

Test-retest reliability and rater agreements of Assessment of Capacity for Myoelectric Control version 2.0

Helen Y. N. Lindner, PhD;^{1–2*} Ann Langius-Eklöf, PhD;³ Liselotte M. N. Hermansson, PhD^{1,4}

¹School of Health and Medical Sciences and ²Centre for Rehabilitation Research, Örebro University Hospital, Örebro, Sweden; ³Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden;

⁴Department of Prosthetics and Orthotics, Örebro University Hospital, Örebro, Sweden

Abstract—The Assessment of Capacity for Myoelectric Control (ACMC) is an observation-based tool that evaluates ability to control a myoelectric prosthetic hand. Validity evidence led to ACMC version 2.0, but the test-retest reliability and minimal detectable change (MDC) of the ACMC have never been evaluated. Investigation of rater agreements in this version was also needed because it has new definitions in certain rating categories and items. Upper-limb prosthesis users ($n = 25$, 15 congenital, 10 acquired; mean age 27.5 yr) performed one standardized activity twice, 2 to 5 wk apart. Activity performances were video-recorded and assessed by two ACMC raters. Data were analyzed by weighted kappa, intraclass correlation coefficient (ICC), and Bland-Altman method. For test-retest reliability, weighted kappa agreements were fair to excellent (0.52 to 1.00), $ICC_{2,1}$ was 0.94, and one user was located outside the limits of agreement in the Bland-Altman plot. MDC_{95} was less than or equal to 0.55 logits (1 rater) and 0.69 logits (2 raters). For interrater reliability, weighted kappa agreements were fair to excellent in both sessions (0.44 to 1.00), and $ICC_{2,1}$ was 0.95 (test) and 0.92 (retest). Intrarater agreement (rater 1) was also excellent ($ICC_{3,1}$ 0.98). Evidence regarding the reliability of the ACMC is satisfactory and MDC_{95} can be used to indicate change.

Key words: ACMC, assessment, capacity, myoelectric control, myoelectric prosthetic hand, prosthesis, prosthetic hand control, rater agreement, test-retest, upper limb.

INTRODUCTION

One important rehabilitation goal for upper-limb myoelectric prosthesis users is to increase their function-

ality so that they can perform their daily activities independently [1–2]. To achieve this goal, prosthesis users have to learn how to control the myoelectric prosthetic hand [3–4], such as using appropriate grip force or maintain holding objects in motion [4]. For evaluation of the client's progress in learning to control the prosthetic hand, a reliable and valid instrument that evaluates the ability to control a myoelectric prosthetic hand is needed.

Different instruments have been used to measure functional outcomes in users of upper-limb prostheses [5–7]. Although some of them have been used for many years in clinical practice, their validity evidence for the upper-limb prosthetic group has not been fully investigated [5,8]. One instrument, the Assessment of Capacity for Myoelectric Control (ACMC) [9–11], has been acknowledged in the field for its rigorous psychometric evaluations [5] and its

Abbreviations: ACMC = Assessment of Capacity for Myoelectric Control, AM-ULA = Activities Measure for Upper-Limb Amputees, CI = confidence interval, ICC = intraclass correlation coefficient, LDAPC = Limb Deficiency and Arm Prosthesis Centre, LOA = limits of agreement, MCID = minimal clinically important difference, MDC = minimal detectable change, MFRM = Many-Facets Rasch Model, MSE = mean squared error, PA = percentage agreement, SE = standard error, SEM = standard error of measurement.

*Address all correspondence to Helen Y. N. Lindner, PhD; Centre for Rehabilitation Research, Örebro University Hospital, SE 701 85 Örebro, Sweden; +46-19-6025882; fax: +46-19-255559. Email: helen.lindner@orebroll.se
<http://dx.doi.org/10.1682/JRRD.2013.09.0197>

potential to measure progress in users of upper-limb myoelectric prostheses [12]. The ACMC is an observation-based assessment that evaluates a person's ability to control a myoelectric prosthetic hand in bimanual activities [10]. The original ACMC (version 1.0) consists of 30 items, and its validations were carried out using Rasch analysis [10–11]. The results from the analyses led to item combination, clarification of item definition, and rating category redefinition [11]. Consequently, the ACMC version 2.0 consists of 22 items with a revised rating scale [13]. In both versions, an important aspect of ACMC that has not been evaluated is the stability of ACMC measures over time. This is usually evaluated in a test-retest situation, assuming that the individuals have not changed between the test and retest sessions [14]. Changes in level of fatigue or other personal factors from the test to retest sessions can affect the assessment scores [15], and rater inconsistency can also affect the scores, which has been shown in a rater agreement study of ACMC version 1.0 [9]. Another important clinical aspect of an instrument is the minimal detectable change (MDC) [16–17]. The MDC suggests the smallest change that can be detected by the instrument beyond measurement error [17–19]. For instruments that have an evaluative purpose, such as ACMC, the MDC is a useful clinical value to suggest whether a change between assessments is due to measurement error or true change. Hence, before ACMC version 2.0 can be recommended to measure change in the ability for myoelectric control, it is necessary to examine the stability of ACMC measures over time in a test-retest situation and estimate the MDC.

