD)	18.6 (18.9)		
ARAT (0–57), mean (SD)	8.6 (13.4)	24.6 (22.3)	27.6 (23.1)
SIS-Hand (0–100), mean (SD)	5.9 (2.8)	36.2 (40.1)	42.8 (42.3)
MAL-QOM (0–5), mean (SD)	0.3 (0.6)	1.4 (1.5)	1.6 (1.6)

MI: Motricity Index; ARAT: Action Research Arm Test; FMA: Fugl-Meyer assessment of the arm; SIS-Hand: Stroke Impact Scale; MAL-QOM: Motor Activity Log-Quality of Movement; SD: standard deviation; *n*: total number of patients.

the 2-way contingency tables, the sensitivity, specificity, and positive and negative predicted values (i.e. the probability that an event is present/not present, when the event is present/absent), and, overall fraction correct (i.e. the probability that an event is correctly classified) were also calculated.

Finally, McNemar's test was used to compare the proportions of matches to mismatches between 0–3 and 3–6 months post-stroke, and the association between the ARAT vs MAL-QOM and SIS-Hand. Only those patients were included from whom data were collected at all time-points. The statistical software SPSS 25.0 (SPSS, Chicago, IL, USA) was used for statistical analysis. The level of statistical significance was set 2-tailed at $p < 0.05$.

RESULTS

Patient characteristics

For the EXPLICIT Stroke trial 159 patients were selected (for flow diagram, see Appendix I) (29). There were no reports of adverse effects from the trial. Table I shows the main characteristics of the included patients at baseline, and mean scores on the used outcome measures for different time-points post-stroke. Missing data-points from patients on one of the time-points resulted in a lower number of total patients in the analyses.

Table II. Contingency table of matches and mismatches on objective vs self-reported outcomes, 0–3 months post-stroke

	ARAT			Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	OFC % (95% CI)	FET <i>p</i>	
	Yes <i>n</i> (%)	No <i>n</i> (%)	Totals <i>n</i>							
Successful change										
MAL-QOM	Yes	72 (50)	3 (2)	75	0.83 (0.73–0.90)	0.95 (0.86–0.99)	0.96 (0.89–0.99)	0.79 (0.70–0.85)	0.88 (0.81–0.92)	0.000
	No	15 (10)	55 (38)	70						
	Totals	87	58	145						
SIS-Hand	Yes	47 (46)	1 (1)	48	0.82 (0.70–0.91)	0.98 (0.88–1.00)	0.98 (0.87–1.00)	0.81 (0.71–0.88)	0.89 (0.81–0.94)	0.000
	No	10 (10)	43 (43)	53						
	Totals	57	44	101						

SIS-Hand: Stroke Impact Scale; MAL-QOM: Motor Activity Log-Quality of Movement; ARAT: Action Research Arm Test; 95% CI: 95% confidence interval; PPV: positive predictive value; NPV: negative predictive value; OFC: overall fraction correct; FET: Fisher's exact test.

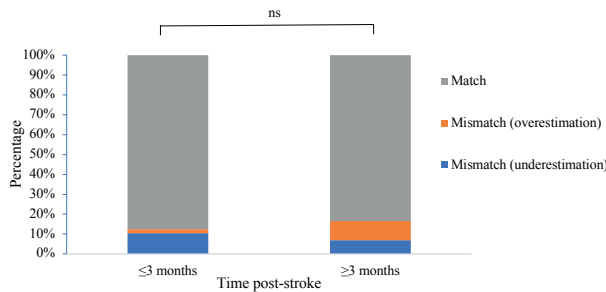


Fig. 1. Proportion of matches to mismatches between Action Research Arm Test (ARAT) and Motor Activity Log-Quality of Movement (MAL-QOM).