Being an observation-based assessment, the reliability of ACMC assessment scores also depends on the consistency of the raters. Even though intra- and interrater agreements have been evaluated in ACMC version 1.0 [9], it is necessary to evaluate ACMC version 2.0's interrater agreement because it has a revised rating scale and revised definitions in several items. The aim of the present study was to evaluate test-retest reliability and rater agreements of ACMC version 2.0.

METHODS

Participants

Potential participants were recruited from two hospitals in Sweden: the Limb Deficiency and Arm Prosthesis Centre (LDAPC) at the Örebro University Hospital and the Center for Arm Amputees at the Red Cross Hospital

in Stockholm. They were regular patients who attended the centers for prosthetic rehabilitation services. In order to examine whether the ACMC could produce consistent measurements, we aimed to recruit users of upper-limb myoelectric prostheses who were able to exercise stable myoelectric control. Therefore, we excluded prosthesis users who were (1) new users or just fitted with a new prosthesis, (2) not able to attend the clinic within 5 wk for the retest session, or (3) undergoing prosthetic training between test and retest sessions. According to Bonett [20], a sample of 21 subjects is required for a reliability study with two raters using intraclass correlation coefficient (ICC) and a desired precision of 0.2 and α value of 0.05. Twenty-five prosthesis users participated in the study between September 2009 and June 2012. Information about the study was given to the potential participants when they arrived at the centers. Formal written consent was obtained directly from the participants or, in the case of small children, from the parents. Demographic data were retrieved from the participants' medical records and are presented in **Table 1**. Ethical approval was obtained from the Uppsala Ethical Committee in Sweden.

Outcome Measures

Assessment of Capacity for Myoelectric Control Version 2.0

ACMC 2.0 consists of 22 items that assess six different aspects related to capacity for myoelectric control: the need for external support, grip force, coordination of both hands, different positions and in motion (timing), repetitive grasp and release, and the need for visual feedback. All items are rated on a four-point rating scale: 0 = not capable, 1 = somewhat capable, 2 = generally capable, and 3 = extremely capable. This gives a maximum raw score of 66. During an ACMC session, the prosthesis user performs a bimanual activity, either self-chosen or standardized. A certified ACMC rater observes how the prosthesis user controls the myoelectric prosthetic hand during the activity and rates the items. The potential influence of the activities on the ACMC ability measures was examined previously and shown to have no significant effect on the result [21]. In the present study, each participant performed a standardized activity in the test and retest sessions.

Prosthesis wearing time, that is, the number of hours the participant wore the prosthesis each day, was reported verbally by the participants or their parents. At the

Table 1.

Participant demographics ($n = 25$). Data shown as n unless otherwise indicated.

Characteristic	Value
Age, Mean (Range, Median)	27.5 (7–72, 23)
Sex	
Male	13
Female	12
Cause of Limb Absence	
Upper-Limb Reduction Deficiency	15
Acquired Amputation	10
Prosthetic Side	
Unilateral	
Right (Congenital)/Dominant (Acquired)	5/3
Left (Congenital)/Nondominant (Acquired)	10/6
Bilateral (Acquired)	
Dominant (Myoelectric), Nondominant (Passive)	1
Prosthetic Level	
Unilateral	
Transhumeral (Acquired)	5
Transradial (Congenital/Trauma)	15/4
Bilateral (Acquired)	
Dominant, Carpal (Myoelectric); Nondominant, Transhumeral (Passive)	1
Prosthetic Experience	
Unilateral	
<1 yr	3
1–4 yr	5
≥5 yr	16
Bilateral	
>5 yr	1
Prosthetic Wearing Time	
>8 h/d, 7 d/wk (Full Time)	12
4–8 h/d, 5–7 d/wk (Part Time)	5
<4 h, 1–7 d/wk (Occasional)	5
At Least Once a Month (Sporadic)	3

LDAPC, prosthetic wearing time is routinely categorized into five levels: 1 = full-time, >8 h/d, 7 d/wk; 2 = part-time, 4–8 h/d, 5–7 d/wk; 3 = occasional, <4 h, 1–7 d/wk; 4 = sporadic, at least once a month; and 5 = nonuser, new user, or stopped wearing for a period.

Assessment of Capacity for Myoelectric Control Version 2.0 Procedure

Each participant ($n = 25$) performed a standardized activity twice, at 2–5 wk apart. In order to achieve an even distribution of the activities for the sample, the allocation technique “minimization” [22] was used to assign the six standardized activities to the participants. The standardized activities were “repotting a plant” (4 participants), “a ready-to-assemble project” (5 participants), “setting a table for four persons” (4 participants), “mix-

ing a store-bought cake/pudding mix” (4 participants), “sorting bills or pictures” (4 participants), and “packing a suitcase for overnight stay” (4 participants). The activity room, the standardized activity, the materials, and their locations were the same in both sessions. The same occupational therapist ($n = 6$) gave the instructions about the activity procedure in both sessions. All performances were recorded on DVDs. Two experienced ACMC raters, rater 1 (5 yr clinical experience) and rater 2 (15 yr clinical experience), assessed the prosthesis users using the ACMC version 2.0 manual [13]. For test-retest reliability and interrater agreements, both raters assessed the participants separately in both test and retest sessions. For intrarater agreement, rater 1 assessed the test session video recordings twice, at a 4–5 wk interval.

Data Analysis

Test-Retest Reliability

Quadratic weighted κ [23] was used to examine test-retest agreement of individual items. Percentage agreement (PA) of each item for both sessions was also calculated. The strength of weighted κ was interpreted according to Fleiss et al.’s guidelines [23]: poor agreement for $\kappa \leq 0.40$, fair to good agreement for $0.41 < \kappa < 0.75$, and excellent agreement for $\kappa \geq 0.75$.

The ACMC is a Rasch-built measure [9–11,21], and the assessment result is routinely reported in Rasch ability measures. Therefore, the item raw scores are converted to Rasch ability measures using the Many-Facets Rasch Model (MFRM) [24]. The Rasch model (a formula) uses odds and natural logarithms to convert ordinal raw scores into interval measures [24–25], which gives a more accurate measure of ability. The unit of Rasch measure is logits (natural log of odds), and logits of greater magnitude represent increasing user ability. Each ability measure is accompanied by standard error (SE), which shows the precision of the measure. Detailed explanation of Rasch analysis and MFRM are described elsewhere [24,26–27]. In the present study, the MFRM calibrated a raw score range of 0–66 for ACMC to an ability range of –6.29 to 7.42 logits. Each participant had two ability measures, one for each session. The participant ability measures were normally distributed (with Shapiro-Wilk test) and were used to compute an ICC_{2,1} (two-way random effects model) [28–29]. The ICC is the ratio of between-groups variance to total variance [29]. As suggested by Kottner et al.’s guidelines for reporting reliability [30], ICC > 0.70 is good for research purposes and > 0.90 is needed for clinical purposes.

The Bland-Altman method [31–32] was used to examine the agreement of ACMC ability measures between both sessions. We calculated the difference in ability measures from both sessions for each participant and the 95 percent limits of agreement (LOA) of the mean difference for the whole group. The 95 percent LOA was estimated by mean difference \pm 1.96 standard deviation of the differences. As recommended by Bland and Altman [33], 95 percent of differences between measurements of the two sessions are expected to lie within the LOA. This difference is visualized in the Bland-Altman plot, where the individual differences are plotted against the mean of the two test sessions [33].

The amount of measurement error was calculated using the standard error of measurement (SEM). SEM quantifies the actual size of measurement variability in the same unit as the ACMC logits. The SEM was calculated using the mean squared error (MSE) from two-way analysis of variance, where $SEM = \sqrt{MSE}$ [34]. The SEM was then used to determine the MDC, which represents the smallest change necessary to exceed the measurement error of two measures to indicate a real change [16–18]. For a 95 percent confidence level of MDC, that is, MDC_{95} , which is an appropriate level for clinical use, the MDC_{95} was calculated using the formula $MDC_{95} = SEM \times 1.96\sqrt{2}$ [18]. The MDC_{95} can be expressed as percentage of the total possible ability range of the instrument [35–37], and as stated earlier, the total possible ability range is 0–66 raw scores or –6.29 to 7.42 logits (i.e., 13.71 logits).

Rater Agreements

Intrarater agreement for the test session was calculated for rater 1 and interrater agreements between the two raters were calculated for each session. PA and quadratic weighted κ statistics were used to examine rater agreement at the item level. Again, Fleiss et al.'s guidelines were used to interpret the weighted κ [23]. The $ICC_{3,1}$ (a two-way mixed effect model) was used to examine intrarater agreement and $ICC_{2,1}$ (a two-way random effect model) was used to examine interrater agreements. Again, Kottner et al.'s guidelines for reporting reliability was used to interpret the magnitude of ICC [30].

All data were analyzed using SPSS 21.0 (IBM Corporation; Armonk, New York) and FACETS Many-Facets software program 3.70.2 (Winsteps; Chicago, Illinois). A syntax for SPSS from IBM SPSS support [38] was downloaded for the calculation of weighted κ in SPSS.

RESULTS

Test-Retest Reliability

Weighted $\kappa \geq 0.75$ was shown in 11 items and their PAs were 74–100 percent; weighted κ 0.52–0.73 was shown in the remaining items and their PAs were ≥ 66 –96 percent (**Table 2**). The item “grasping with support” was scored with one rating category only, and hence, no weighted κ was calculated. This item was performed extremely capably, i.e., scored as 3, by all participants in both sessions. This is because this item is easily performed by prosthesis users who have acquired the basic prosthetic control. The average weighted κ was 0.76 (excluded 1 item with no weighted κ) and average PA was 85 percent. Test-retest $ICC_{2,1}$ was 0.94 (95% confidence interval [CI] 0.86–0.97).

The participant ability range for the test session was –0.71 to 2.79 logits (mean 0.97 logits, SE 0.23 logits); for the retest session it was –0.82 to 2.61 logits (mean 0.96 logits, SE 0.23 logits). In the Bland-Altman plot (**Figure**), the upper and lower LOA were 0.86 and –0.88, respectively. All except one participant, a non-full-time prosthesis user, were within the 95 percent LOA.

The SEM was 0.19 logits (rater 1), 0.20 logits (rater 2), and 0.25 logits (both raters together). This gave an MDC_{95} of 0.52 logits (rater 1), 0.55 logits (rater 2), and 0.69 logits (two raters). All MDC_{95} values were ≤ 5 percent of the total ability logit range of –6.29 to 7.42 (13.71 logits).