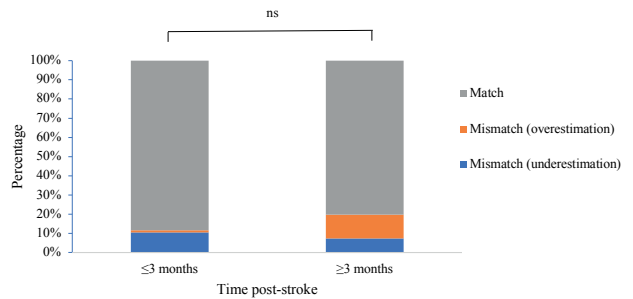


Fig. 2. Proportion of matches to mismatches between Action Research Arm Test (ARAT) and Stroke Impact Scale (SIS-Hand).

Proportion of matches between observational and self-reported measures in the first 6 months post-stroke

For the time-window 0–3 months post-stroke, 88% of the patients showed matches on the ARAT vs MAL-QOM, and 89% on the ARAT vs SIS-Hand (Table II). A successful change on the ARAT is equally likely to have a successful change on the MAL-QOM and SIS-Hand ($p < 0.05$). The sensitivity, specificity, and positive and negative predicted values were comparable for the MAL-QOM and SIS-Hand in comparison with the ARAT. In the time-period 3–6 months post-stroke, a successful change on the ARAT was equally likely to match a successful change on the MAL-QOM and SIS-Hand ($p < 0.05$) (Table III). Eighty-three percent of the patients had a match on the ARAT vs MAL-QOM score, and 81% had a match on the ARAT vs SIS-Hand. The sensitivity, specificity, and positive and negative predicted values were all slightly lower for the ARAT vs SIS-Hand than the ARAT vs MAL-QOM.

False negatives, i.e. underestimations, were measured in 15 out of 18 patients (83%) with mismatches for the ARAT vs MAL-QOM in the time-window 0–3 months post-stroke (Table II, Fig. 1). Three patients (2%) could be classified as false-positives, i.e. overestimations. For the ARAT vs SIS-Hand, underestimations were measured in 10 out of 11 patients (91%). One patient (1%) could be classified as a false-positive (Table II, Fig. 2). False-positives, i.e. overestimations, were more common (14 out of 24 patients: 58%) in the mismatch proportion in the time-window 3–6 months post-stroke

for the ARAT vs MAL-QOM (Table III, Fig. 1). For the ARAT vs SIS-Hand, false-positives were measured in 14 out of 22 patients (64%) (Table III, Fig. 2).

Within the first 3 months and beyond the first 3 months post-stroke, the accuracy (overall fraction correct) was comparable for the MAL-QOM and SIS-Hand in relation to the ARAT (Tables II and III).

Comparison between 0–3 and 3–6 months post-stroke in proportion of matches to mismatches

For the ARAT and MAL-QOM, 3–6 months, the number of matches had decreased from 126 (87.5%) to 121 (84%), which was a non-significant difference ($p = 0.487$) (Fig. 1). For the ARAT and SIS-Hand, the proportion of matches decreased from 88% at 0–3 months post-stroke to 81% at 3–6 months post-stroke, which was a non-significant difference ($p = 0.210$) (Fig. 2). This change was a consequence of 7 matches within 3 months post-stroke changing into mismatches beyond 3 months. The sensitivity, specificity, and positive predicted values, except the negative predicted value, were higher in the time-window 0–3 months post-stroke.

DISCUSSION

These results show that stroke patients had significantly more matches than mismatches between observational and self-reported measures of improvements in the upper

Table III. Contingency table of matches and mismatches on objective vs self-reported outcomes, 3–6 months post-stroke

Successful change	ARAT		Totals	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	OFC % (95% CI)	FET p
	Yes n (%)	No n (%)							
MAL-QOM	Yes	18 (12)	14 (10)	0.64 (0.44–0.81)	0.88 (0.81–0.93)	0.56 (0.42–0.69)	0.91 (0.86–0.95)	0.83 (0.77–0.89)	0.000
	No	10 (7)	104 (71)						
	Totals	28	118						
SIS-Hand	Yes	10 (9)	14 (12)	0.56 (0.31–0.78)	0.85 (0.77–0.92)	0.42 (0.27–0.57)	0.91 (0.86–0.95)	0.81 (0.72–0.87)	0.000
	No	8 (7)	82 (72)						
	Totals	18	96						