Rater Agreement

Weighted $\kappa \geq 0.75$ was shown in 16 items (PA 68%–100%) for the test session and in 11 items (PA 72%–100%) for the retest session (**Table 2**). Weighted κ 0.44–0.74 (PA 56%–96%) was shown in the remaining items in both sessions. One item in the test session and three items in the retest session were scored with one rating category only, and hence, no weighted κ was calculated. The average weighted κ for the test session was 0.82 and for the retest session it was 0.76 (excluding the items with no weighted κ in both sessions). The $ICC_{2,1}$ between the raters was 0.95 (95% CI 0.87–0.98) for the test session and 0.92 (95% CI 0.80–0.96) for the retest session.

For intrarater agreement of rater 1, the weighted κ values of the test session were all >0.80 and the PAs for each item were ≥ 96 percent. The $ICC_{3,1}$ for the test session was 0.98 (95% CI 0.95–0.99).

Table 2.

Assessment of Capacity for Myoelectric Control (ACMC): rater agreements in percentage agreement (PA) and weighted κ (95% confidence interval [CI]).

ACMC Item	Test-Retest Agreement (Both Raters)		Interrater Agreement For Test Session		Interrater Agreement For Retest Session	
	PA	Weighted κ (95% CI)	PA	Weighted κ (95% CI)	PA	Weighted κ (95% CI)
Need For External Support						
Grasping with Support *	100	— [†]	100	— [†]	100	— [†]
Power Grip Without Support	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)
Precision Grip Without Support	86	0.78 (0.50–1.02)	96	0.85 (0.61–1.08)	96	0.90 (0.67–1.13)
Holding with Support	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)	100	— [†]
Holding Without Support	98	0.92 (0.80–1.08)	92	0.88 (0.60–1.17)	84	0.94 (0.80–1.07)
Releasing with Support	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)	100	— [†]
Releasing Without Support	96	0.73 (0.36–1.08)	96	0.65 (0.02–1.28)	96	0.65 (0.02–1.28)
Grip Force						
Appropriate Grip Force *	78	0.54 (0.27–0.81)	76	0.78 (0.58–1.00)	64	0.44 (0.27–0.61)
Coordination Of Both Hands						
Coordinating During Grasping	74	0.75 (0.59–0.91)	68	0.86 (0.75–0.96)	80	0.80 (0.65–0.9)
Coordinating During Releasing *	66	0.59 (0.33–0.84)	64	0.63 (0.29–0.98)	64	0.59 (0.23–0.97)
Different Positions and In Motion (Timing)						
Grasping In Different Positions	90	0.95 (0.90–1.00)	84	0.90 (0.79–1.01)	80	0.89 (0.80–0.97)
Releasing In Different Positions	90	0.94 (0.89–0.99)	88	0.86 (0.72–1.01)	88	0.83 (0.67–0.99)
Timing During Grasping *	74	0.67 (0.47–0.87)	76	0.80 (0.66–0.96)	80	0.83 (0.70–0.98)
Timing During Releasing *	66	0.52 (0.31–0.73)	80	0.82 (0.64–0.97)	80	0.73 (0.50–0.95)
Holding In Motion	82	0.81 (0.63–0.99)	92	0.97 (0.93–1.01)	84	0.74 (0.46–1.02)
Repetitive Grasp And Release						
Repetitive Grasp And Release *	84	0.60 (0.28–0.81)	80	0.89 (0.80–0.99)	76	0.76 (0.53–0.99)
Need for Visual Feedback						
Grasping Without Visual Feedback [‡]	74	0.57 (0.31–0.82)	60	0.46 (0.06–0.84)	60	0.64 (0.31–0.97)
Appropriate Grip Force, Without Visual Feedback	86	0.58 (0.20–0.95)	72	0.78 (0.54–1.02)	72	0.83 (0.67–0.98)
Holding Without Visual Feedback	76	0.60 (0.35–0.86)	80	0.67 (0.34–0.98)	76	0.44 (0.00–0.89)
Holding in Motion, Without Visual Feedback	78	0.75 (0.57–0.92)	80	0.75 (0.51–0.99)	88	0.86 (0.68–1.04)
Releasing Without Visual Feedback	72	0.56 (0.33–0.78)	56	0.58 (0.27–0.89)	56	0.53 (0.16–0.82)
Repetitive Grasp and Release, Without Visual Feedback *	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)	100	1.00 (1.00–1.00)

*Newly combined item with newly combined definition.

[†]Both raters used only 1 rating category for this item; hence, no weighted κ was calculated.

[‡]Clarification of item definition.

DISCUSSION

This study is the first test-retest reliability assessment of ACMC and the second rater agreement evaluation of ACMC. Overall, the results from ICC, weighted κ , and the Bland-Altman plot support the test-retest reliability of

ACMC version 2.0. The ICCs and weighted κ values also support rater agreements.

Evidence of test-retest reliability in upper-limb prosthetic outcome measures is sparse [7]. This is partly because only a handful of outcome measures have been validated with upper-limb prosthesis users [6] and partly

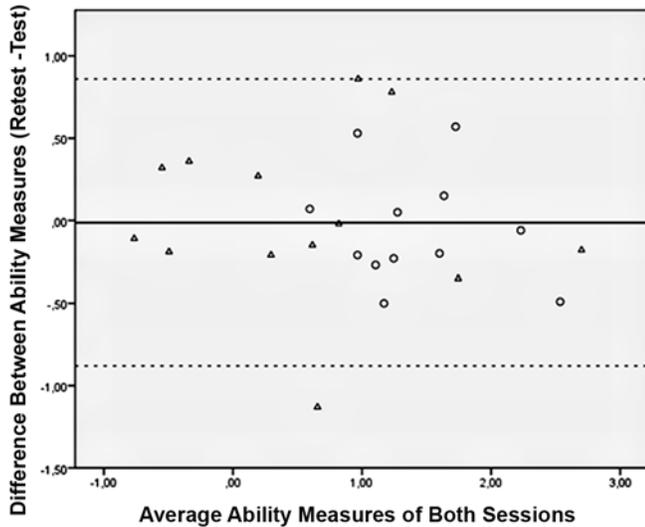


Figure.

Bland-Altman plot between test and retest sessions. O = full-time user, Δ = non-full-time user. Difference between test and retest sessions was plotted against average of 2 test sessions for each participant. It plots average logit measures of 2 test sessions (x-axis) against difference between test and retest logit measures (y-axis). Solid line in center represents mean of differences. Two dashed lines indicate 95% limits of agreement.

due to the difficulty in recruiting users for the retest session because many of them do not live near the prosthetic clinics. Since the development of ACMC, two more outcome measures that measure prosthetic function have been developed [39–40] and test-retest reliability was also evaluated in these two measures [40–41]. The ICC that we obtained is similar to the ICC in one of the measures [40] and higher than the other outcome measure [41]. This could be due to the wide range of ability in our sample because the ICC value is highly dependent on between-subject variance. In our study, the test-retest reliability of ACMC was examined among prosthesis users with different causes of limb absence, a wide age range, and varying prosthetic experience and wearing time. This was advantageous because we covered a wide range of subjects at whom the test is aimed. The results from the item weighted κ values were also excellent but were not as high as the test-retest ICC values. This is probably because some of the items were rated using only two of the rating scale categories instead of all four,

and hence, a low weighted κ value was obtained. The Bland-Altman method is independent of between-subject variance [31] and the majority of participants were with the 95 percent LOA, suggesting good agreement of participant ability measures from both sessions. Consequently, these three statistical methods provide evidence about the test-retest reliability of ACMC, which demonstrates that ACMC can produce consistent results.

The Bland-Altman plot was used to visualize the agreement of the test and retest sessions. One non-full-time prosthetic user fell outside the LOA of the Bland-Altman plot, indicating a good agreement between the test and retest sessions. In general, non-full-time users tended to be scattered slightly wider than full-time users in the plot. Although the sample was too small to draw any conclusion, prosthetic wearing time possibly played a role in the stability of ACMC measures. Wearing the prosthesis for >8 h/d gives the user the opportunity to use the device more often than non-full-time users, which may contribute to a more stable level of ability to control the prosthesis. Furthermore, for prosthesis users, it may take a lot of mental and physical effort to control a myoelectric prosthesis [42], and if the user is stressed or tired, his or her ACMC score could fluctuate unsystematically. Full-time users may have learned to live with their prostheses, and the related stress or tiredness could be lower than in non-full-time users. Further research on the relationship between prosthetic wearing time and ACMC ability measures could provide a better understanding of this relationship.

This is the first study to calculate the MDC regarding the use of ACMC to measure the ability to control a myoelectric prosthetic hand. The results showed MDC values <5 percent of the total possible ability range. The newly developed Activities Measure for Upper-Limb Amputees (AM-ULA) used raw scores for the calculation of MDC, and an MDC of 4.4 was reported [40]. It is not easy to compare the MDC values of the ACMC and the AM-ULA because these two instruments are quite different in their assessment procedures. Although unilateral users rarely use the prosthesis to perform one-hand activities [43–46], the AM-ULA requests unilateral users to perform such activities with the prosthesis so that the assessment can be compared with the use of a sound limb and a ceiling effect is also avoided. In contrast, ACMC assesses how the prosthesis is normally used to assist the sound hand to perform bimanual activities. It has been reported that prosthesis users prefer to be assessed in

their usual way of using the prosthesis [47], and consistent results are observed if they are allowed to perform the activities in their usual way [43]. Both assessment procedures are useful for different purposes, and it is important for clinicians to be aware of the different assessment procedures before they evaluate their clients. Now that the MDC for ACMC is calculated, further research can estimate the minimal clinically important difference (MCID) [48–50]. The MCID is the threshold at which a person or a group has just begun to experience what is an important improvement [50].

The results of rater agreements served two purposes in the present study. One purpose was to show whether the redefined rating categories and newly combined items were well understood. The other purpose was to assess overall reliability of the data and provide a glimpse of the influence of rater agreement on test-retest agreement. For the first purpose, the results confirmed that the redefined and newly combined items were well understood. Compared with rater reliability of ACMC version 1.0 [9], the average weighted κ was higher in the present study. One reason could have been that the rating category definitions and item definitions are clearer in ACMC 2.0. A relatively low PA and weighted κ were found for two items: “grasping without visual feedback” and “releasing without visual feedback.” It was not easy to assess these two items from the video recordings. From clinical experience, we know that these two items are easier to see from live ACMC assessments than from recordings. Further research comparing live and recorded ACMC assessments would improve our knowledge about the use of these two items. For the second purpose, the average weighted κ values of interrater agreement for each session were higher than the average weighted κ of the test-retest agreement. This supports the assumption that part of the variation originated from the prosthesis users, as discussed earlier.