SIS-Hand: Stroke Impact Scale; MAL-QOM: Motor Activity Log-Quality of Movement; ARAT: Action Research Arm Test; CI: confidence interval; PPV: positive predictive value; NPV: negative predictive value; OFC: overall fraction correct; FET: Fisher’s exact test.

limb during the first 6 months after stroke, which is in accordance with earlier findings (10, 16, 17, 30). Contrary to our hypothesis, the proportion of matches remained stable between 0–3 months and 3–6 months post-stroke, and is therefore not significantly dependent on the timing of assessment within the subacute stage post-stroke. Patients with mismatches within the first 3 months post-stroke were more likely to underestimate their self-reported performance (86%), whereas between 3 and 6 months they tended to overestimate their self-reported performance (61%) on the MAL-QOM and SIS-Hand domain, compared with their actual improvements on the ARAT.

The significantly high correspondence between observational and self-reported improvements in the upper limb is in line with cross-sectional studies in which significant associations between observational and self-reported measures were found (10, 16, 17, 30). These findings suggest that the patient's perspective is usable in the evaluation of upper limb rehabilitation, which supports patient involvement in rehabilitation as encouraged in patient-centred care.

In line with the present study, van Delden et al. (10) used the MCID of the change scores to determine an improvement in scores, and found a significant discrepancy in the proportions of matches, compared to mismatches, between the ARAT and MAL-QOM, where matches were more prevalent. However, in contrast, they found no significant difference in the proportion of matches and mismatches between the ARAT and SIS-Hand. This may be explained by the differences in severity and timing of assessments of upper limb paresis between both studies. Van Delden et al. (10) only included patients with noticeably preserved motor function (i.e. control of the paretic wrist and fingers) in contrast to the current study, in which 63.5% of patients could not voluntarily extend the thumb and/or 2 or more fingers. Since preserved control in hand function (i.e. finger extension) early post-stroke is a favourable sign for good outcome of the paretic upper limb, more spontaneous motor recovery is expected and needs to be perceived and subsequently quantified by the patient, which can complicate self-reports and may result in mismatches (31). Otherwise, the group of patients that no longer recovers remains stable, which can facilitate self-reports and may result in matches.

Neither of the self-reported measures (MAL-QOM and SIS-Hand) was found to be superior in terms of the number of matches compared with the observational measure (ARAT) during the 6-month period after stroke. However, beyond the first 3 months the sensitivity, specificity, and positive and negative predicted values were slightly lower than within the first 3 months post-stroke. The MAL-QOM and SIS-Hand have not been compared previously (14, 15). However, a possible explanation

for higher sensitivity and specificity values in the first 3 months post-stroke is a higher degree of neuroplasticity early after stroke. Since larger percentage changes are required for a self-reported measure as the MAL-QOM to exceed the measurement error, the sensitivity and specificity values can be lower beyond the first 3 months post-stroke when less recovery is expected (23, 32).

The proportion of matches remained similar between the early subacute phase (i.e. the first 3 months post-stroke) and the late subacute phase (3–6 months post-stroke). In measures of self-reported physical function, response shifts seem to occur (i.e. changes within patients regarding internal standards, values or conceptualization of health-related quality of life) over time post-stroke. "Evaluation-based" items, such as when the patient needs to evaluate their difficulty in task performance, are most susceptible to response shifts (33). The SIS-Hand and MAL-QOM contain evaluation-based items. Although patients who have recalibrated what difficulty means to them, it corresponded with the observationally detected improvements. These findings, while preliminary, provide further support for the use of PROMs in the assessment of upper limb capacity. However, caution must be applied, as the findings might be different for patients with severe communication or cognitive problems. This group of patients was not included in our study.

The mismatches (over- and under-estimations) between observational and self-reported outcome measures could be attributed to different causes. Underestimations of self-reported capacity in the first 3 months may be associated with a more pronounced disturbed self-awareness (20), a limited insight into one's own functioning (19), more negativity-prone thoughts, and lack of information about the rehabilitation phase. Another possibility is that the improved, but affected, upper limb capacity is insufficiently used in daily activities, so that functional recovery is not fully experienced. Overestimations of self-reported capacity beyond the first 3 months may be explained by a less disturbing perception of upper limb impairments or better adaptation to the new situation (33).

Other reasons for mismatches between observational- and self-reported outcome measures might be that standardized testing does not account for complex and stressful real-world situations, in contrast to perceived self-reported outcomes. The reverse conditions may also be possible, where patients adapt to their own environment and use compensatory strategies to manage daily life, despite poorer performance in a single (test) environment (34, 35).

Study limitations

This study has some limitations. Firstly, there was a restricted sample size, and the outcome measures were ar-

bitrarily chosen based on the presence in the EXPLICIT trial (24). Secondly, there is no consensus about the most appropriate methodological method to identify (clinically meaningful) improvement (e.g. MCID values, cut-off scores). We chose to use the MCID values (based on clinical experience and expertise; 10% of the total range of the scale) of the outcome measures to determine if a given improvement between 2 time-points was smaller (unsuccessful change) or larger (successful change) than these values (27). In addition, different methods and algorithms are also used to calculate MCIDs (i.e. distribution-based or anchor-based approaches, clinical experience and expertise). Thirdly, patients with severe communication problems and cognitive deficits were excluded from the EXPLICIT trial. In particular, this group of patients run the risk of inaccurate self-reports, which limits the generalizability of the results (9–11).

Conclusion

Self-reported questionnaires used for monitoring upper limb recovery are accurate compared with observationally measured improvement in the early and late subacute phase after stroke. The current study suggests that the timing of assessment post-stroke does not affect the accuracy of self-reports in the sub-acute stages. Self-reported measures can provide additional insights into the impact of disability on the patient, beyond what is provided by observational measures alone. Self-reported measures in addition to observational measures can help to design optimized rehabilitation strategies for patients who underestimate their capacity (training in the use of the affected hand, positive psychology, self-efficacy, expectation management). For the patients who overestimate their capacity, training in body-image may be warranted.

In order to include patients with severe communication problems or cognitive deficits in PROMS, further research should focus on determining whether alternative self-reported measures and data reported by a proxy are equally as accurate as observational measures.

ACKNOWLEDGEMENTS

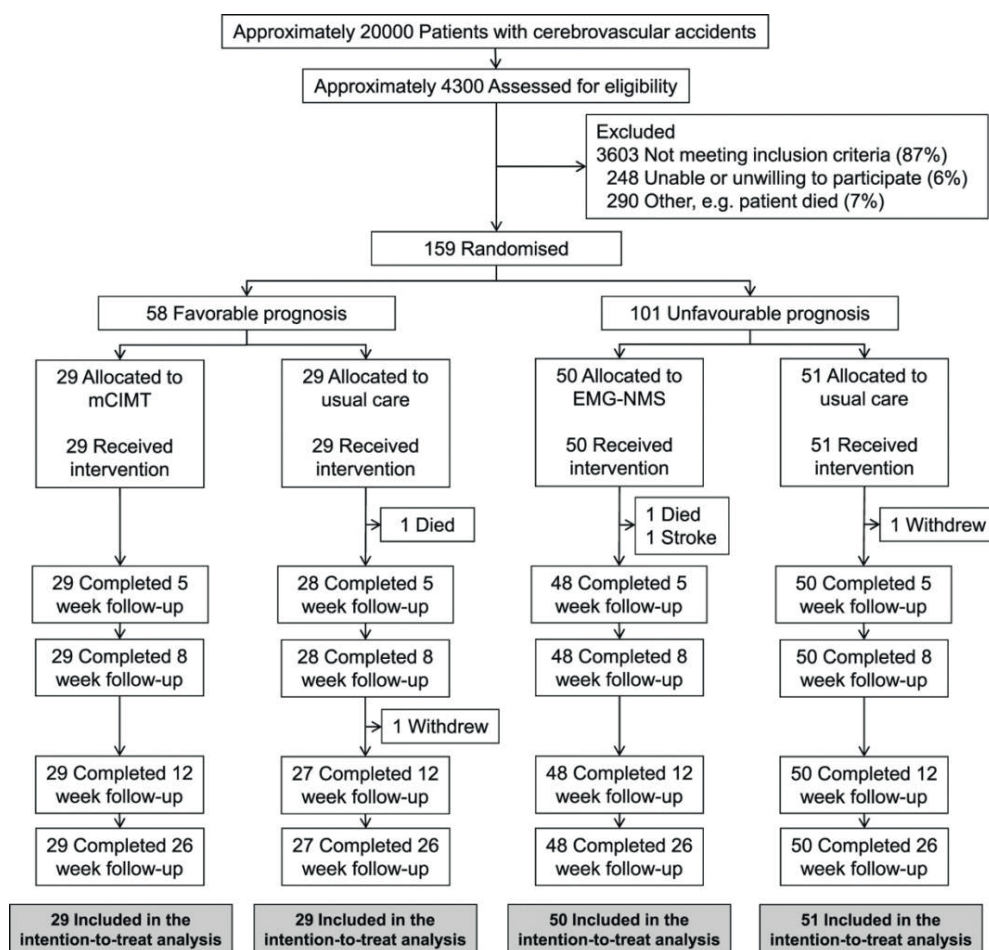
This work was supported by the Netherlands Organization for Scientific Research (VICI 016.130.662).