One methodological concern was the recruitment of prosthesis users. The sample recruited to examine the test-retest reliability of an instrument must be sufficiently stable so that errors from the instrument itself or the measurement procedure can be estimated. For the present study, different criteria for the recruitment of stable prosthesis users were set. ACMC is designed for users of upper-limb myoelectric prostheses with different personal characteristics. Therefore, we decided to recruit users with different prosthetic wearing times and years of experience. Non-full-time users seemed to be less stable

than full-time users, which could have introduced more errors into their measurements. Another methodological concern was the interval between the test and retest sessions. We chose to wait for at least 2 wk before retesting the prosthesis users because we wanted to avoid any carryover effect, such as improvement in skill [15]. Some users rescheduled their retest sessions to within 5 wk, and we decided to include them in the study. The longer interval could also have contributed to larger variation between the two sessions. A third methodological concern was whether to collect data from live clinical situations or to use video recordings. We chose the latter because it gave us the opportunity to watch the performances repeatedly. However, video recordings may affect the behavior or performance of the prosthesis users differently in different sessions, thus influencing the test-retest results. Furthermore, the two raters were involved in video recording of some of the participants, which could have influenced their scores on some participants. This probably affected the interrater agreements of several items, as discussed earlier.

We used both ICC and weighted κ statistics to analyze different aspects of ACMC reliability. Weighted κ values were used for categorical data because they take into account the magnitude of the discrepancy in categorical data [23]. However, the weighted κ values depend on the number of categories used to rate the item [51], and this was shown in some of the items with relatively low κ values but high PAs. It has been suggested that ICC is equivalent to weighted κ [52]. However, we chose weighted κ at the item level and calculated the average weighted κ value because we wanted to compare the weighted κ results with a previous rater agreement study of ACMC [9]. Nevertheless, the results showed that the agreement for all items in the test-retest sessions was fair to excellent. The use of both ICC and item weighted κ values provided different evidence about the test-retest reliability of ACMC, which gave a better picture of the reliability of ACMC. The SEM was calculated for the first time in ACMC, and the assessments were rated by two experienced raters with different clinical experience. Experienced raters score more consistently than inexperienced raters [9]; thus, it is possible that the SEM is larger for less-experienced raters because their error rate is higher.

Despite the study limitations, the results of the present study demonstrate different aspects of the reliability of ACMC version 2.0. Based on these results, we can

recommend ACMC as a tool to follow the progress of users in controlling their myoelectric prostheses. The MDC is clinically useful for ACMC raters as a guideline to indicate whether the change is real.

CONCLUSIONS

Evidence regarding the stability of ACMC version 2.0 measures over time is satisfactory and the MDC value can be clinically useful. Further research is needed to determine the MCID and the responsiveness of ACMC.

ACKNOWLEDGMENTS

Author Contributions:

Analysis and interpretation of data: H. Y. N. Lindner.

Drafting of manuscript: H. Y. N. Lindner, L. M. N. Hermansson.

Critical revision of manuscript for important intellectual content:

A. Langius-Eklöf, L. M. N. Hermansson.

Statistical analysis: H. Y. N. Lindner.

Study supervision: A. Langius-Eklöf, L. M. N. Hermansson.

Financial Disclosures: The authors have declared that no competing interests exist.

Funding/Support: This material was based on work supported by a doctoral grant from Health Care Sciences Postgraduate School, Karolinska Institute, Solna, Sweden.

Additional Contributions: A special thanks to the upper-limb prosthesis users who gave their time for this study.

Institutional Review: Formal written consent was obtained directly from the participants or, in the case of small children, from the parents. The study was approved by the Uppsala Ethical Committee in Sweden (No. 231).

Participant Follow-Up: The authors do not plan to inform participants of the publication of this study. However, the study abstract in Swedish will be published on the Örebro University Hospital Web site.