The authors have no conflicts of interest to report.

REFERENCES

- Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. *Stroke* 2002; 33: 1840–1844.
- World Health Organization (WHO). International Classification of Functioning, Disability and Health. Geneva: World Health Organization; 2003, p. 1–15. Available from: <http://www.who.int/classifications/icf/icfchecklist.pdf?ua=1>.
- Kwakkel G. Standardised measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable (SRRR). *Int J Stroke* 2017; 12: 451–461.
- Ekstrand E, Rylander L, Lexell J, Brogårdh C. Perceived ability to perform daily hand activities after stroke and associated factors: a cross-sectional study. *BMC Neurol* 2016; 16: 208.
- Reeves M, Lisabeth L, Williams L, Katzan I, Kapral M, Deutsch A, et al. Patient-reported outcome measures (PROMs) for acute stroke: rationale, methods and future directions. *Stroke* 2018; 49: 1549–1556.
- Stewart JC, Cramer SC. Patient-reported measures provide unique insights into motor function after stroke. *Stroke* 2013; 44: 1111–1116.
- Uswatte G, Taub E, Morris D, Light K, Thompson PA. The Motor Activity Log-28: assessing daily use of the hemiparetic arm after stroke. *Neurology* 2006; 67: 1189–1194.
- Blennerhassett JM, Avery RM, Carey LM. The test-retest reliability and responsiveness to change for the Hand Function Survey during stroke rehabilitation. *Aust Occup Ther J* 2010; 57: 431–438.
- Coenen M, Kus S, Rudolf KD, Müller G, Berno S, Dereske-witz C, et al. Do patient-reported outcome measures capture functioning aspects and environmental factors important to individuals with injuries or disorders of the hand? *J Hand Ther* 2013; 26: 332–342.
- van Delden AL, Peper CL, Beek PJ, Kwakkel G. Match and mismatch between objective and subjective improvements in upper limb function after stroke. *Disabil Rehabil* 2013; 35: 1961–1967.
- Hochstenbach J, Prigatano G, Mulder T. Patients' and relatives' reports of disturbances 9 months after stroke: subjective changes in physical functioning, cognition, emotion, and behavior. *Arch Phys Med Rehabil* 2005; 86: 1587–1593.
- Persson HC, Danielsson A, Sunnerhagen KS. A cross sectional study of upper extremity strength ten days after a stroke; relationship between patient-reported and objective measures. *BMC Neurol* 2015; 15: 178.
- Essers B, Meyer S, De Bruyn N, Van Gils A, Boccuni L, Tedesco Triccas L, et al. Mismatch between observed and perceived upper limb function: an eye-catching phenomenon after stroke. *Disabil Rehabil* 2019; 41: 1545–1551.
- Van Der Lee JH, Beckerman H, Knol DL, De Vet HCW, Bouter LM. Clinimetric properties of the motor activity log for the assessment of arm use in hemiparetic patients. *Stroke* 2004; 35: 1410–1404.
- Beebe JA, Lang CE. Relationships and responsiveness of six upper extremity function tests during the first 6 months of recovery after stroke. *J Neurol Phys Ther* 2009; 33: 96–103.
- Dromerick AW, Lang CE, Birkenmeier R, Hahn MG, Sahrman SA, Edwards DF. Relationships between upper-limb functional limitation and self-reported disability 3 months after stroke. *J Rehabil Res Dev* 2006; 43: 401.
- van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Devillé WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients. *Stroke* 1999; 30: 2369–2375.
- Wu C, Chuang L, Lin K, Chen H, Tsay P. Randomized trial of distributed constraint-induced therapy versus bilateral arm training for the rehabilitation of upper-limb motor control and function after stroke. *Neurorehabil Neural Repair* 2011; 25: 130–139.
- Zawadzka E, Domańska Ł. Assessment of select dimensions of patients' emotional functioning at different time periods after stroke. *Appl Neuropsychol* 2014; 21: 87–93.
- Prigatano GP, Johnson SC. The three vectors of consciousness and their disturbances after brain injury. *Neuropsychol Rehabil* 2003; 13: 13–29.
- Cooper CL, Phillips LH, Johnston M, Whyte M, Macleod MJ. The role of emotion regulation on social participation following stroke. *Br J Clin Psychol* 2015; 54: 181–199.
- Dancause N, Nudo RJ. Shaping plasticity to enhance recovery after injury. *Prog Brain Res* 2011; 192: 273–295.