REFERENCES

- Smurr LM, Gulick K, Yancosek K, Ganz O. Managing the upper extremity amputee: A protocol for success. *J Hand Ther.* 2008;21(2):160–75, quiz 176. [PMID:18436138] <http://dx.doi.org/10.1197/j.jht.2007.09.006>
- Alexander M, Matthews D. *Pediatric rehabilitation: Principles and practices.* 4th ed. New York (NY): Demos Medical; 2009. p. 210–29.
- Bouwsema H, Kyberd PJ, Hill W, van der Sluis CK, Bongers RM. Determining skill level in myoelectric prosthesis use with multiple outcome measures. *J Rehabil Res Dev.* 2012;49(9):1331–48. [PMID:23408215] <http://dx.doi.org/10.1682/JRRD.2011.09.0179>
- Hubbard S, Stocker D, Heger H. “Training” powered upper-limb prostheses. In: Muzumdar A, editor. *Powered upper-limb prostheses: Control, implementation and clinical implication.* Berlin (Germany): Springer-Verlag; 2004. p. 147–74.
- Hill W, Kyberd P, Norling Hermansson L, Hubbard S, Stavadahl Ø, Swanson S. Upper Limb Prosthetic Outcome Measures (ULPOM): A working group and their findings. *J Prosthet Orthot.* 2009;21(9):P69–82. <http://dx.doi.org/10.1097/JPO.0b013e3181ae970b>
- Lindner HY, Nätterlund BS, Hermansson LM. Upper limb prosthetic outcome measures: Review and content comparison based on International Classification of Functioning, Disability and Health. *Prosthet Orthot Int.* 2010;34(2):109–28. [PMID:20470058] <http://dx.doi.org/10.3109/03093641003776976>
- Wright FV. Prosthetic outcome measures for use with upper limb amputees: A systematic review of the peer-reviewed literature, 1970–2009. *J Prosthet Orthot.* 2009;21(9):P3–63. <http://dx.doi.org/10.1097/JPO.0b013e3181ae9637>
- McFarland LV, Hubbard Winkler SL, Heinemann AW, Jones M, Esquenazi A. Unilateral upper-limb loss: Satisfaction and prosthetic-device use in veterans and service-members from Vietnam and OIF/OEF conflicts. *J Rehabil Res Dev.* 2010;47(4):299–316. [PMID:20803400] <http://dx.doi.org/10.1682/JRRD.2009.03.0027>
- Hermansson LM, Bodin L, Eliasson AC. Intra- and interrater reliability of the assessment of capacity for myoelectric control. *J Rehabil Med.* 2006;38(2):118–23. [PMID:16546769] <http://dx.doi.org/10.1080/16501970500312222>
- Hermansson LM, Fisher AG, Bernspång B, Eliasson AC. Assessment of capacity for myoelectric control: A new Rasch-built measure of prosthetic hand control. *J Rehabil Med.* 2005;37(3):166–71. [PMID:16040474]
- Lindner HY, Linacre JM, Norling Hermansson LM. Assessment of capacity for myoelectric control: Evaluation of construct and rating scale. *J Rehabil Med.* 2009;41(6):467–74. [PMID:19479160] <http://dx.doi.org/10.2340/16501977-0361>
- Simon AM, Lock BA, Stubblefield KA. Patient training for functional use of pattern recognition-controlled prostheses. *J Prosthet Orthot.* 2012;24(2):56–64. [PMID:22563231] <http://dx.doi.org/10.1097/JPO.0b013e3182515437>
- Hermansson LM, Hill W, Lindner HY. *Assessment of Capacity for Myoelectric Control version 2.0: Training manual.* Örebro (Sweden): Örebro University Hospital; 2011.

14. Miller LA, McIntire SA, Lovler RL. Foundations of psychological testing: A practical approach. 4th ed. Thousand Oaks (CA): SAGE Publications; 2013. p. 156–89.
15. Yu CH. Test-retest reliability. In: Kempf-Leonard IK, editor. Encyclopedia of social measurement. San Diego (CA): Academic Press; 2005. p. 777–84.
16. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *J Rheumatol*. 2001;28(2):400–405. [PMID:11246687]
17. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;4:54. [PMID:16925807] <http://dx.doi.org/10.1186/1477-7525-4-54>
18. King MT. A point of minimal important difference (MID): A critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171–84. [PMID:21476819] <http://dx.doi.org/10.1586/erp.11.9>
19. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol*. 2010;63(7):804–6. [PMID:20399609] <http://dx.doi.org/10.1016/j.jclinepi.2009.12.015>
20. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*. 2002; 21(9):1331–35. [PMID:12111881] <http://dx.doi.org/10.1002/sim.1108>
21. Lindner HY, Eliasson AC, Hermansson LM. Influence of standardized activities on validity of Assessment of Capacity for Myoelectric Control. *J Rehabil Res Dev*. 2013;50(10):1391–1400. [PMID:24699974] <http://dx.doi.org/10.1682/JRRD.2012.12.0231>
22. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005;330(7495):843. [PMID:15817555] <http://dx.doi.org/10.1136/bmj.330.7495.843>
23. Fleiss JL, Levin B, Cho P. Statistical methods for rates and proportions. 3rd ed. New York (NY): Wiley; 2003. p. 598–626.
24. Eckes T. Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments. New York (NY): Peter Lang Publishing Group; 2011.
25. Wright BD, Stone MH. Best test design. Chicago (IL): MESA; 1979. p. 69.
26. Linacre JM, Wright BD. Construction of measures from many-facet data. *J Appl Meas*. 2002;3(4):486–512. [PMID:12486312]
27. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah (NJ): Lawrence Erlbaum Associates; 2007. p. 29–48.
28. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46. <http://dx.doi.org/10.1037/1082-989X.1.1.30>
29. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–28.
30. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011; 64(1):96–106. [PMID:21130355] <http://dx.doi.org/10.1016/j.jclinepi.2010.03.002>
31. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;20(5):337–40. [PMID:2257734] [http://dx.doi.org/10.1016/0010-4825\(90\)90013-F](http://dx.doi.org/10.1016/0010-4825(90)90013-F)
32. Bland JM, Altman DG. Applying the right statistics: Analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22(1):85–93. [PMID:12858311] <http://dx.doi.org/10.1002/uog.122>
33. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2): 135–60. [PMID:10501650] <http://dx.doi.org/10.1191/096228099673819272>
34. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231–40. [PMID:15705040]
35. Buffart LM, Roebroeck ME, Janssen WG, Hoekstra A, Hovius SE, Stam HJ. Comparison of instruments to assess hand function in children with radius deficiencies. *J Hand Surg Am*. 2007;32(4):531–40. [PMID:17398365] <http://dx.doi.org/10.1016/j.jhssa.2007.01.011>
36. Lu WS, Wang CH, Lin JH, Sheu CF, Hsieh CL. The minimal detectable change of the simplified stroke rehabilitation assessment of movement measure. *J Rehabil Med*. 2008;40(8):615–19. [PMID:19020694] <http://dx.doi.org/10.2340/16501977-0230>
37. Romero S, Bishop MD, Velozo CA, Light K. Minimum detectable change of the Berg Balance Scale and Dynamic Gait Index in older persons at risk for falling. *J Geriatr Phys Ther*. 2011;34(3):131–37. [PMID:21937903] <http://dx.doi.org/10.1519/JPT.0b013e3182048006>
38. IBM. Weighted kappa, Kappa for ordered categories [Internet]. Armonk (NY): IBM Corporation; 2011 [cited 2012 Oct 30]. Available from: <https://www-304.ibm.com/support/docview.wss?uid=swg21477357>
39. Bagley AM, Molitor F, Wagner LV, Tomhave W, James MA. The Unilateral Below Elbow Test: A function test for

- children with unilateral congenital below elbow deficiency. *Dev Med Child Neurol*. 2006;48(7):569–75.
[\[PMID:16780626\]](#)
<http://dx.doi.org/10.1017/S0012162206001204>
40. Resnik L, Adams L, Borgia M, Delikat J, Disla R, Ebner C, Walters LS. Development and evaluation of the Activities Measure for Upper-Limb Amputees. *Arch Phys Med Rehabil*. 2013;94(3):488–94.
41. Buffart LM, Roebroek ME, van Heijningen VG, Pesch-Batenburg JM, Stam HJ. Evaluation of arm and prosthetic functioning in children with a congenital transverse reduction deficiency of the upper limb. *J Rehabil Med*. 2007;39(5):379–86. [\[PMID:17549329\]](#)
<http://dx.doi.org/10.2340/16501977-0068>
42. Bongers RM, Kyberd PJ, Bouwsema HM, Kenney LP, Plettenburg DH, Van der Sluis CK. Bernstein's levels of construction of movements applied to upper limb prosthetics. *J Prosthet Orthot*. 2012;24(2):67–76.
<http://dx.doi.org/10.1097/JPO.0b013e3182532419>
43. Black N, Biden EN, Rickards J. Using potential energy to measure work related activities for persons wearing upper limb prostheses. *Robotica*. 1999;23(3):319–27.
<http://dx.doi.org/10.1017/S0263574704001341>
44. Light CM, Chappell PH, Kyberd PJ. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity. *Arch Phys Med Rehabil*. 2002;83(6):776–83.
[\[PMID:12048655\]](#)
<http://dx.doi.org/10.1053/apmr.2002.32737>
45. van Lunteren A, van Lunteren-Gerritsen GH, Stassen HG, Zuithoff MJ. A field evaluation of arm prostheses for unilateral amputees. *Prosthet Orthot Int*. 1983;7(3):141–51.
[\[PMID:6647010\]](#)
46. Bhaskaranand K, Bhat AK, Acharya KN. Prosthetic rehabilitation in traumatic upper limb amputees (an Indian perspective). *Arch Orthop Trauma Surg*. 2003;123(7):363–66.
[\[PMID:12827395\]](#)
<http://dx.doi.org/10.1007/s00402-003-0546-4>
47. Hebert JS, Lewicke J. Case report of modified Box and Blocks test with motion capture to measure prosthetic function. *J Rehabil Res Dev*. 2012;49(8):1163–74.
[\[PMID:23341309\]](#)
<http://dx.doi.org/10.1682/JRRD.2011.10.0207>
48. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–15.
[\[PMID:2691207\]](#)
[http://dx.doi.org/10.1016/0197-2456\(89\)90005-6](http://dx.doi.org/10.1016/0197-2456(89)90005-6)
49. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): A literature review and directions for future research. *Curr Opin Rheumatol*. 2002;14(2):109–14. [\[PMID:11845014\]](#)
<http://dx.doi.org/10.1097/00002281-200203000-00006>
50. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Arch Phys Med Rehabil*. 2008;89(9):1693–1700. [\[PMID:18760153\]](#)
<http://dx.doi.org/10.1016/j.apmr.2008.02.022>
51. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology*. 1996;7(2):199–202. [\[PMID:8834562\]](#)
<http://dx.doi.org/10.1097/00001648-199603000-00016>
52. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613–19.
<http://dx.doi.org/10.1177/001316447303300309>

Submitted for publication September 5, 2013. Accepted in revised form November 26, 2013.

This article and any supplementary material should be cited as follows:

Lindner HY, Langius-Eklöf A, Hermansson LM. Test-retest reliability and rater agreements of Assessment of Capacity for Myoelectric Control version 2.0. *J Rehabil Res Dev*. 2014;51(4):635–44.
<http://dx.doi.org/10.1682/JRRD.2013.09.0197>