23. Biernaskie J. Efficacy of rehabilitative experience declines with time after focal ischemic brain injury. *J Neurosci* 2004; 24: 1245–1254.
24. Kwakkel G, Meskers CGM, van Wegen EE, Lankhorst GJ, Geurts ACH, van Kuijk AA, et al. Impact of early applied upper limb stimulation: the EXPLICIT-stroke programme design. *BMC Neurol* 2008; 8: 49.
25. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair* 2008; 22: 78–90.
26. van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. Improving the Action Research Arm test: a unidimensional hierarchical scale. *Clin Rehabil* 2002; 16: 646–653.
27. 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: 14–19.
28. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil* 2003; 84: 950–963.
29. Kwakkel G, Winters C, van Wegen EEH, Nijland RHM, van Kuijk AAA, Visser-Meily A, et al. Effects of unilateral upper limb training in two distinct prognostic groups early after stroke. *Neurorehabil Neural Repair* 2016; 30: 804–816.
30. Wu CY, Chuang LL, Lin KC, Lee S Da, Hong WH. Responsiveness, minimal detectable change, and minimal clinically important difference of the nottingham extended activities of daily living scale in patients with improved performance after stroke rehabilitation. *Arch Phys Med Rehabil* 2011; 92: 1281–1287.
31. Nijland RHM, Van Wegen EEH, Harmeling-Van Der Wel BC, Kwakkel G. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS cohort study. *Stroke* 2010; 41: 745–750.
32. Simpson LA. Functional recovery following stroke: capturing changes in upper extremity function. *Neurorehabil Neural Repair* 2013; 27: 240–250.
33. Barclay-Goddard R, Lix LM, Tate R, Weinberg L, Mayo NE. Health-related quality of life after stroke: does response shift occur in self-perceived physical function? *Arch Phys Med Rehabil* 2011; 92: 1762–1769.
34. Dennis A, Bosnell R, Dawes H, Howells K, Cockburn J, Kischka U, et al. Cognitive context determines dorsal premotor cortical activity during hand movement in patients after stroke. *Stroke* 2011; 42: 1056–1061.
35. Ten Brink AF, Visser-Meily JMA, Nijboer TCW. Dynamic assessment of visual neglect: the Mobility Assessment Course as a diagnostic tool. *J Clin Exp Neuropsychol* 2018; 40: 161–172.



Appendix I. Inclusion flow diagram.